Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantum Error Correction by Purification

Jonathan Raghoonanan

jonathanraghoonanan@gmail.com

New York University Shanghai https://orcid.org/0000-0002-6912-7443

Tim Byrnes

New York University Shanghai

Article

Keywords:

Posted Date: March 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9119304/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Quantum Error Correction by Purification

Jonathan Raghoonanan[1,2] and Tim Byrnes[1,3,4,2,*]

[1] *New York University Shanghai; NYU-ECNU Institute of Physics at NYU Shanghai, 567 West Yangsi Road, Shanghai, 200124, China.*
[2] *Department of Physics, New York University, New York, NY 10003, USA*
[3] *State Key Laboratory of Precision Spectroscopy, School of Physical and Material Sciences, East China Normal University, Shanghai 200062, China*
[4] *Center for Quantum and Topological Systems (CQTS), NYUAD Research Institute, New York University Abu Dhabi, UAE*

(Dated: March 12, 2026)

We present a general-purpose quantum error correction primitive based on state purification via the SWAP test, which we refer to as *purification quantum error correction* (PQEC). This method operates on N noisy copies, requires minimally $O(M \log_2 N)$ data qubits to process the M-qubit inputs. In a similar way to standard QEC, the purification steps may be interleaved within a quantum algorithm to suppress the logical error rate. No postselection is performed and no knowledge of the state is required. We analyze its performance under a variety of error channels and find that PQEC is highly effective at boosting fidelity and reducing logical error rates, particularly for the depolarizing channel. Error thresholds for the local depolarizing channel are found to be 75% for any register size. For local dephasing, the error threshold is reduced to 50% but may be boosted using twirling.

I. INTRODUCTION

Quantum error correction (QEC) is essential for reliable quantum computation: without active suppression of noise, circuit depth is limited by decoherence, control errors, and leakage [1–4]. Stabilizer codes—notably the surface code—form the dominant pathway toward scalable architectures due to locality and high thresholds [5, 6]. Yet, the resource overheads for full fault tolerance (magic state factories, repeated rounds of syndrome extraction, and real-time decoding) remain a central challenge for current devices, motivating alternative techniques [7, 8]. Alongside fully fledged fault-tolerant QEC, a wide range of lighter-weight strategies has been an alternative direction of interest. An early example of this is the use of decoherence-free subspaces and noiseless subsystems, where information is encoded in symmetry-protected sectors that are immune to specific collective noise models [8–10]. Dynamical decoupling uses sequences of fast control pulses to average out system–bath couplings and extend coherence times without encoding into a code space [11]. More recently, quantum error-mitigation protocols infer zero-noise expectation values from noisy data via extrapolation in noise strength or quasi-probability cancellation, and have become a leading approach for near-term devices where full fault tolerance is out of reach [12, 13].

A closely related line of work, traditionally framed in the context of long-distance communication and resource preparation rather than computation, is entanglement purification [14, 15]. In a typical purification protocol, one starts with several noisy copies of a quantum state, and using only local operations and classical communication (LOCC), a single higher fidelity copy of the target state is obtained after post-selection. In most purification protocols, the *target state is fixed and known*, such as a specific Bell pair, GHZ state, or graph state. Furthermore, they typically employ postselection where particular measurement outcomes are desirable and otherwise the resultant state is discarded. For example, the original protocols of Bennett, Wootters, and co-workers distilled high-fidelity Bell pairs from many noisy copies, together with twirling operations to an isotropic Bell-diagonal form [16]. Deutsch, Sanpera, and co-workers incorporated such purification steps into quantum privacy amplification for quantum key distribution over noisy channels [17]. Subsequent work generalized these ideas to multipartite entangled states [18, 19]. Purification is considered essential for quantum repeaters and networking architectures [20–22]. Modern variants optimize gate counts, success probabilities, and robustness for concrete network architectures [23, 24], but the basic setting remains one in which a known resource state is distilled offline from many noisy copies.

Historically, purification has been mainly considered in the context of quantum state initialization due to the fact that the target state is fixed and known. However, a more general *unknown-state* purification can be performed, where an arbitrary unknown state is purified. This can be performed when a typical requirement of purification is relaxed, such as only using LOCC operations. Here, the task is to recover a nearly pure copy of an unknown state from many noisy copies, without prior knowledge of which pure state was prepared. Early attempts include optimal collective processing and universal purification procedures, but are generally limited to single qubits, asymptotic regimes, and may require having limited knowledge of the initial state [25, 26]. More recently, Childs and co-workers showed that a recursive, SWAP-test-based protocol can efficiently boost the fidelity of

* tim.byrnes@nyu.edu

an unknown pure state given multiple noisy copies, for qudits of arbitrary dimension [27]. Here, the SWAP test consists of a quantum primitive that determines the fidelity between two states [28]. In the purification context, by post-selecting on the symmetric outcome of the test, one can iteratively amplify overlap with the dominant eigenvector of the input density operator. Because the protocol acts on whatever pure state occupies the dominant eigenvalue, it naturally aligns with the idea of protecting unknown encoded information rather than a fixed resource state.

In parallel, several recent purification-based quantum error mitigation (QEM) approaches exploit the expectation that the ideal output of a computation is approximately pure. A prominent example is given by Huggins, McClean and co-workers [29], which uses collective measurements on multiple copies to estimate expectation values with respect to an effectively purified state. The approach is called *virtual distillation* because the state itself does not become purified, but the use of a modified observable makes the effective (or virtual) state ρ^N, while physically only the state $\rho^{\otimes N}$ is present. The state ρ^N approaches the dominant eigenvector of ρ as N increases, and thus corresponds to a purified state. Related multi-copy purification/verification techniques—including postselected variants—have been experimentally benchmarked in quantum simulation settings and shown to reduce error by orders of magnitude, with sampling overhead becoming the dominant practical limitation at scale [30]. These works primarily target the mitigation of *measured observables* without necessarily outputting a purified state for further coherent processing.

In this paper, we merge several of the concepts above — namely unknown-state purification, virtual distillation, and quantum error correction — to show that the SWAP-based purification can implement unknown-state purification as an error-correction primitive. We refer to this scheme as *purification quantum error correction* (PQEC). Figure 1 illustrates the basic logical flow: noisy copies arrive sequentially and are pairwise merged via SWAP tests to form progressively purified states. Layers can be interwoven between unitary steps in a larger algorithm such that errors may be corrected during the evolution of a quantum circuit. Rather than relying upon codewords as in traditional QEC, the protocol processes a sequence of noisy copies and, by pairing and merging successes, the purity of the state improves by consuming copies of the state. Importantly, no postselection needs to be performed and all measurement outcomes are used to obtain the final estimate. While Fig. 1 suggests an exponential number of data qubits, this is only the simplest implementation; resources can be compressed to $O(M \log_2 N)$ coherent data qubits to handle N lots of M-qubit input copies of the noisy state, achieving the same purity. We analyze the performance of PQEC in a QEC framework and find quantities of interest such as the logical error rate and the error threshold.

II. SWAP-BASED PURIFICATION

In this section, we review the SWAP-based purification approach.

A. The SWAP test

The core operation in our purification protocol is the SWAP test applied to two identical noisy states [27]. Let A, B be two quantum registers, each with a D-dimensional Hilbert space ($D = 2^M$ for M qubits). In the SWAP test, as shown in Fig. 1(b), an ancilla qubit controls the SWAP operation on the two registers. The measurement outcomes on the ancilla partition the AB registers into symmetric and antisymmetric subspaces. The $|0\rangle$ outcome projects AB onto the symmetric subspace

$$\Pi_+ |\psi\rangle_A \otimes |\phi\rangle_B = \frac{1}{2}\left(|\psi\rangle_A \otimes |\phi\rangle_B + |\phi\rangle_A \otimes |\psi\rangle_B\right), \quad (1)$$

while $|1\rangle$ projects onto the antisymmetric subspace,

$$\Pi_- |\psi\rangle_A \otimes |\phi\rangle_B = \frac{1}{2}\left(|\psi\rangle_A \otimes |\phi\rangle_B - |\phi\rangle_A \otimes |\psi\rangle_B\right). \quad (2)$$

The projectors onto the symmetric/antisymmetric subspaces are given by

$$\Pi_\pm = \tfrac{1}{2}\big(I \pm \mathrm{SWAP}\big), \quad (3)$$

where $\mathrm{SWAP} = [(I_A I_B + X_A X_B + Y_A Y_B + Z_A Z_B)/2]^{\otimes M}$, $\vec{\sigma} = (X, Y, Z)$ are Pauli matrices, and I is the identity matrix. The projectors satisfy $\Pi_+ + \Pi_- = I$, $\Pi_\pm^2 = \Pi_\pm$. For M-qubit registers, CSWAP is realized as M parallel Fredkin gates (one per physical qubit pair) [1, 3, 28].

B. SWAP gadget

In the analysis that follows we will make use of the SWAP gadget, which involves performing the SWAP test projection (3), following by discarding the second register. The resulting state for an initial product state $\rho \otimes \rho'$ is [27, 28]

$$\begin{aligned}\rho_\pm =& \frac{\mathrm{Tr}_B\left[\Pi_\pm(\rho\otimes\rho')\Pi_\pm\right]}{P_\pm} \\ =& \frac{1}{4P_\pm}\Big(\mathrm{Tr}_B[\rho\otimes\rho'] + \mathrm{Tr}_B[\rho'\otimes\rho] \\ & \pm \mathrm{Tr}_B[\mathrm{SWAP}(\rho\otimes\rho')] \pm \mathrm{Tr}_B[(\rho\otimes\rho')\mathrm{SWAP}]\Big) \\ =& \frac{\rho + \rho' \pm (\rho\rho' + \rho'\rho)}{4P_\pm}, \end{aligned} \quad (4)$$

where we used the "SWAP trick": $\mathrm{Tr}[\mathrm{SWAP}(\rho\otimes\rho')] = \mathrm{Tr}[\rho\rho']$ [3]. The probability of the outcomes are

$$P_\pm = \frac{1}{2}(1 \pm \mathrm{Tr}[\rho\rho']). \quad (5)$$

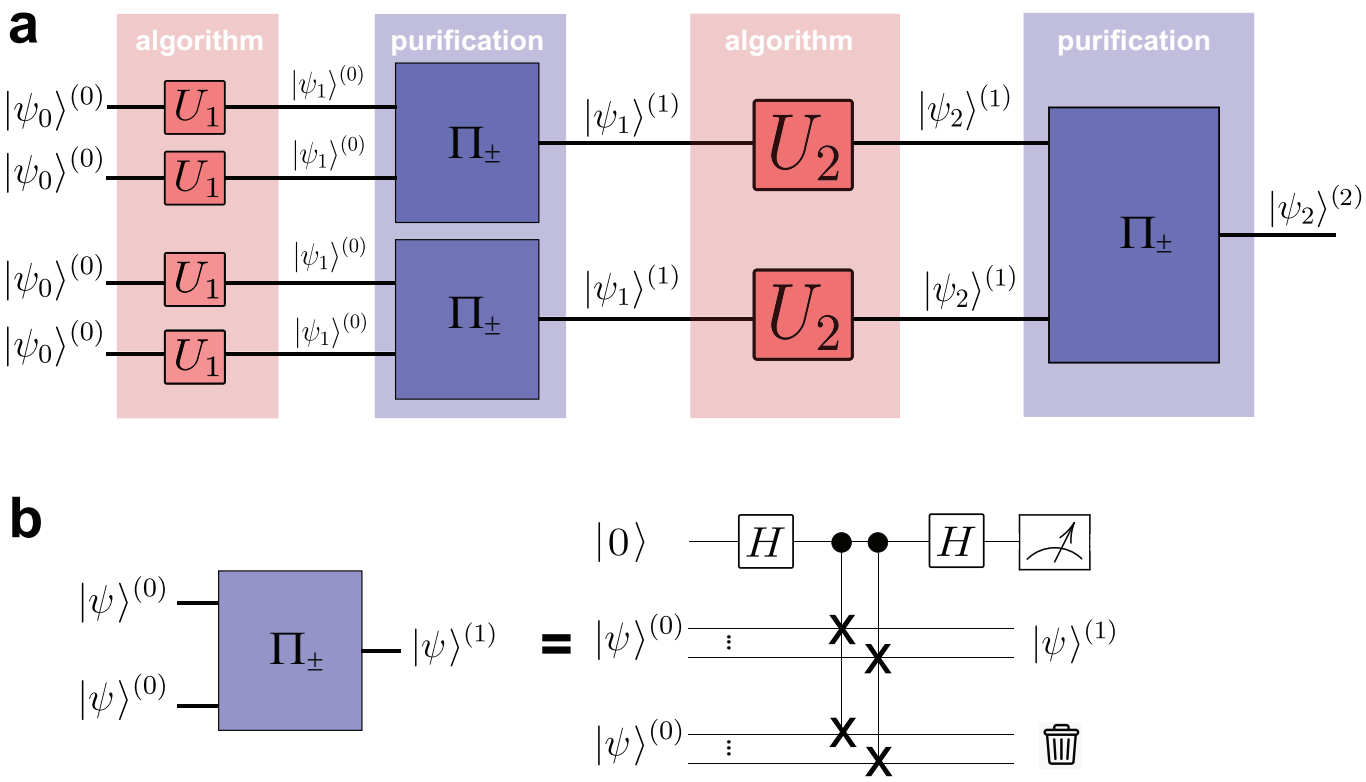


FIG. 1. Purification quantum error correction implementation in a quantum circuit. (a) Depth-efficient configuration of SWAP-based purifiers being interleaved within a M-qubit quantum algorithm consisting of a sequence of unitary gates U_n. The whole quantum circuit is given by the unitary $U = \prod_n U_n$. The state $|\psi_n\rangle^{(\ell)}$ refers to the state at the nth stage of the quantum algorithm, with ℓ rounds of purification. (b) The SWAP gadget and its definition. The Π_+ measurement corresponds to the ancilla outcome $|0\rangle$ and Π_- corresponds to $|1\rangle$. The second copy of the state is discarded (traced out). Thick horizontal lines denote qubit registers consisting of M qubits, thin horizontal lines are single qubits.

