Efficient diagnostics for quantum error correction

Pavithran Iyer,[1,2,3,*,†] Aditya Jain,[1,2,3,4,*] Stephen D. Bartlett,[5] and Joseph Emerson[1,2,3,4]

[1]*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[2]*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[3]*Quantum Benchmark Inc., Kitchener, Ontario N2H 5G5, Canada*
[4]*Keysight Technologies Canada, Kanata, ON K2K 2W5, Canada*
[5]*Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney, New South Wales 2006, Australia*

(Received 27 August 2021; revised 29 September 2022; accepted 10 October 2022; published 27 December 2022)

Fault-tolerant quantum computing will require accurate estimates of the resource overhead under different error correction strategies, but standard metrics such as gate fidelity and diamond distance have been shown to be poor predictors of logical performance. We present a scalable experimental approach based on Pauli error reconstruction to predict the performance of concatenated codes. Numerical evidence demonstrates that our method significantly outperforms predictions based on standard error metrics for various error models, even with limited data. We illustrate how this method assists in the selection of error correction schemes.

DOI: 10.1103/PhysRevResearch.4.043218

I. INTRODUCTION

Noise is pervasive in quantum processing and must be overcome to achieve the disruptive capabilities of quantum computing. Fault tolerance (FT) guarantees reliable logical quantum computation in the presence of noise under prescribed conditions often oversimplified as achieving a threshold on gate error rates. However, achieving low logical error rates (LERs) in practice is challenging, in part because of the large overheads required in terms of the number of additional qubits and gates. Optimizing quantum error correction (QEC) strategies for a particular platform requires accurate prediction of its expected logical performance. For instance, in the presence of biased noise [1–6], tailored codes have been shown to outperform traditional codes that are designed to correct unstructured noise. However, bias is only one of the exponentially many parameters that describe the noise on n physical qubits. This work addresses the lack of tools for predicting the logical performance of a fault tolerant architecture based on a description of noise at the physical level.

The existing framework for choosing a FT scheme is centered around the threshold theorem [7,8] which provides a threshold on the physical noise strength below which reliable quantum computation can be guaranteed. However, directly applying the theorem to realistic noise has several challenges. The FT threshold is derived under oversimplified conditions that implicitly model a physical noise process as an incoherent error model with the same diamond distance. This leads to loose estimates of the logical performance when the noise has coherence or strong correlations. Another is that diamond distance, which is usually invoked for assessing error rates in FT proofs, cannot be measured in a scalable way [9]. It has been shown that the resource overheads for a fault tolerant architecture depend critically on the precise relationship between the architecture and the underlying error model. While there are several well-studied error metrics, none of them can accurately predict the LER of a QEC [10]. In this work we address this crucial deficiency prevalent in all these metrics.

Here, we present a new figure of merit specifically tailored to predict the performance of a class of error correcting codes, namely concatenated codes, which can be measured efficiently using experimental protocols. As opposed to average gate fidelity and diamond distance, our approach captures the interplay between the physical noise model and the choice of a fault tolerant architecture. Our method leverages randomized compiling (RC) [11] to create an effective Pauli noise on the physical qubits, and then uses noise reconstruction (NR) techniques [12–14] to estimate Pauli error probabilities. Using these experimental data, we design a *logical estimator* that predicts the total probability of Pauli errors that a code cannot correct. While exactly computing this quantity is inefficient for a generic code, we introduce an efficient approximation for concatenated codes. We provide a bound on the efficiency and demonstrate the accuracy of our method through numerical simulations in several noise scenarios of interest. Finally, as an application, we demonstrate how the logical estimator pinpoints the selection of a suitable error correcting code for differing noise environments.

II. BACKGROUND

A wide class of Markovian noise processes is formally described by completely positive trace preserving (CPTP) maps