III. PURIFICATION QEC

In this section, we introduce the basic formulation of PQEC. We show how the SWAP gadget can be used within a quantum circuit to successively purify the output with the binary tree structure as shown in Fig. 1(a).

A. Basic idea: Single SWAP purification

The purification effect of the SWAP gadget can be seen by setting the input states are identical $\rho' = \rho$ in (4)

$$\rho_\pm = \frac{\rho \pm \rho^2}{2P_\pm}. \tag{6}$$

For the + outcome, the resulting state is a mixture of the original state ρ and ρ^2. The state ρ^2 may be considered a purified version of the original state ρ as it concentrates the distribution towards the largest eigenvalue [29]

$$\mathcal{P}(\rho) := \frac{\rho^2}{\mathrm{Tr}(\rho^2)} = \frac{\sum_i \lambda_i^2 |\lambda_i\rangle\langle\lambda_i|}{\sum_i \lambda_i^2}, \tag{7}$$

where the density matrix eigendecomposition is $\rho = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$.

One simple way to take advantage of the purification would be to postselect on the ρ_+ outcome and use this as the purified state. While this indeed has the effect of reducing smaller eigenvalues originating from errors [27], it has the drawback of having an exponentially suppressed success probability for many rounds of purification. Without postselection, taking the average over the two outcomes precisely removes the purified component ρ^2

$$\rho' = P_+\rho_+ + P_-\rho_- = \rho, \tag{8}$$

hence is not effective with respect to correcting errors.

Suppose, instead of adding the two contributions, one *subtracted* them; we would then obtain

$$P_+\rho_+ - P_-\rho_- = \rho^2 \tag{9}$$

thereby gaining access to the purified component. This is not an average state in the same way that (8) is; nevertheless, it can be accessed in the following way. For example, if one retains the ancilla qubit in the SWAP test circuit

of Fig. 2(b), we have the state

$$\begin{aligned}\rho' &= P_+|0\rangle\langle 0|\otimes\rho_+ + P_-|1\rangle\langle 1|\otimes\rho_- \\ &= \frac{1}{2}(I\otimes\rho + Z\otimes\rho^2).\end{aligned} \tag{10}$$

This gives a convenient way of accessing the ρ and ρ^2 components of the resulting state. If we consider any observable O of the system, we then have

$$\langle I\otimes O\rangle = \mathrm{Tr}(I\otimes O\rho') = \mathrm{Tr}(O\rho), \tag{11}$$

$$\langle Z\otimes O\rangle = \mathrm{Tr}(Z\otimes O\rho') = \mathrm{Tr}(O\rho^2), \tag{12}$$

due to the traceless property of Z. We may then "pick out" the ρ^2 component of the resultant state by modifying the observable with a Z operator. Alternatively, one may omit the ancilla by keeping track of the outcome of the SWAP measurement and putting the minus sign in (9) when averaging over outcomes (discussed further in Sec. III C).

B. Multiple purification rounds

We now show the effect of multiple purification rounds, as illustrated in Fig. 1(a). For now, let us disregard the quantum algorithm part of the purification, in other words set $U_n = I$. Consider the effect of one of the SWAP gadgets somewhere in the binary tree. The input states depend on all the prior measurements that have been performed on them. After ℓ rounds of purification, there are $2^\ell - 1$ measurement outcomes. For example, for the $\ell = 2$ circuit shown in Fig. 1(a), there are a total of 3 SWAP tests, giving 3 outcomes. We denote the outcomes by

$$\vec{\sigma}_\ell = (\sigma_1, \sigma_2, \ldots, \sigma_{2^\ell-1}), \tag{13}$$

where $\sigma_i \in \pm 1$. The conditional state after ℓ rounds is then written as $\rho_{\vec{\sigma}_\ell}$. Two states $\rho_{\vec{\sigma}_\ell}, \rho_{\vec{\sigma}'_\ell}$ being purified after ℓ rounds will not, in general, be the same state due to the different measurement outcomes. Utilizing (4), the state after measurement is

$$\rho_{\vec{\sigma}_{\ell+1}} = \frac{\rho_{\vec{\sigma}_\ell} + \rho_{\vec{\sigma}'_\ell} + \sigma(\rho_{\vec{\sigma}_\ell}\rho_{\vec{\sigma}'_\ell} + \rho_{\vec{\sigma}'_\ell}\rho_{\vec{\sigma}_\ell})}{4P_\sigma}, \tag{14}$$

where $\vec{\sigma}_{\ell+1} = (\vec{\sigma}_\ell, \vec{\sigma}'_\ell, \sigma)$. Eq. (14) provides a recursive definition that can be used to find the resultant state after several purification rounds.

For example, after $\ell = 2$ rounds starting from the input state $\rho^{\otimes 4}$, the final state may be evaluated as

$$\begin{aligned}\rho_{\sigma_1\sigma_2\sigma_3} =& \frac{1}{8}\Big[\left(\frac{1}{P_{\sigma_1}P_{\sigma_3}} + \frac{1}{P_{\sigma_2}P_{\sigma_3}}\right)\rho \\ &+ \left(\frac{\sigma_1}{P_{\sigma_1}P_{\sigma_3}} + \frac{\sigma_2}{P_{\sigma_2}P_{\sigma_3}} + \frac{\sigma_3}{P_{\sigma_1}P_{\sigma_2}P_{\sigma_3}}\right)\rho^2 \\ &+ \left(\frac{\sigma_2\sigma_3 + \sigma_1\sigma_3}{P_{\sigma_1}P_{\sigma_2}P_{\sigma_3}}\right)\rho^3 + \left(\frac{\sigma_1\sigma_2\sigma_3}{P_{\sigma_1}P_{\sigma_2}P_{\sigma_3}}\right)\rho^4\Big].\end{aligned} \tag{15}$$

We again see that the final state is a mixture of various degrees of purification ρ^m. Similar to the $\ell = 1$ case of (8), taking an average of all the outcomes simply produces the original state

$$\sum_{\sigma_1\sigma_2\sigma_3} P_{\sigma_1}P_{\sigma_2}P_{\sigma_3}\rho_{\sigma_1\sigma_2\sigma_3} = \rho. \tag{16}$$

Note that all other terms evaluate to zero since $\sum_{\sigma_i}\sigma_i = 0$. We may extract the higher powers of ρ by multiplying by suitable factors of σ_i and using the fact that $\sigma_i^2 = 1$. The highest power of ρ can be extracted using

$$\mathcal{P}_2(\rho) := \sum_{\sigma_1\sigma_2\sigma_3} P_{\sigma_1}P_{\sigma_2}P_{\sigma_3}\sigma_1\sigma_2\sigma_3\rho_{\sigma_1\sigma_2\sigma_3} = \rho^4. \tag{17}$$

This is the same logic as that formulated with ancilla qubits, as given in (9). Here, the sign rule when performing the average is the total parity $\sigma_1\sigma_2\sigma_3$, which can be included when averaging over all the SWAP measurement outcomes.

In general, we find that averaging the conditional density matrix over all outcomes gives the same state as the initial state

$$\sum_{\vec{\sigma}_\ell} P_{\vec{\sigma}_\ell}\rho_{\vec{\sigma}_\ell} = \rho, \tag{18}$$

where the probability of the outcome $\vec{\sigma}_\ell$ is

$$P_{\vec{\sigma}_\ell} = \prod_{i=1}^{2^\ell-1} P_{\sigma_i}. \tag{19}$$

Meanwhile, the most purified component of the density matrix may be extracted according to

$$\sum_{\vec{\sigma}_\ell} P_{\vec{\sigma}_\ell}\Omega_{\vec{\sigma}_\ell}\rho_{\vec{\sigma}_\ell} = \rho^{2^\ell}, \tag{20}$$

where we defined the total parity of the measurement outcomes as

$$\Omega_{\vec{\sigma}_\ell} = \prod_{i=1}^{2^\ell-1}\sigma_i. \tag{21}$$

An important property of (20) is that the coefficient of ρ^{2^ℓ} is unity, which shows that one is able to extract the purified component with the same amplitude as the original state (18).

As the normalized version of the ℓ-fold purified state, we define

$$\mathcal{P}_\ell(\rho) := \frac{\rho^N}{\mathrm{Tr}(\rho^N)} = \frac{\sum_i \lambda_i^N|\lambda_i\rangle\langle\lambda_i|}{\sum_i\lambda_i^N}, \tag{22}$$

where the number of copies in this case is $N = 2^\ell$. As with (7), this concentrates the distribution towards the largest eigenvalue [29].

We note that the above procedure attains the equivalent effect of virtual distillation from Ref. [29] with $N = 2^\ell$ copies. The primary difference here is that the SWAP test purifications are explicitly performed, such that the projected state contains a purified component (6), rather than employing a modified observable operator.

C. Extracting observables

We now discuss how observables are evaluated from the purified state above. First, any observable O using the purified version of the state (22) evaluates as [29]

$$\langle O\rangle_\ell = \frac{\mathrm{Tr}(O\rho^N)}{\mathrm{Tr}(\rho^N)}, \tag{23}$$

where $N = 2^\ell$ is the number of copies of the original state in this case. Substituting (20) into (23), we have

$$\langle O\rangle_\ell = \frac{\sum_{\vec{\sigma}_\ell} P_{\vec{\sigma}_\ell}\Omega_{\vec{\sigma}_\ell}\langle O\rangle_{\vec{\sigma}_\ell}}{\sum_{\vec{\sigma}_\ell} P_{\vec{\sigma}_\ell}\Omega_{\vec{\sigma}_\ell}}, \tag{24}$$

where

$$\langle O\rangle_{\vec{\sigma}_\ell} := \mathrm{Tr}(O\rho_{\vec{\sigma}_\ell}) \tag{25}$$

is the expectation value of the observable with respect to the state with the measurement outcome $\rho_{\vec{\sigma}_\ell}$.

From (24) we may estimate the overhead associated with performing PQEC. In addition to the statistics required to obtain the expectation value of the observable O in (25), one additionally needs statistics to estimate both the numerator and denominator in (24). The standard error of (24) is given by (see Appendix A)

$$\epsilon \approx \frac{\sqrt{\mathrm{Var}(\Omega_{\vec{\sigma}_\ell}(\langle O\rangle_{\vec{\sigma}_\ell} - \langle O\rangle_\ell)}}{\sqrt{N_{\mathrm{samp}}}\mathrm{Tr}(\rho^N)} \tag{26}$$

where N_{samp} is the number of samples, and the variance is taken with respect to the probability distribution $P_{\vec{\sigma}_\ell}$. For the case that O is normalized so that $\|O\|_\infty \leq 1$ (e.g. Pauli strings), then the numerator of (26) can be bounded by 1 (see Appendix A). We may then rearrange for N_{samp} to obtain the overhead

$$N_{\mathrm{samp}} \lesssim \frac{1}{\epsilon^2(\mathrm{Tr}(\rho^N))^2}. \tag{27}$$

We see that there is an additional overhead that increases with the impurity of the state. For example, in the limit of perfectly pure input states, only the symmetric outcome Π_+ is obtained, and the probability distribution $P_{\vec{\sigma}_\ell}$ is completely peaked at this outcome. For highly mixed states, there is a larger variation in $P_{\vec{\sigma}_\ell}$, resulting in a larger overhead in performing the sampling over PQEC measurements.

D. Purifying quantum circuits

Up to this point, we have not considered the algorithmic part of the purification U in Fig. 1(a). With the binary tree structure, since each round of purification occurs simultaneously, it is trivial to incorporate a quantum algorithm within the purification structure. As shown,

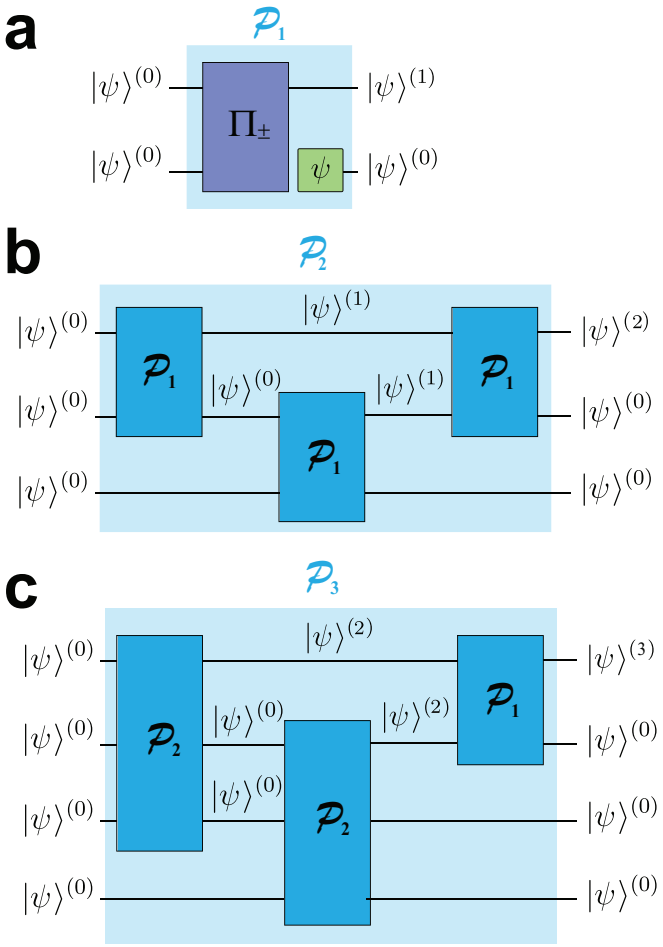


FIG. 2. Qubit-efficient implementation of the PQEC scheme. Purification sequences for (a) $\ell = 1$; (b) $\ell = 2$; (c) $\ell = 3$ are shown. $\mathcal{P}_\ell$ denotes a circuit that consumes $N = 2^\ell$ noisy preparations of $|\psi\rangle^{(0)}$ and outputs a single purified state $|\psi\rangle^{(\ell)}$ that has successfully passed through ℓ rounds of SWAP purification. In (a), the two registers are first SWAP purified according to the circuit given in Fig. 1(b). The discarded register is then reinitialized to the $\ell = 0$ level purified state $|\psi\rangle^{(0)}$.

interleaving the unitaries that compose the quantum algorithm makes parallel updates to the copies as

$$\rho \to \rho' = U_n \rho U_n^\dagger. \tag{28}$$

Since the copies are rotated bilaterally, the state entering each SWAP purification is $\rho' \otimes \rho'$, and the results of the previous section still hold, up to a redefinition of $\rho \to \rho'$. The final resulting state is a purified version of the circuit output state $U\rho U^\dagger$. In this way, we may interleave the quantum algorithm into the purification circuit, in a similar way to QEC.

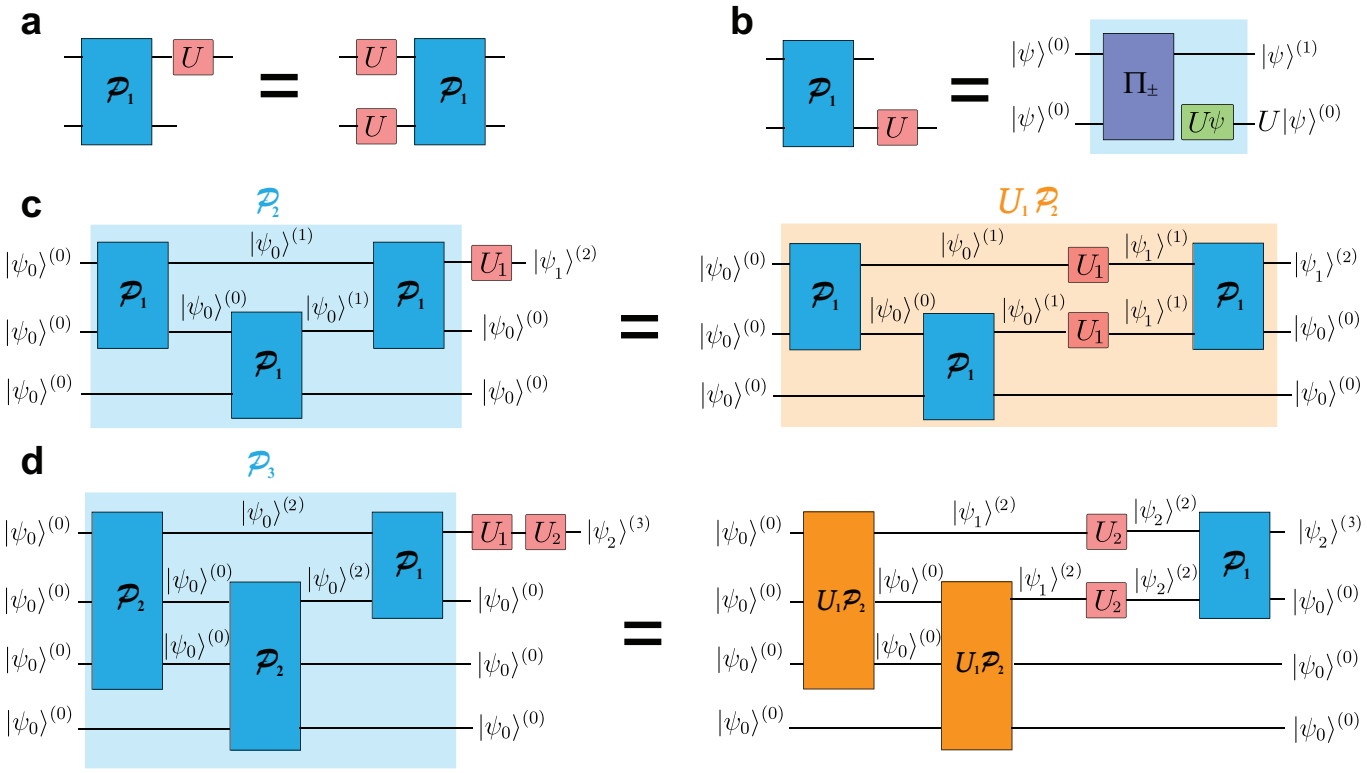


FIG. 3. Interleaving SWAP-based purification within a quantum algorithm. (a) Circuit identity for interchanging the order of the purification step $\mathcal{P}_1$ and an arbitrary unitary U. (b) Absorbing a unitary on the reinitialization register of $\mathcal{P}_1$. (c) Circuit rearrangement for a single unitary gate acting on the $\ell = 2$ purified state. (d) Circuit rearrangement for two unitary gates acting on the $\ell = 3$ purified state. The definition of the block $U_1\mathcal{P}_1$ is shown in (c).

IV. QUBIT-EFFICIENT PQEC

A. Equivalent purification circuit

The arrangement of purification steps shown in Fig. 1(a) is not the only way that PQEC can be implemented. In fact, this arrangement is highly inefficient from the perspective of the number of qubits required. To achieve ℓ rounds of purification, the layer consumes $N = 2^\ell$ noisy instances of the relevant register in parallel and hence uses

$$M\,2^\ell = O(MN) \tag{29}$$

data qubits (plus $O(2^\ell)$ ancillas) during that segment. It is, however, efficient in terms of circuit depth, as the purification steps are parallelized. The depth of the circuit is bounded by the depth of ℓ SWAP-test stages arranged in a binary tree. Figure 1 should be interpreted in this way: the algorithm runs in layers, where each layer is followed by at most ℓ purification rounds. The resulting level-ℓ output states are dispatched forward.

A more qubit-efficient PQEC scheme is shown in Fig. 2. We first only show the purification steps without interleaving the algorithm. The key idea is to reuse quantum registers to reduce the qubit count. Let $\mathcal{P}_\ell$ denote a circuit that consumes $N = 2^\ell$ noisy preparations of $|\psi\rangle^{(0)}$ and outputs a purified state $|\psi\rangle^{(\ell)}$ that has passed through ℓ rounds of the SWAP purification. Since only one of the registers has the purified output, the remaining registers are reused to prepare copies of an unpurified state $|\psi\rangle^{(0)}$. In general, $\mathcal{P}_\ell$ uses $\mathcal{P}_{\ell-1}$ twice to produce two copies of $|\psi\rangle^{(\ell-1)}$. By induction on ℓ, this construction requires only $(\ell{+}1)$ M-qubit registers (plus $O(1)$ ancillas for the SWAP tests) to consume $N = 2^\ell$ noisy initial states and output $|\psi\rangle^{(\ell)}$. At no point are more than $(\ell{+}1)M$ data qubits live, so the coherent data footprint satisfies

$$(\ell{+}1)\,M \;=\; O\big(M \log_2 N\big), \tag{30}$$

rather than (29) for the explicit binary tree configuration. The protocol still realizes the same binary tree of SWAP merges as in the depth–efficient construction, but it stores only one node per depth at any given time; all other intermediate nodes are generated, purified, and recycled on the fly. The trade-off is additional circuit depth: implementing $\mathcal{P}_\ell$ requires two invocations of $\mathcal{P}_{\ell-1}$ plus a final SWAP purification, so the PQEC depth grows linearly in ℓ.

The $O(M \log N)$ qubit scaling is consistent with, and close to, general lower bounds on the resources required for black-box state amplification and purification. Recent optimality results by Grier, Schaeffer, and co-workers [27, 31] show that boosting the overlap with a target state by a

fixed factor, using only coherent access to noisy copies, necessarily incurs costs that grow at least polylogarithmically with the desired accuracy. Our protocol, when implemented as shown in Fig. 2, realizes state amplification with the minimal $O(M \log_2 N)$ number of coherent qubits needed to implement the full set of SWAP merges.

We note that there are also intermediate realizations between the depth-efficient configuration of Fig. 1 and qubit-efficient configuration of Fig. 2. One may spend qubits to keep the per-layer depth small or recycle qubits aggressively at the cost of deeper PQEC segments.

B. Interleaving quantum algorithms

The qubit-efficient version of PQEC may be interleaved in a quantum circuit using the circuit identities shown in Fig. 3(a)(b). The identity shown in Fig. 3(a) follows from the fact that the SWAP projector commutes with a bilateral rotation

$$[\Pi_{\pm}, U \otimes U] = 0. \tag{31}$$

In the $\ell = 1$ purification sequence as shown in Fig. 2(a), one of the registers after purification is discarded (traced out), which is why only one U appears on the right side of the $\mathcal{P}_1$ box in Fig. 3(a). The identity in Fig. 3(b) simply states that, for the lower register of $\mathcal{P}_1$, instead of preparing a new copy of the initial state $|\psi\rangle^{(0)}$, the state $U|\psi\rangle^{(0)}$ is prepared.

The circuit identities are then used in the following way. First start with the bare purification circuit as given in Fig. 2. Then add the unitary sequence for the desired quantum circuit after the purification circuit, such as that shown on the left side of Fig. 3(c)(d). This trivially implements a purified version of the circuit, where the final output is $|\psi_n\rangle^{(\ell)}$, where n is the number of unitary gates in the circuit. One then uses the identities of Fig. 3(a)(b) to put the unitaries throughout the purification circuit. There are many ways that this can be done; a particular example is shown in Fig. 3(c)(d). In a typical QEC scenario, one would like to boost the fidelity after a certain time duration to overcome decoherence, or after particular gates that are known to be noisy.

The simple binary tree layout of Fig. 1(a) can be derived in a similar way. Starting from the pure binary tree purification with the unitary gates placed at the end, one uses Fig. 3(a) to commute the unitaries to the positions as shown in Fig. 1(a).

V. ERROR-MODEL INDEPENDENT PERFORMANCE

The remaining part of this paper will be concerned with analyzing the performance of PQEC under various circumstances. The error model that is considered makes a large difference to the performance of PQEC, and for this reason we will consider various cases separately. In this section, we show results for the general case, and also introduce the concept of an error threshold for PQEC.

A. Noisy single qubit

As a first example of PQEC in action, it is illustrative to consider a single noisy qubit. An arbitrary noisy qubit state can be written

$$\rho = \tfrac{1}{2}(I + \vec{r} \cdot \vec{\sigma}). \tag{32}$$

Substituting this into (7), we obtain the state after a single round of PQEC (see Appendix B)

$$\mathcal{P}(\rho) = \frac{1}{2}\left(I + \frac{2}{1+|\vec{r}|^2}\,\vec{r} \cdot \vec{\sigma}\right). \tag{33}$$

Equivalently, we can consider the coordinates within the Bloch sphere to be updated as

$$\vec{r} \to \frac{2}{1+|\vec{r}|^2}\,\vec{r}. \tag{34}$$

Thus, for qubits, the PQEC step implements a radial rescaling in the Bloch sphere. The Bloch direction is preserved and monotonically increases the radius $|\vec{r}|$ whenever $0 < |\vec{r}| < 1$, with fixed points at $|\vec{r}| = 0$ (maximally mixed) and $|\vec{r}| = 1$ (pure states). From this behavior of PQEC we can expect that it should be particularly effective against depolarizing channels. PQEC acts to revert the effect of such noise, and with sufficient rounds of purification, it acts to restore the state to its fully polarized form.

B. Purification

In general, the purification step (7) acts to increase the purity of a given state. Eq. (7) shows that the map is *spectral*; the eigenbasis is unchanged while only the spectrum is updated. The ordering of eigenvalues is preserved, and their *relative contrast increases*. For a density matrix $\rho = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$ with $\lambda_i > \lambda_j$, the purified state (7) has eigenvalues

$$\frac{\lambda_i'}{\lambda_j'} = \left(\frac{\lambda_i}{\lambda_j}\right)^2 > \frac{\lambda_i}{\lambda_j}. \tag{35}$$

Thus, each purification round increases the purity of the state: it pushes probability weight toward the larger eigenvalues and away from the smaller ones, driving the spectrum away from the maximally mixed point and toward the pure-state vertices. Moreover, the map strictly increases purity, except at the fixed points of completely pure states or maximally mixed states (see Appendix C for details).

C. Fidelity bounds

Consider ρ, a noisy version of $|\psi\rangle\langle\psi|$ produced by a physical noise map applied prior to purification. The fidelity after the PQEC map (7) is

$$F' := \langle\psi|\,\mathcal{P}(\rho)\,|\psi\rangle = \frac{\langle\psi|\,\rho^2\,|\psi\rangle}{\mathrm{Tr}(\rho^2)}, \tag{36}$$

where $\mathrm{Tr}\big(\rho^2\big)$ is the purity.

We can obtain general bounds on the output fidelity by analyzing the term $\langle\psi|\,\rho^2\,|\psi\rangle$. Using the Cauchy–Schwarz inequality for vectors we have

$$F^2 = |\langle\psi|v\rangle|^2 \;\le\; \langle v|v\rangle = \langle\psi|\,\rho^2\,|\psi\rangle\,. \tag{37}$$

where we defined $|v\rangle := \rho\,|\psi\rangle$ and the original fidelity is $F := \langle\psi|\,\rho\,|\psi\rangle$. Another bound may be obtained by noticing that a density operator satisfies $0 \le \rho \le I$, hence $\rho^2 \le \rho$ in the operator (Loewner) order. Taking the expectation value gives

$$\langle\psi|\,\rho^2\,|\psi\rangle \;\le\; \langle\psi|\,\rho\,|\psi\rangle = F. \tag{38}$$

Thus, for every input state ρ, the output fidelity obeys the two-sided sandwich bounds

$$\frac{F^2}{\mathrm{Tr}(\rho^2)} \le F' \le \frac{F}{\mathrm{Tr}(\rho^2)}. \tag{39}$$

Specific noise models and state families sharpen this picture by expressing F, $\mathrm{Tr}\big(\rho^2\big)$, and $\langle\psi|\,\rho^2\,|\psi\rangle$ as explicit functions of the noise parameters.

D. Error threshold

One of the important parameters that characterizes any QEC is the error threshold. The threshold theorem of quantum computing states that for physical error rates that are sufficiently low, it is possible to reduce logical errors rates to arbitrarily low levels, by scaling up the code size. Here, the code size involves creating a version of the code with a larger code distance, and typically involves increasing the number of qubits used to encode the logical states. In this way, low logical errors are obtained by the additional overhead of a larger number of qubits.

In the case of PQEC, the code size is determined by the number of purification rounds ℓ (equivalently, the number of copies $N = 2^\ell$) as a distance-like resource parameter. We may define the notion of an error threshold in the context of PQEC in the following way:

Definition 1 (Error threshold for PQEC)**.** *For a given noise model, we define the purification threshold p_{th} as the largest physical error rate such that, for all $p < p_{\mathrm{th}}$, the logical error rate with PQEC satisfies:*

$$\forall\, p < p_{\mathrm{th}} : \quad \lim_{\ell\to\infty} \gamma_L(\ell, p) = 0 \tag{40}$$

where $\gamma_L(\ell, p)$ is the logical error rate and ℓ denotes the number of purification rounds per PQEC cycle.

To determine the error threshold in practice, we follow the procedure as follows:

0) Initialize the state in a suitable pure state $\rho = |\psi_0\rangle\langle\psi_0|$. Set the cycle number $t = 0$.

1) Apply the error channel

$$\rho \to \mathcal{E}(\rho). \tag{41}$$

2) Perform ℓ rounds of purification

$$\rho \to \mathcal{P}_\ell(\rho). \tag{42}$$

3) Measure the fidelity $F = \langle\psi_0|\rho|\psi_0\rangle$.

4) Update $t \to t + 1$ and go to step 1.

To extract γ_L, we assume that the fidelity F follows an exponential decay with the number of cycles and evaluate:

$$\gamma_L = -\left.\frac{dF}{dt}\right|_{t=0} \approx F(t=0) - F(t=1), \tag{43}$$

where t is the number of cycles. This gives the decay rate of the fidelity with ℓ rounds PQEC, which should be lower than the bare error channel if the error correction is performed successfully. From this we apply (40) to find the error threshold p_{th}.

VI. GLOBAL DEPOLARIZING ERRORS

As seen in Sec. V A, we expect that PQEC is particularly effective under depolarizing channels. In this section, we examine the performance of PQEC under a global depolarizing channel.

A. Error channel

The global depolarizing channel is defined as

$$\mathcal{E}(\rho) = (1-p)\rho + p\frac{I}{D}, \tag{44}$$

where $p \in [0, 1]$ is the error probability. Note that we consider the channel (44) to be applied *per quantum register*. For example, in Fig. 1, the quantum channel acts on each copy of $|\psi_n\rangle^{(\ell)}$.

B. Purification

We first examine how the purification operation (22) acts to improve the fidelity after the channel (44) acts on a initial state $|\psi_0\rangle\langle\psi_0|$. The global depolarizing channel is a particularly simple case to analyze because the channel leaves it a Werner state at all times

$$\rho = (1-\lambda)|\psi_0\rangle\langle\psi_0| + \lambda\frac{I}{D}. \tag{45}$$

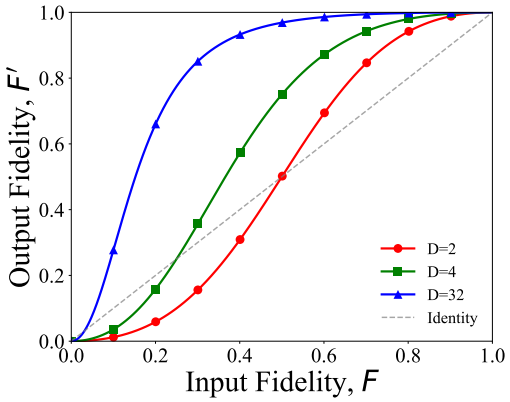


FIG. 4. Output fidelity versus input fidelity after a single pufication round for the Werner channel (44), for several register dimensions D. Curves show the exact map of Eq. (48) with the gray dashed line indicating the identity $F' = F$. The purification step strictly improves fidelity when $F > \frac{1}{D}$, and the gain is largest at lower F where the curvature is steepest.

The PQEC purification also leaves the state in Werner form since (22) involves a mixture of ρ and ρ^2, and the term ρ^2 for a Werner state is itself a Werner state. The fidelity of the Werner state (45) is

$$F = \langle\psi_0|\rho|\psi_0\rangle = 1 - \lambda(1 - \frac{1}{D}). \tag{46}$$

Substituting (45) into (6), we obtain the effect of applying one round of purification

$$\lambda \mapsto \lambda' = \frac{\lambda^2}{D(1-\lambda)^2 - \lambda(\lambda - 2)}. \tag{47}$$

This can be equally written in terms of the fidelity using the relation (46), giving

$$F \mapsto F' = \frac{F^2}{F^2 + \frac{(1-F)^2}{D-1}}. \tag{48}$$

In Fig. 4, we show the output fidelity for various system sizes. By solving for the crossing point $F' = F$, we find that the fidelity improves in the region $F > 1/D$. In terms of the Werner mixing parameter, this corresponds to $\lambda < 1$. For large registers $D \gg 1$, the purification becomes extremely effective and quickly approaches $F' \approx 1$ after a single round. We show in Appendix D that if $F > 1/D$, iterating (48) is strictly increasing and $F_* = 1$ is an attractive fixed point.

The above shows that by repeated purification rounds, it is always possible to fully recover any Werner state to the ideal state $|\psi_0\rangle$ as long as $\lambda < 1$. Thus any Werner state that is not a completely mixed state can be used to recover unit fidelity.

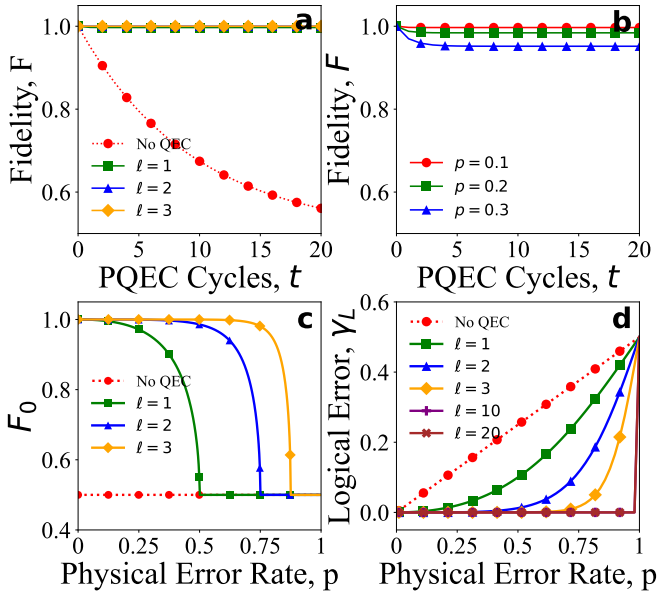


FIG. 5. Error threshold behavior for a single qubit ($D = 2$) in a global depolarizing channel with PQEC. (a)(b) Fidelity evolution under the depolarizing Bloch vector contraction (44) and ℓ rounds of purification (34), where t is the number of cycles. (a) Fidelity evolution versus t with $p = 0.1$ and various ℓ. (b) Fidelity evolution versus t for $\ell = 1$ and various p as shown. (c) The steady state ($t \to \infty$) fidelity F_0 for various ℓ. (d) The logical error rate γ_L evaluated using (43) for various ℓ.

C. Error thresholds

To determine the error threshold under the global depolarizing channel, we perform the procedure given in Sec. V D. Under the global depolarizing channel, the Werner state parameter undergoes the update rule

$$\lambda \mapsto (1-p)\lambda + p. \tag{49}$$

PQEC then updates the Bloch vector according to (47). The fidelity is then evaluated using (46).

A typical sequence is shown in Fig. 5(a)(b). We see that the fidelity follows an exponential decay due to the depolarizing channel being applied per cycle. Without any application of PQEC ($\ell = 0$), the fidelity approaches the completely mixed state. As expected, adding additional rounds of purification reduces the error rate. Interestingly, the fidelities converge to higher values in the limit of a large number of cycles, improving with ℓ. For example, in the case of $\ell = 1$, the steady state fidelity may be found by determining the fixed point of the combined effect of (47) and (49), giving

$$\begin{aligned} F_0 &:= \lim_{t\to\infty} F(t) \\ &= \frac{1}{2}\left(1 + \sqrt{1 - \frac{4(D-1)p^2}{D^2(1-p)^2}}\right). \end{aligned} \tag{50}$$

The steady-state fidelities as a function of p is shown in Fig. 5(c). In a typical QEC, for a large number of cycles, the fidelities approach that of a completely mixed state $F_0 = 1/D$. The saturation effect acts to further preserve the fidelity in addition to suppressing the logical decay rate.

Figure 5(d) shows the logical error rate γ_L as a function of physical error rate p for various numbers of rounds ℓ of PQEC. We see that, as expected, PQEC reduces the logical error rate with ℓ. With $\ell = 20$, the logical error rate is reduced to nearly zero for the full range of $p \in [0, 1]$. The error threshold is found by determining the crossing point of the curves in Fig. 5(d), which in this case is

$$p_{\text{th}} = 1, \tag{51}$$

for the global depolarizing channel. This means that any error rate $p < 1$ can be reduced to a logical error rate of zero, assuming that a sufficient number of copies are available such that PQEC can be scaled up sufficiently.

Similar results to Fig. 5 are obtained for higher dimensions D. The error threshold is found to be the same as (51) for any dimension.

VII. LOCAL DEPOLARIZING ERRORS

The global depolarizing channel analyzed in the previous section has the advantage of computational simplicity, due to the state being of Werner form at all times. It is, however, not the most realistic error model, since it is equivalent to a sum of all possible Clifford errors, which involves multiqubit bit/phase flip errors. In order to examine a more realistic case, we now examine local depolarizing errors. Here, each qubit is subjected to an identically and independently distributed single-qubit depolarizing channel.

A. Error channel

The local depolarizing channel on the mth qubit is defined as

$$\mathcal{E}_m(\rho) = E_m^{(0)} \rho E_m^{(0)\dagger} + \sum_{j=x,y,z} E_m^{(j)} \rho E_m^{(j)\dagger}, \tag{52}$$

where the Kraus operators are defined as

$$\begin{aligned} E_m^{(0)} &= \sqrt{1-p} I_m \\ E_m^{(j)} &= \sqrt{\tfrac{p}{3}} \sigma_m^{(j)} \end{aligned} \tag{53}$$

and $p \in [0, 1]$ is the error probability. The full register noise is then the product channel

$$\mathcal{E}(\rho) = \bigotimes_{m=1}^{M} \mathcal{E}_m(\rho). \tag{54}$$

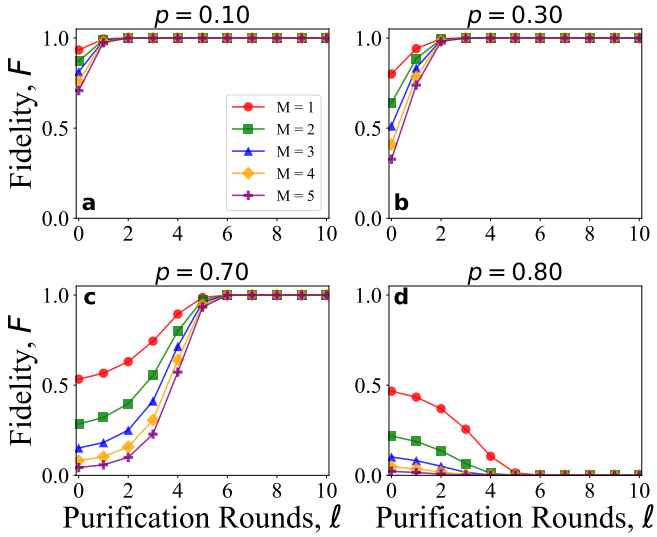


FIG. 6. Fidelity improvement of locally depolarized states under successive SWAP purification rounds. All plots show fidelity, F, versus rounds of SWAP purification, ℓ, for various qubit numbers M. The initial state is the state $|\psi_0\rangle = |+\rangle^{\otimes M}$ subjected to the channel (52) with the error probability (a) $p = 0.1$; (b) $p = 0.3$; (c) $p = 0.7$; (d) $p = 0.8$.

B. Purification

1. *Symmetric product states*

We first show the effect of PQEC applied to the M-qubit symmetric product state, subjected to local depolarization. For concreteness, we choose $|\psi_0\rangle = |+\rangle^{\otimes M}$; however, due to the isotropic nature of the local depolarizing channel, the same results hold for any completely symmetric product state $|\phi\rangle^{\otimes M}$. While local depolarization preserves the separable nature of the target state, the purification step (7) does not. For this reason, we resort to numerical and approximate methods to analyze different system sizes M.

Figure 6 shows the fidelity after ℓ purification rounds and various physical error probabilities p. The local depolarization noise model is applied only to the initial state, and successive purification rounds are applied to the noisy state. We see that, for $p < 3/4$, the fidelity converges to 1 for all system sizes. Larger system sizes tend to require more purification rounds to reach the same fidelity, but this may be attributed to starting at a lower fidelity. For moderate p (e.g. $p \lesssim 0.5$), only a few rounds are needed to reach $F_\ell \gtrsim 0.99$. As p approaches $3/4$ from below, more rounds are required. Once the error probability is larger than $p \geq 3/4$, the fidelities worsen with ℓ.

The critical value of p may be explained in the following way. The initial state, $|+\rangle^{\otimes M}$, after applying (54) becomes

$$\mathcal{E}(|+\rangle\langle+|^{\otimes M}) = \left((1 - \frac{4p}{3})|+\rangle\langle+| + \frac{4p}{3}\frac{I}{2}\right)^{\otimes M}. \tag{55}$$

For an error probability of $p = 3/4$, the state is a completely mixed state. As may be seen from (7), a completely mixed state $\rho \propto I$ is unaffected by purification, and hence is not recoverable. Similarly to the global depolarizing channel, any state with a higher fidelity than the completely mixed state can be recovered with sufficient rounds of purification. For $p > 3/4$, PQEC acts to actively take the state away from the original state towards another state. For example, for $p = 1$ the state (55) is $(\frac{1}{3}|+\rangle\langle+| + \frac{2}{3}|-\rangle\langle-|)^{\otimes M}$, and the convergence is towards the $|-\rangle^{\otimes M}$ state.

We may see the critical value of p more clearly by plotting the fidelities versus p as shown in Fig. 7(a)(b). For $p < 3/4$, each purification level increases fidelity, and as ℓ grows, the fidelity can be made arbitrarily close to 1. The critical value remains $p = 3/4$ for all values of M.

2. *Small error expansion*

The results of the previous section was limited to permutationally symmetric product states. In order to see the performance of PQEC on more general states, we perform a small error expansion under the local depolarizing channel.

For an arbitrary pure state $|\psi_0\rangle$, we may expand the density matrix in the M-qubit Pauli basis as

$$|\psi_0\rangle\langle\psi_0| = \frac{1}{2^M}\sum_P r_P P \tag{56}$$

where the sum runs over 4^M Pauli strings and the real coefficients $r_P := \langle\psi|P|\psi\rangle$ satisfy $\sum_P r_P^2 = 2^M$. Each Pauli string has a weight $w(P)$, which is equal to the number of non-identity Pauli operators. The coefficients, r_P, may then be grouped by weight, so that we can define

$$a_k = \frac{1}{2^M}\sum_{P:\, w(P)=k} r_P^2, \tag{57}$$

By performing a small-p expansion, we evaluate that the fidelity of a general state decays under the local depolarizing channel as (see Appendix E)

$$F(\rho) = 1 - \frac{4}{3}\bar{k}\,p + O(p^2) \tag{58}$$

where $\bar{k} := \sum_k k a_k$. Under one round of PQEC with purification (7), the output fidelity is

$$F(\mathcal{P}(\rho)) = \frac{\langle\psi_0|\rho^2|\psi_0\rangle}{\mathrm{Tr}(\rho^2)} = 1 - O(p^2). \tag{59}$$

Thus, for a single purification round, the leading logical error $O(p)$ is removed entirely. This shows the effectiveness of the PQEC against local depolarizing errors for the general case in the small-p regime.

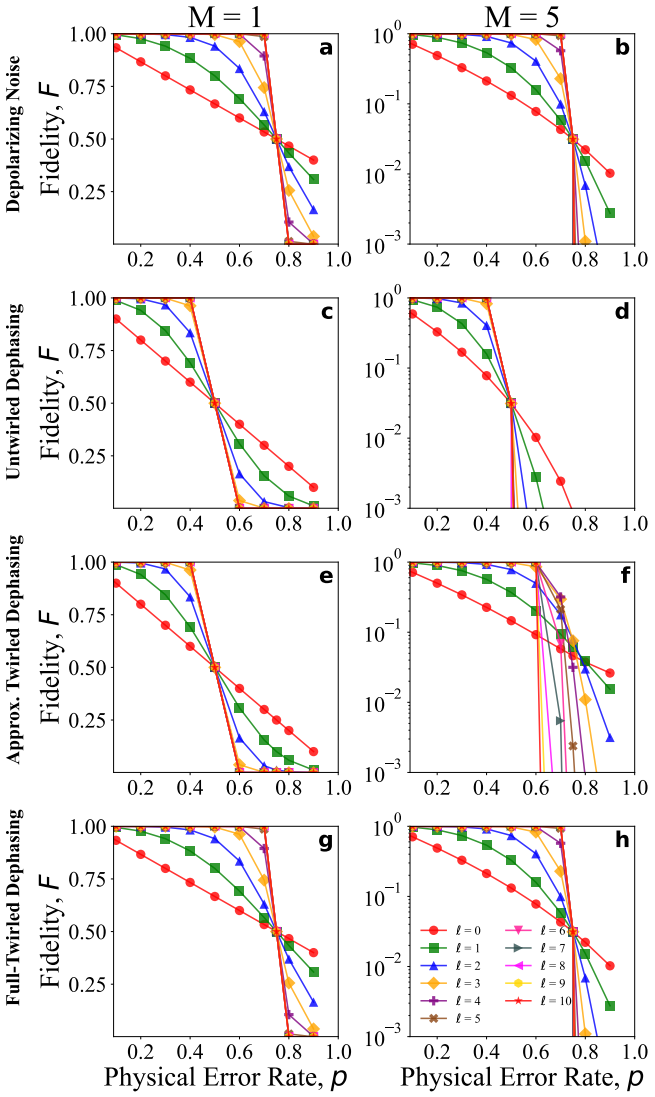


FIG. 7. Fidelity evolution of various noise models under one cycle of PQEC ($t = 1$). All plots show fidelity, F, versus iterations of physical error rate, p, for various numbers of purification rounds, ℓ. The initial state is $|\psi_0\rangle = |+\rangle^{\otimes M}$; The columns represent $M = 1$ (left) and $M = 5$ (right). The top row (a)(b) is subjected to the local depolarization channel (52); the second row (c)(d) is subjected to local dephasing without twirling (61); the third row (e)(f) is subjected to local dephasing with 20% twirling (65), where the applied twirling gates are randomly selected from the total set of 3^M options; the fourth row (g)(h) is subjected to local dephasing with full twirling (65). The protocol exhibits ultrahigh error tolerance for depolarization noise, with effective correction persisting up to $p_{\max} \sim 0.75$. The fully-twirled dephasing matches the depolarization result as the channel is effectively isotropic after twirling.

C. Error thresholds

To analyze the error threshold behavior, we again follow the procedure in Sec. V D using the local depolarizing

channel (54) starting from the initial state $|\psi_0\rangle = |+\rangle^{\otimes M}$.

Figure 8(a)(b) shows the fidelity evolution as a function of error and purification cycles for $M = 5$. Similarly to the global depolarizing channel (see Fig. 5), the fidelity follows an exponential decay that saturates to a finite value for the physical error values shown. For the physical error probabilities in Figs. 8(a)(b), increasing the code size reduces the logical error rate (43).

Figure 8(c)(d) shows the logical error rate γ_L as a function of the physical error rate p. For both $M = 1$ and $M = 5$, we observe a crossover of the curves, giving the error threshold at

$$p_{\text{th}} = 3/4. \tag{60}$$

For $p < p_{\text{th}}$, increasing the number of purification rounds ℓ substantially suppresses the logical error rate γ_L, reflecting that each PQEC cycle yields a net purification gain that slows logical-error accumulation over time. As $p \to p_{\text{th}}$, the benefit of increasing ℓ diminishes, and progressively larger ℓ is required to maintain a small effective decay rate. For $p > p_{\text{th}}$, adding additional purification rounds produces a worse result, characteristic of the error rate being over the threshold. We note the same threshold was attained for the case of a non-stabilizer code [32].

We note that the $M = 1$ case is physically the same as the global depolarizing channel for a single qubit $D = 2$ The difference of Fig. 8(c) to Fig. 5(d) arises only due to the difference in definition of the error channel. For the global depolarizing channel, $p = 1$ corresponds to a completely mixed state, whereas for the local depolarizing channel, $p = 3/4$ is the corresponding point. The $M = 5$ case, shown in Fig. 8(d), is distinct from the global depolarizing channel, since only local depolarization is applied.

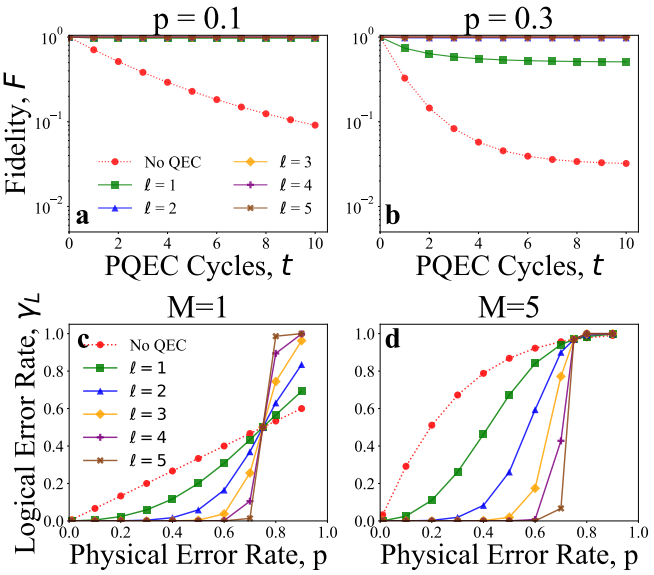


FIG. 8. (Top) Fidelity evolution of locally depolarized states under PQEC. Plots show fidelity, F, versus cycles of PQEC purification, t, for various numbers of purification rounds, ℓ. The initial state is the state $|\psi_0\rangle = |+\rangle^{\otimes M}$ with $M = 5$, subjected to the channel (52) with the error probability (a) $p = 0.1$; (b) $p = 0.3$. (Bottom) Logical error rate γ_L, versus physical error rate p, for local depolarizing noise (52) on the target state $|+\rangle^{\otimes M}$. Curves are shown for (c) $M = 1$ and (d) $M = 5$ for various purification rounds, ℓ. The no correction threshold corresponds to $\ell = 0$.

VIII. LOCAL ANISOTROPIC ERRORS

In this section, we consider local anisotropic errors under PQEC to analyze its performance. We show how Clifford twirling can be used to convert it into an effective depolarizing channel that is more naturally handled by our protocol.

A. Error channel

As a canonical example of an anitropic error, we consider a dephasing channel which can be defined on the mth qubit as

$$\mathcal{E}_m^{(z)}(\rho) = E_m^{(0)}\rho E_m^{(0)\dagger} + E_m^{(1)}\rho, E_m^{(1)\dagger} \tag{61}$$

where the Kraus operators are

$$\begin{aligned} E_m^{(0)} &= \sqrt{1-p}I_m \\ E_m^{(1)} &= \sqrt{p}Z_m. \end{aligned} \tag{62}$$

On an M-qubit register, we take the noise to be

$$\mathcal{E}_z(\rho) = \bigotimes_{m=1}^{M} \mathcal{E}_m^{(z)}(\rho), \tag{63}$$

which we will refer to as a local dephasing model.

B. Conversion of dephasing to depolarizing errors

As discussed in Sec. V A, we expect PQEC to be most effective with depolarizing errors, since the purification acts to repolarize the state. The problem of anisotropic errors for PQEC may be illustrated with a single qubit ($M = 1$). Using the parameterization (32), the effect of applying the dephasing channel (63) is to leave the z component invariant while contracting the transverse components

$$\vec{r} \longmapsto ((1-2p)r_x, (1-2p)r_y, r_z). \tag{64}$$

This results in a change in the direction of the Bloch vector (see Fig. 9). Applying the purification step (34) acts to repolarize the vector. However, since the purification does not change the direction of the Bloch vector, the correction step does not revert the state to the original vector $\vec{r}$ (see Appendix F for more details). This results

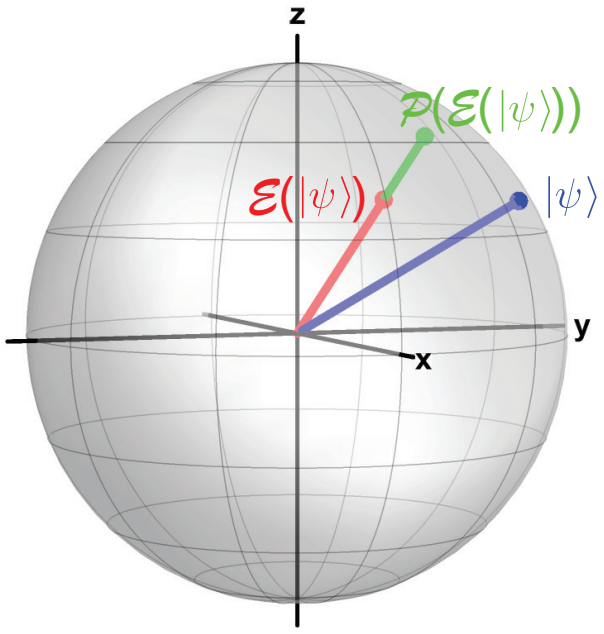


FIG. 9. A single PQEC cycle for a single-qubit state under dephasing. The initial state $|\psi\rangle$ has Bloch sphere parameters $\theta = \pi/3, \phi = \pi/4$. The dephased state $\mathcal{E}(|\psi\rangle)$ is applied according to (63) with $p = 0.3$. Finally one round of purification according to (34) is applied to give $\mathcal{P}(\mathcal{E}(|\psi\rangle))$. The purified state lies in the same direction as $\mathcal{E}(|\psi\rangle)$.

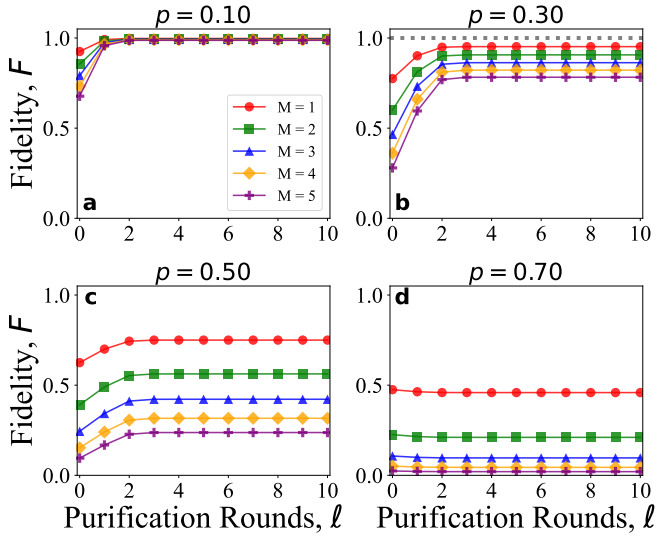


FIG. 10. Fidelity improvement of locally dephased states under PQEC, without twirling. All plots show fidelity, F, versus rounds of purification, ℓ, for various qubit numbers M. The initial state is $|\psi_0\rangle = (\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle)^{\otimes M}$, with parameters $\theta = \pi/3, \phi = \pi/4$, and is subjected to the channel (61) with the error probability (a) $p = 0.1$; (b) $p = 0.3$; (c) $p = 0.5$; (d) $p = 0.7$.

in an imperfect correction operation, and results in a saturation of the fidelity for PQEC.

We now show how Clifford "frame randomization" rotations turns axis-biased dephasing into an effective depolarizing channel, without changing the PQEC map. We focus on local Z-dephasing (63), but our results apply to any anisotropic local error channel. The dephasing channel may be converted to a depolarizing channel by averaging over the X, Y, Z-directions, by performing a basis transformation

$$\mathcal{E}_m(\rho) = \frac{1}{3}\sum_{U\in\{I,H,HS\}} U^\dagger \mathcal{E}_m^{(z)}(U\rho U^\dagger)U. \tag{65}$$

where H is the Hadamard operator and $S = \mathrm{diag}(1, i)$ is the phase operator acting on the mth qubit. Each U permutes the Bloch axes, so averaging over the three rotations maps the local Z dephasing to a local depolarizing channel. On M qubits, substituting (65) into (54) gives the full local depolarizing channel. This would give 3^M different types of unitary operations to average over, corresponding to all combinations of $\{I, H, HS\}^{\otimes M}$.

In practice, the twirling process can be implemented by applying any one of the 3^M unitaries randomly before and after the dephasing errors take place. Another approach would be to split the time that the error channel is exposed into 3^M intervals, and within each interval apply the corresponding unitary. This converts the dephasing channel to a depolarizing channel such that the results of Sec. VII are recovered exactly. However, it has the drawback of averaging over a large number (i.e., $O(3^M)$) of samples to obtain good statistics. This makes the full twirling process highly costly from both a real implementation and numerical point of view.

More economically, a subset of the full set of 3^M gates can be applied for approximate twirling, while still improving results beyond the no-twirl baseline (see Sec. VIII D). For this reason, we also examine approximate twirling protocols to convert local dephasing errors to a more isotropic form that has better performance with PQEC. We consider the approximate twirling channel

$$\mathcal{E}_\mathcal{T}(\rho) = \frac{1}{T}\sum_{U\in\mathcal{T}} U^\dagger \mathcal{E}_z(U\rho U^\dagger)U. \tag{66}$$

where $T = \dim\mathcal{T}$ and $\mathcal{T}$ is a restricted set of Clifford unitaries to perform twirling over.

C. Purification

Figure 10 shows the fidelity evolution under multiple rounds of purification for the state $|\psi_0\rangle = (\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle)^{\otimes M}$, for various M and physical dephasing error probabilities p. No twirling is performed. We see that as before, below a critical error probability, the purification acts to improve the fidelity. The primary difference to the depolarizing case is that the fidelities saturate to a value less than $F = 1$. This is precisely the effect discussed in Fig. 9, where the purification does

not completely revert the state. However, for small error probabilities the fidelities nevertheless converge to values close to unity, showing that the purification can still be effective, while not perfect. For error probabilities above $p > 1/2$, we do not expect good performance, since at $p = 1/2$, the x and y components of the Bloch vector (64) are completely erased. We see in Fig. 10(c) the fidelities do in fact increase because the purification converges to $|0\rangle^{\otimes M}$, which is closer to the original state than the initially dephased state.

To overcome the saturation effect, we apply the twirling operation of (65) to each qubit by averaging over 3^M unitary operations. According to the discussion in Sec. VIII B, this converts the local dephasing channel to a local depolarizing channel with the same parameter p. This prediction is borne out numerically. Performing the full twirling with the dephasing channel reproduces precisely the same graphs as shown in the genuine depolarizing channel Fig. 6 (we do not show the twirled plots as they are visually identical). Interpreting Fig. 6 as the results for twirled dephasing and comparing to Fig. 10, we see that the twirling generally acts to improve the fidelity in a wider range $p < 3/4$. For small errors (e.g. $p = 0.1$) there is only a marginal improvement but for larger errors the increase in fidelity is more marked.

Figure 7(c)-(f) shows the shift in the critical error probability with and without twirling under the dephasing channel. The plots show the fidelity F versus p for various ℓ and system sizes M. With the initial state set at $|+\rangle^{\otimes M}$, and applying dephasing without twirling, additional rounds of purification act to improve the fidelity only if $p < 1/2$. When twirling is added, the critical point shifts to $p = 3/4$. The critical point is unchanged for different M. However, for larger M, the fidelity for the same number of purification rounds is reduced. We may attribute this to the fact that for the same error rate p, a larger system tends to have a higher probability that there is an error. For example, the probability that there is no error in the whole system under the channel (63) is $(1-p)^M$, which reduces exponentially with M.

D. Error thresholds

We now consider the error threshold analysis by following the procedure in Sec. V D using the local dephasing channel (63) starting from the initial state $|\psi_0\rangle = |+\rangle^{\otimes M}$.

Figure 11 shows results for local dephasing without twirling for $M = 1$ and $M = 5$. Figure 11(a)(b) shows the fidelity evolution as a function of error and purification cycles for $M = 5$. Similarly to the global and local depolarizing channels (see Figs. 5 and 8), the fidelity follows an exponential decay that saturates to a finite value for the physical error values shown. As before, increasing the code size reduces the logical error rate (43).

Figure 11(c)(d) shows the logical error rate γ_L as a function of the physical error rate p. For both $M = 1$ and

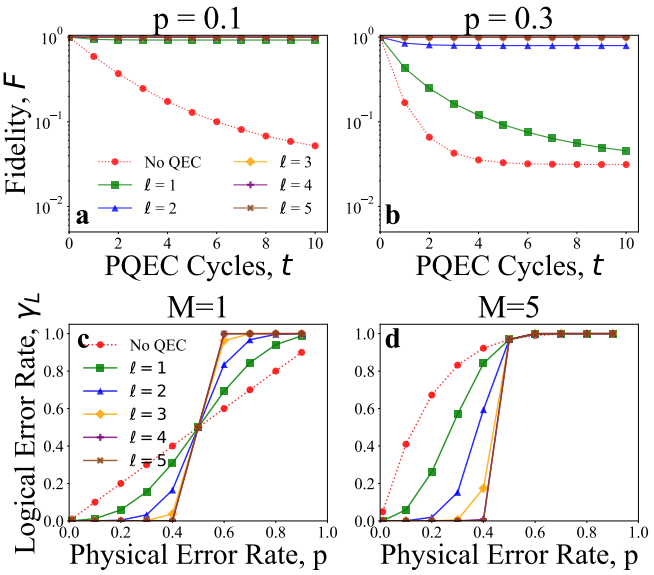


FIG. 11. (Top) Fidelity improvement of locally dephased states with no twirling under PQEC. Plots show fidelity, F, versus iterations of PQEC purification, t, for various numbers of purification rounds, ℓ. The initial state is the state $|\psi_0\rangle = |+\rangle^{\otimes M}$ with $M = 5$, subjected to the channel (63) with the error probability (a) $p = 0.1$; (b) $p = 0.3$. (Bottom) Logical error rate, γ_L, versus physical error rate, p, for pure dephasing noise without twirling on the target state $|+\rangle^{\otimes M}$ with noise channel (63). Curves are shown for $M = 1$ (c) and $M = 5$ (d) for various purification rounds, ℓ. The no correction threshold corresponds to $\ell = 0$. The threshold is seen at $p_{\rm th} = 0.5$.

$M = 5$, we observe a crossover of the curves, giving the error threshold at

$$p_{\rm th} = 1/2. \tag{67}$$

For $p < p_{\rm th}$, increasing the number of purification rounds ℓ substantially suppresses the logical error rate γ_L, reflecting that each PQEC cycle yields a net purification gain that slows logical error accumulation over time. As $p \to p_{\rm th}$, the marginal benefit of increasing ℓ diminishes, and progressively larger ℓ is required to maintain a small effective decay rate. For $p > p_{\rm th}$, adding additional purification rounds produces a worse result, characteristic of the error rate being over the threshold.

Because the full twirling (65) maps local dephasing to an isotropic Pauli channel, the resulting PQEC dynamics, and subsequently all threshold curves for dephasing with complete twirling, are exactly identical to those obtained in the depolarization analysis of Fig. 8 for every (p, ℓ), the fidelity relaxation to its steady state, the steady-state values themselves, and the associated logical-error behavior (with threshold at $p = 0.75$), coincide point-wise with the depolarizing case.

We also consider approximate twirling, where only a fraction of the full 3^M Clifford gates is applied. Figure 12 shows results for local dephasing with only 20% twirling applied for $M = 1$ and $M = 5$ (rounded up to a whole

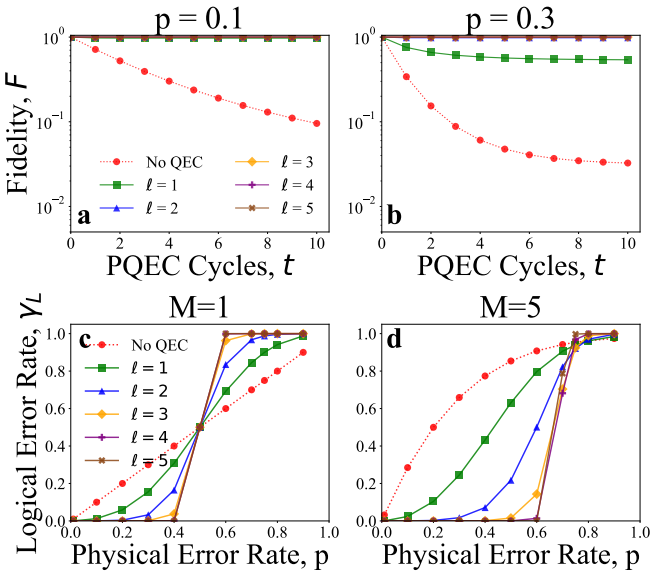


FIG. 12. (Top) Fidelity improvement of locally dephased states with approximate twirling under PQEC. Plots show fidelity, F, versus iterations of PQEC purification, t, for various numbers of purification rounds, ℓ. The initial state is the state $|\psi_0\rangle = |+\rangle^{\otimes M}$ with $M = 5$, subjected to the channel (65) with the error probability (a) $p = 0.1$; (b) $p = 0.3$. (Bottom) Logical error rate, γ_L, versus physical error rate, p, for approximately twirled-dephasing noise on the target state $|+\rangle^{\otimes M}$ with noise channel (65). Curves are shown for $M = 1$ (c) and $M = 5$ (d) for various purification rounds, ℓ. The no correction threshold corresponds to $\ell = 0$. Approximate twirling corresponds to using 20% of the total 3^M total gate combinations, rounded up to a whole number.

number of gates); the applied twirling gates are randomly selected from the total set of 3^M options. Figure 12(a)(b) shows the fidelity evolution as a function of error and purification cycles for $M = 5$. Similarly to the global and local depolarizing channels (see Figs. 5 and 8), the fidelity follows an exponential decay that saturates to a finite value for the physical error values shown. As before, increasing the code size reduces the logical error rate (43).

Figure 12(c)(d) shows the logical error rate γ_L as a function of the physical error rate p. For $M = 1$, approximate twirling here does not change anything from the untwirled case, as it corresponds to only applying a single Clifford gate; this simply rotates the dephasing axis but does not make the noise channel any less anisotropic. Hence, the crossover is $p_{\text{th}} = 0.5$, identical to pure dephasing without twirling as shown in Fig 11(c). For $M = 5$, the crossover yields an approximate error threshold $0.7 < p_{\text{th}} \lesssim 0.8$. For $p < p_{\text{th}}$, increasing the number of purification rounds ℓ substantially suppresses the logical error rate γ_L, reflecting that each PQEC cycle yields a net purification gain that slows logical error accumulation over time. As $p \to p_{\text{th}}$, the marginal benefit of increasing ℓ diminishes, and progressively larger ℓ is required to maintain a small effective decay rate. For $p > p_{\text{th}}$, adding additional purification rounds produces a worse result, characteristic of the error rate being over the threshold.

We see that approximate twirling is effective at increasing the error threshold beyond the no-twirl crossover at $p_{\text{th}} = 0.5$, with full twirling achieving $p_{\text{th}} = 0.75$, consistent with the isotropic (depolarization) case. One can accordingly tune the number of twirling gates applied, in order to negotiate the threshold with the resource demands of applying more gates.

In both noise families, the qualitative picture is therefore the same: below the relevant threshold, PQEC can in principle drive the single-layer fidelity arbitrarily close to unity by increasing ℓ (in the absence of additional noise within the PQEC layer). However, above the threshold, the protocol cannot asymptotically recover the target state, even with additional purification rounds.

IX. SUMMARY AND CONCLUSIONS

We have introduced a scheme for performing quantum error correction based on the SWAP-based purification primitive, which we refer to as purification quantum error correction (PQEC). Operationally, PQEC uses physical redundancy in the *copy* domain rather than encoding redundancy in codewords. Given N identically prepared noisy copies of a state, it repeatedly compares and merges pairs, keeping the symmetric outcomes and recursively purifying the surviving states. What is attained is the same as quantum error correction: it reduces the error of an *unknown* principal density matrix eigenstate by consuming additional noisy copies, and it does so via shallow circuits and classical feedforward. The fact that the purification works on unknown states is crucial for QEC tasks, because in a quantum circuit the mid-circuit state is not necessarily known. Furthermore, there is no postselection employed in our scheme and all measurement outcomes of the SWAP test are used to access the purified component of the conditional state that is generated. The main distinction from standard QEC is where the redundancy resides (copies vs codewords) and the atypical extraction procedure to obtain the purified state.

The most natural type of errors that are handled by PQEC are depolarizing errors, where the error threshold was evaluated to be 75% for local depolarizing errors. While dephasing errors are expected not to work as well due to the anisotropy (Fig. 9), nevertheless the difference in performance for small error probabilities is not drastically different between depolarizing and dephasing errors. Furthermore, dephasing errors can be converted to a depolarizing channel by twirling giving identical results for full local twirling. Compared to the error thresholds of standard QEC codes (e.g. $\sim 1\%$ for surface codes [33]), PQEC error thresholds are remarkably high. This can be attributed to the different mechanism of correction in PQEC. The quoted thresholds all correspond to physical error probabilities, where the error channel converts an arbitrary state into a completely mixed state. Hence for PQEC, the main important factor is that the copies of

the state are not entirely destroyed by the noise. Given that the input states have some level of coherence and a sufficient number of copies is available, it is always possible to recover the original state through successive rounds of PQEC.

Let us compare the qubit resources of PQEC with a standard QEC. Using the simplest binary tree implementation of PQEC as shown in Fig. 1(a), the total qubit count is $n \sim O(MN)$, where $N = 2^\ell$ is the number of copies. On first glance this would appear to have a heavy qubit count, but errors are also suppressed exponentially with N, due to (22). Then the logical error rate scales as $\gamma_L \sim e^{-N} \sim e^{-n/M}$. This is comparable with the scaling of standard QEC, where there is exponential scaling of logical errors with the code distance. For example, in the surface code, the number of physical qubits scales as the square of the code distance so that $\gamma_L \sim e^{-\sqrt{n}}$. The scaling in terms of the number of qubits can be even improved further using the qubit-efficient PQEC formulation of Sec. IV, although this comes at the cost of deeper circuits.

In addition to the extremely high error thresholds, one major advantage that PQEC has over standard QECs is that fault-tolerant gates are extremely simple to implement, since the encoding is merely in the form of copies. In standard QEC, for fault-tolerance one requires transversal logical gates, where the gates on the encoded states are arranged in a way such that errors do not propagate. According to the Eastin-Knill theorem, a universal set of transversal gates (i.e. including non-Clifford gates) is incompatible with the ability of a code to detect arbitrary errors [34]. In our case, since the redundancy is simply present in the form of a product state, the gates that are applied to each copy are also in product form. This means that errors that occur on a particular copy register do not propagate between the copies. The copies do not meet until the purification process itself, where errors are removed (Fig. 1). Thus, in PQEC, performing universal logical gates in a fault-tolerant way — such that the errors do not multiply throughout the quantum computer — is far less problematic.

In this paper, we have mainly considered the quantum computing scenario, where one is interested in the output of a fixed quantum algorithm. In addition to the mid-circuit correction that we have discussed, it may also be used to prepare high fidelity initial states (e.g., as a Bell-pair or magic state factory). There are other applications where PQEC can also be useful. Due to the logarithmic coherent-memory requirement, it is naturally suited to settings where identical or nearly identical copies are produced over time. Examples include quantum sensor networks [35], distributed quantum computing [22, 24], quantum metrology with spin ensembles [32, 36, 37], and quantum communication and repeater architectures [16, 21]. In such regimes, PQEC can run in the background as an error-correction layer, continuously distilling higher-fidelity states from noisy input streams while keeping the coherent footprint small.

ACKNOWLEDGMENTS

This work is supported by the SMEC Scientific Research Innovation Project (2023ZKZD55); the National Natural Science Foundation of China (92576102); the Science and Technology Commission of Shanghai Municipality (22ZR1444600); the NYU Shanghai Boost Fund; the China Foreign Experts Program (G2021013002L); the NYU-ECNU Institute of Physics at NYU Shanghai; the NYU Shanghai Major-Grants Seed Fund; and Tamkeen under the NYU Abu Dhabi Research Institute grant CG008.

Data and Code Availability. All data supporting the findings of this study are available from the corresponding author upon reasonable request.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
[2] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Reports on Progress in Physics **76**, 076001 (2013).
[3] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).
[4] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, Physical Review A **97**, 10.1103/physreva.97.032306 (2018).
[5] D. Gottesman, Stabilizer codes and quantum error correction (1997), arXiv:quant-ph/9705052 [quant-ph].
[6] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Physical Review A **86**, 10.1103/physreva.86.032324 (2012).
[7] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Physical Review Letters **106**, 10.1103/physrevlett.106.180504 (2011).
[8] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Physical Review A **94**, 10.1103/physreva.94.052325 (2016).
[9] D. Lidar, *Quantum Information and Computation for Chemistry* (Wiley, 2014).
[10] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Physical Review A **80**, 10.1103/physreva.80.012304 (2009).
[11] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999).
[12] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).
[13] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Reviews of Modern Physics **95**,

10.1103/revmodphys.95.045005 (2023).
[14] W. Dür and H. J. Briegel, Entanglement purification and quantum error correction, Reports on Progress in Physics **70**, 1381–1424 (2007).
[15] P.-S. Yan, L. Zhou, W. Zhong, and Y.-B. Sheng, Advances in quantum entanglement purification, Science China Physics, Mechanics & Astronomy **66**, 250301 (2023).
[16] 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, Physical Review Letters **76**, 722–725 (1996).
[17] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Quantum privacy amplification and the security of quantum cryptography over noisy channels, Phys. Rev. Lett. **77**, 2818 (1996).
[18] M. Murao, M. B. Plenio, S. Popescu, V. Vedral, and P. L. Knight, Multiparticle entanglement purification protocols, Phys. Rev. A **57**, R4075 (1998).
[19] S. Ahmad, S. Li, J. Raghoonanan, K. Zhou, V. Ivannikov, and T. Byrnes, Distillation of supersinglet states (2025), arXiv:2509.20962 [quant-ph].
[20] J.-W. Pan, C. Simon, C. Brukner, and A. Zeilinger, Entanglement purification for quantum communication, Nature **410**, 1067–1070 (2001).
[21] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, Quantum repeaters: The role of imperfect local operations in quantum communication, Phys. Rev. Lett. **81**, 5932 (1998).
[22] H. J. Kimble, The quantum internet, Nature **453**, 1023–1030 (2008).
[23] J. Mauricio Torres, J. Zsolt Bernád, and R. Gómez-Rosas, Performance of entanglement purification including maximally entangled mixed states, Journal of Physics A: Mathematical and Theoretical **57**, 315302 (2024).
[24] J. Illiano, M. Caleffi, A. Manzalini, and A. S. Cacciapuoti, Quantum internet protocol stack: A comprehensive survey, Computer Networks **213**, 109092 (2022).
[25] J. I. Cirac, A. K. Ekert, and C. Macchiavello, Optimal purification of single qubits, Physical Review Letters **82**, 4344–4347 (1999).
[26] M. Keyl and R. F. Werner, The rate of optimal purification procedures (1999), arXiv:quant-ph/9910124 [quant-ph].
[27] A. M. Childs, H. Fu, D. Leung, Z. Li, M. Ozols, and V. Vyas, Streaming quantum state purification, Quantum **9**, 1603 (2025).
[28] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Quantum fingerprinting, Phys. Rev. Lett. **87**, 167902 (2001).
[29] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, Physical Review X **11**, 041036 (2021).
[30] T. E. O'Brien, G. Anselmetti, F. Gkritsis, V. E. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, K. Anderson, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, J. C. Bardin, A. Bengtsson, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, J. Campero, Z. Chen, B. Chiaro, D. Chik, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, S. Demura, I. Drozdov, A. Dunsworth, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, V. S. Ferreira, L. Flores Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, R. Gosula, A. Grajales Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, P. Heu, M. R. Hoffmann, S. Hong, T. Huang, A. Huff, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, T. Khattar, M. Khezri, M. Kieferová, S. Kim, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. T. Lill, W. Liu, W. P. Livingston, A. Locharla, F. D. Malone, S. Mandrà, O. Martin, S. Martin, J. R. McClean, T. McCourt, M. McEwen, X. Mi, A. Mieszala, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, R. Movassagh, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, S. Omonije, A. Opremcak, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, C. Rocque, P. Roushan, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, J. Skruzny, W. C. Smith, R. D. Somma, G. Sterling, D. Strain, M. Szalay, D. Thor, A. Torres, G. Vidal, B. Villalonga, C. Vollgraff Heidweiller, T. White, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, N. Zobrist, D. Bacon, S. Boixo, Y. Chen, J. Hilton, J. Kelly, E. Lucero, A. Megrant, H. Neven, V. Smelyanskiy, C. Gogolin, R. Babbush, and N. C. Rubin, Purification-based quantum error mitigation of pair-correlated electron simulations, Nature Physics **19**, 1787–1792 (2023).
[31] D. Grier, D. Leung, Z. Li, H. Pashayan, and L. Schaeffer, Streaming quantum state purification for general mixed states (2025), arXiv:2503.22644 [quant-ph].
[32] M. Grafe, K. Zhou, Z. Tekin, Z. Lin, S. Li, F. Zhang, V. Ivannikov, and T. Byrnes, Ultrahigh threshold nonstabilizer nonlinear quantum error correcting code (2025), arXiv:2506.10445 [quant-ph].
[33] A. G. Fowler, A. M. Stephens, and P. Groszkowski, High-threshold universal quantum computation on the surface code, Physical Review A—Atomic, Molecular, and Optical Physics **80**, 052312 (2009).
[34] B. Eastin and E. Knill, Restrictions on transversal encoded quantum gate sets, Physical review letters **102**, 110502 (2009).
[35] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, Optimal and secure measurement protocols for quantum sensor networks, Phys. Rev. A **97**, 042337 (2018).
[36] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, Spin squeezing and reduced quantum noise in spectroscopy, Phys. Rev. A **46**, R6797 (1992).
[37] J. Ma, Y. Shen, J. Huang, and C. Lee, Quantum metrology via floquet-engineered two-axis twisting and turn dynamics (2025), arXiv:2409.08524 [quant-ph].

Appendix A: PQEC Overhead

Eq. (24) is evaluated using experimental shots $k \in \{1, \ldots, N_{\text{samp}}\}$. From each shot, we obtain (i) a SWAP-outcome string $\vec{\sigma}_\ell^{(k)}$ and hence a parity weight $\Omega_k := \Omega_{\vec{\sigma}_\ell^{(k)}} \in \{\pm 1\}$, and (ii) a single-shot measurement outcome o_k from measuring the observable O on the output register (i.e. $o_k \in \text{spec}(O)$). Define the sample means

$$\widehat{A} := \frac{1}{N_{\text{samp}}} \sum_{k=1}^{N_{\text{samp}}} \Omega_k \, o_k, \quad \widehat{B} := \frac{1}{N_{\text{samp}}} \sum_{k=1}^{N_{\text{samp}}} \Omega_k, \quad \text{(A1)}$$

so that $\mathbb{E}[\widehat{A}] = \text{Tr}\big(O\rho^N\big)$ and $\mathbb{E}[\widehat{B}] = \text{Tr}\big(\rho^N\big)$, and we estimate $\langle O \rangle_\ell$ by $\widehat{\langle O \rangle}_\ell := \widehat{A}/\widehat{B}$. By a standard delta-method (large-N_{samp}) expansion for ratio estimators, the asymptotic variance is

$$\text{Var}\big(\widehat{\langle O \rangle}_\ell\big) \;\approx\; \frac{1}{N_{\text{samp}} \, \text{Tr}(\rho^N)^2} \, \text{Var}\Big(\Omega_{\vec{\sigma}_\ell} \, [\, \langle O \rangle_{\vec{\sigma}_\ell} - \langle O \rangle_\ell \,]\Big), \quad \text{(A2)}$$

and hence the standard error is

$$\epsilon \;\approx\; \frac{\sqrt{\text{Var}\big(\Omega_{\vec{\sigma}_\ell} \, [\, \langle O \rangle_{\vec{\sigma}_\ell} - \langle O \rangle_\ell \,]\big)}}{\sqrt{N_{\text{samp}}} \, |\, \text{Tr}(\rho^N)|}. \quad \text{(A3)}$$

Since $\Omega^2 = 1$ and $\mathbb{E}[\Omega(o - \langle O \rangle_\ell)] = 0$, one also has the simplification $\text{Var}\big(\Omega[\, o - \langle O \rangle_\ell \,]\big) = \mathbb{E}[(o - \langle O \rangle_\ell)^2]$. In particular, if O is normalized so that $\|O\|_\infty \leq 1$ (e.g. Pauli strings), then $\text{Var}(\Omega[\, o - \langle O \rangle_\ell \,]) \leq 1$ giving

$$\epsilon \;\lesssim\; \frac{1}{\sqrt{N_{\text{samp}}} \, |\, \text{Tr}(\rho^N)|}. \quad \text{(A4)}$$

Appendix B: Pauli basis expansion

We connect the spectral purification argument of Sec. V B to the Pauli-basis expansion. First expand mixed states using the M-qubit Pauli expansion

$$\rho = \frac{1}{2^M} \sum_P r_P P, \quad \text{(B1)}$$

where the sum runs over all 4^M Pauli strings $P \in \{I, X, Y, Z\}^{\otimes M}$ and $r_P := \text{Tr}(\rho P) \in \mathbb{R}$. In these coordinates, the purity is

$$\text{Tr}\big(\rho^2\big) = \frac{1}{2^M} \sum_P r_P^2, \quad \text{(B2)}$$

Depolarizing noise contracts all non-identity Pauli components uniformly, $r_P \mapsto (1-p) \, r_P$ for $P \neq I$, producing an *isotropic* shrinkage of the generalized Bloch vector and a corresponding reduction in purity. By contrast, the PQEC map $\mathcal{P}$ does not, in general, act as an isotropic (radial) expansion in Pauli space. Indeed, although $\text{Tr}\big(\rho^2\big) = 2^{-M} \sum_P r_P^2$ implies that $\mathcal{P}$ increases the Euclidean radius $\sum_{P \neq I} r_P^2$ whenever it increases purity, the map $\mathcal{P}$ is nonlinear and spectral: it preserves the eigenbasis and updates only the eigenvalues according to (C1). An isotropic rescaling $r_P \mapsto \alpha \, r_P$ (with a single α for all $P \neq I$) would correspond to an affine, rotationally invariant action on operator space, whereas (C1) shows that $\mathcal{P}$ depends on the full set $\{\lambda_i\}$ through the normalization $1 + \sum_j \lambda_j^2$, and therefore generically changes not only the length but also the direction of the Pauli-coefficient vector $(r_P)_{P \neq I}$.

Nevertheless, in the isotropic regime, the dynamics are confined to highly symmetric families of spectra, so the action of $\mathcal{P}$ is well captured by a one-parameter "radial" picture: depolarizing (or fully twirled dephasing) drives the spectrum toward uniformity by shrinking $\sum_j \lambda_j^2$, while purification counteracts this by polarizing the spectrum (increasing the contrast among the λ_i via (35)), which necessarily increases $\text{Tr}\big(\rho^2\big) = \sum_i \lambda_i^2$ and hence increases $\sum_{P \neq I} r_P^2$. This explains why our protocol is naturally well matched to depolarizing noise and why fully twirled dephasing exhibits identical performance: the noise acts isotropically in Pauli space, and purification reliably increases the Bloch radius, even though it is not, in general, an isotropic expansion map on the individual r_P components.

Appendix C: Purification and purity

Here we show that, in general, the purification step (6) acts to increase the purity of a given state. Eq. (6) shows that the map is *spectral*: if $\rho = U \, \text{diag}(\boldsymbol{\lambda}) \, U^\dagger$ with eigenvalues $\boldsymbol{\lambda} = (\lambda_i)$, then

$$\rho' = \mathcal{P}(\rho) = U \, \text{diag}(\boldsymbol{\lambda}') \, U^\dagger, \qquad \lambda_i' = \frac{\lambda_i^2}{\sum_j \lambda_j^2}. \quad \text{(C1)}$$

Thus, the eigenbasis is unchanged, and only the spectrum is updated. Moreover, because the function $g(x) = x^2$ is strictly increasing on $[0, 1]$, the ordering of eigenvalues is preserved, and their *relative contrast increases* according to (35).

The purity from (C1) is given by

$$\text{Tr}\Big(\rho'^2\Big) = \sum_i (\lambda_i')^2 = \frac{\sum_i \lambda_i^4}{\big(\sum_j \lambda_j^2\big)^2}. \quad \text{(C2)}$$

The map strictly increases purity, except at the fixed points (pure states and the maximally mixed state). A direct way to see this is to work on the spectrum:

Let $\rho = \sum_i \lambda_i \, |i\rangle\langle i|$ with eigenvalues $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$, and define power sums $s_n := \sum_i \lambda_i^n$. Under the map (22), the eigenvalues update as $\lambda_i' = \lambda_i^2 / s_2$.

Therefore,

$$\mathrm{Tr}\big[(\rho')^2\big] = \sum_i (\lambda_i')^2 = \frac{s_4}{s_2^2}, \qquad \mathrm{Tr}\big(\rho^2\big) = s_2, \tag{C3}$$

and hence

$$\mathrm{Tr}\big[(\rho')^2\big] - \mathrm{Tr}\big(\rho^2\big) = \frac{s_4 - s_2^3}{s_2^2}. \tag{C4}$$

We can see that $s_4 \geq s_2^3$ by using Hölder's inequality. Equality holds iff the spectrum is flat on its support, i.e. $\rho \propto \Pi$ for some projector Π (in particular, for pure states and for the maximally mixed state).

Appendix D: Fidelity convergence under SWAP purification

In this section we show monotonic improvement and convergence of the fidelity $F \to 1$ under PQEC for the global depolarizing channel of Sec. VI.

For the Werner state (45), ρ commutes with $\rho_0 = |\psi\rangle\langle\psi|$. Then the spectrum of ρ is

$$\lambda_1 = F, \qquad \lambda_2 = \cdots = \lambda_D = \frac{1-F}{D-1}. \tag{D1}$$

It follows that $\mathrm{Tr}\,\rho^2 = F^2 + \frac{(1-F)^2}{D-1}$ and $\langle\psi|\,\rho^2\,|\psi\rangle = F^2$. Under the PQEC update (7) the fidelity update is therefore

$$F' = g(F) := \frac{F^2}{F^2 + \dfrac{(1-F)^2}{D-1}}, \qquad D \geq 2. \tag{D2}$$

This recursive relation has the following properties.

Theorem 1 (Monotone behavior and convergence in the isotropic family)**.** *For any $D \geq 2$, the recursion $F' = g(F)$ in Eq.* (D2) *has fixed points at $F_\star = 1$ and $F_\star = 1/D$ (and also $F_\star = 0$ for the extended recursion on $F \in [0,1]$). Moreover: (i) if $F > 1/D$ then F' is strictly increasing and converges to 1; (ii) $F = 1/D$ is a fixed point (the maximally mixed state in the Werner family). The fixed point at $F = 1$ is locally attractive with* quadratic *rate.*

Proof. (i) $g(F) > F$ for $F \in (1/D, 1)$: For $F \in (0,1)$, compute

$$\begin{aligned} g(F) - F &= \frac{F^2}{F^2 + \frac{(1-F)^2}{D-1}} - F \\ &= \frac{F(1-F)\,(DF-1)}{(D-1)F^2 + (1-F)^2}. \end{aligned} \tag{D3}$$

The denominator is strictly positive and $F(1-F) > 0$ on $(0,1)$, hence $\mathrm{sign}(g(F) - F) = \mathrm{sign}(DF - 1)$. Therefore $g(F) > F$ iff $F > 1/D$, with equality at $F \in \{1/D, 1\}$.

(ii) Boundedness: $g(F) \in [0,1]$ for $F \in [0,1]$, and $g(F) < 1$ unless $F = 1$, since the denominator in (D2) exceeds the numerator unless $1 - F = 0$.

(iii) Convergence and fixed point: If $F_0 > 1/D$, then $F_{n+1} = g(F_n) > F_n$ and the sequence is monotone increasing and bounded above by 1, hence convergent to some limit $L \leq 1$. By continuity, $L = g(L)$, so L must be a fixed point. Since $L \geq F_0 > 1/D$ and $g(F) > F$ on $(1/D, 1)$, the only consistent limit is $L = 1$. If $F_0 = 1/D$, then $F_n \equiv 1/D$ for all n.

(iv) Local rate near $F = 1$: Write $F = 1 - \varepsilon$ with $\varepsilon \ll 1$. Then

$$F' = \frac{(1-\varepsilon)^2}{(1-\varepsilon)^2 + \frac{\varepsilon^2}{D-1}} = 1 - \frac{\varepsilon^2}{D-1} + O(\varepsilon^3), \tag{D4}$$

so $1 - F'$ is quadratic in ε. This shows superlinear (quadratic) convergence to $F = 1$. □

Appendix E: Small error expansion for local depolarizing channel

Consider the Pauli expansion of an arbitrary pure state $\rho_0 = |\psi_0\rangle\langle\psi_0|$ as given in (56). Define the weight $w(P)$ of a Pauli string P as the number of non-identity single-qubit factors in P:

$$w(P) := \big|\{m \in \{1, \ldots, M\} : P_m \neq I\}\big|. \tag{E1}$$

Under local depolarizing noise, each non-identity factor is contracted by $\eta = 1 - \frac{4p}{3}$, so a string of weight $w(P)$ acquires attenuation $\eta^{w(P)}$. The noisy pre-purification state is therefore

$$\rho \;=\; \mathcal{E}_p(\rho_0) \;=\; \frac{1}{2^M} \sum_P \eta^{\,w(P)} r_P P. \tag{E2}$$

Using Pauli orthonormality, we obtain

$$F(\rho) = \langle\psi_0|\rho|\psi_0\rangle = \frac{1}{2^M} \sum_P \eta^{\,w(P)} r_P^2, \tag{E3}$$

$$\mathrm{Tr}\big(\rho^2\big) = \frac{1}{2^M} \sum_P \eta^{\,2w(P)} r_P^2. \tag{E4}$$

It is convenient to group the squared coefficients by Pauli weight. Define the Pauli-weight distribution of the target

$$a_k \;:=\; \frac{1}{2^M} \sum_{P:\, w(P)=k} r_P^2, \qquad \sum_{k=0}^{M} a_k = 1. \tag{E5}$$

Then Eqs. (E3)–(E4) become

$$F(\rho) \;=\; \sum_{k=0}^{M} a_k\, \eta^k, \tag{E6}$$

$$\mathrm{Tr}\big(\rho^2\big) \;=\; \sum_{k=0}^{M} a_k\, \eta^{2k}. \tag{E7}$$

The coefficients $\{a_k\}$ encode how "globally" the target $|\psi_0\rangle$ is supported in Pauli space: states whose weight distribution is biased toward large k suffer faster degradation of both $F(\rho)$ and $\mathrm{Tr}(\rho^2)$ as p increases.

For weak depolarization ($p \ll 1$), we can expand $\eta^k = (1 - \frac{4p}{3})^k$ to first order, which gives

$$F(\rho) = 1 - \frac{4}{3}\,\bar{k}\,p + O(p^2), \tag{E8}$$

$$\mathrm{Tr}(\rho^2) = 1 - \frac{8}{3}\,\bar{k}\,p + O(p^2), \tag{E9}$$

where

$$\bar{k} := \sum_k k\,a_k \tag{E10}$$

is the average Pauli weight of the target $|\psi_0\rangle$.

We now evaluate the action of the PQEC map (7). The corresponding output fidelity is

$$F(\mathcal{P}(\rho)) \;=\; \frac{\langle\psi_0|\,\rho^2\,|\psi_0\rangle}{\mathrm{Tr}(\rho^2)}. \tag{E11}$$

To obtain its small-p behavior, write $\rho = \rho_0 + \delta\rho$ with $\delta\rho = O(p)$ and $\mathrm{Tr}(\delta\rho) = 0$. Then

$$\begin{aligned}\langle\psi_0|\,\rho^2\,|\psi_0\rangle &= \langle\psi_0|\,(\rho_0+\delta\rho)^2\,|\psi_0\rangle \\ &= 1 + 2\,\langle\psi_0|\,\delta\rho\,|\psi_0\rangle + \langle\psi_0|\,\delta\rho^2\,|\psi_0\rangle\,,\end{aligned} \tag{E12}$$

and

$$\begin{aligned}\mathrm{Tr}(\rho^2) &= \mathrm{Tr}\big[(\rho_0+\delta\rho)^2\big] \\ &= 1 + 2\,\mathrm{Tr}(\rho_0\delta\rho) + \mathrm{Tr}(\delta\rho^2) \\ &= 1 + 2\,\langle\psi_0|\,\delta\rho\,|\psi_0\rangle + \mathrm{Tr}(\delta\rho^2),\end{aligned} \tag{E13}$$

where we used $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\mathrm{Tr}(\rho_0\delta\rho) = \langle\psi_0|\,\delta\rho\,|\psi_0\rangle$. Therefore the *linear* terms in $\delta\rho$ cancel in the ratio (E11), giving

$$\begin{aligned}F(\mathcal{P}(\rho)) &= 1 - \Big(\mathrm{Tr}(\delta\rho^2) - \langle\psi_0|\,\delta\rho^2\,|\psi_0\rangle\Big) + O(p^3) \\ &= 1 - O(p^2).\end{aligned} \tag{E14}$$

In other words, under the PQEC map, the leading deviation of the purified fidelity is *quadratic* in the physical error rate: the $O(p)$ infidelity present in $F(\rho)$ is removed by normalization in (E11), and the remaining contribution is governed by the second-order term $\delta\rho^2$ (with a state- and channel-dependent coefficient). This is consistent with the spectral picture that $\mathcal{P}(\rho)$ effectively squares eigenvalues and renormalizes, producing superlinear convergence in the high-fidelity regime.

Appendix F: Anisotropic errors

When considering the behavior of PQEC in the context of depolarization vs. dephasing errors, distinct behavior arises. Consider a single qubit with target Bloch unit vector $\hat{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. With local Z-dephasing with error probability p_z contracts the transverse components by $\beta_z := 1 - 2p_z$, giving

$$\vec{r} = (\beta_z \sin\theta\cos\phi,\ \beta_z\sin\theta\sin\phi,\ \cos\theta). \tag{F1}$$

Two consequences follow:

(i) *Tilt toward the dephasing axis.* Unless $\theta \in \{0, \pi\}$, the direction of $\vec{n}$ moves toward $\pm Z$.

(ii) *PQEC cannot repopulate axes.* Under the PQEC update (7), a single qubit state's Bloch sphere vector updates as (34). This preserves direction and only stretches the *existing* components. If X/Y coherence was strongly suppressed, the map cannot "rotate" amplitude back into X/Y; it can only scale up what is left. Quantitatively, with $F = \frac{1}{2}(1 + \hat{n}\cdot\vec{r})$ and $\mathrm{Tr}(\rho^2) = \frac{1}{2}(1 + \|\vec{r}\|^2)$,

$$\hat{n}\cdot\vec{r} = \beta_z\sin^2\theta + \cos^2\theta = 1 - (1-\beta_z)\sin^2\theta \tag{F2}$$

$$\|\vec{r}\|^2 = \beta_z^2\sin^2\theta + \cos^2\theta. \tag{F3}$$

After one (effective) purification round,

$$F' = \frac{1}{2}\left(1 + \frac{2}{1+\|\vec{r}\|^2}\,[\,\beta_z\sin^2\theta + \cos^2\theta\,]\right). \tag{F4}$$

For equatorial targets ($\theta = \pi/2$), the input coherence is just β_z, and the per-round gain is linear:

$$F = \tfrac{1}{2}(1+\beta_z) \quad\longrightarrow\quad F' = \tfrac{1}{2}\left(1 + \frac{2\beta_z}{1+\beta_z^2}\right) \tag{F5}$$

$$= \tfrac{1}{2}\big(1 + 2\beta_z + O(\beta_z^3)\big). \tag{F6}$$

When $\beta_z \ll 1$ (strong dephasing) each round increases the transverse coherence by at most a constant factor. The core inefficiency remains: *anisotropic* loss converts target amplitude into a component (along Z) that PQEC will not rotate back; it can only amplify the residual transverse component that survives the dephasing.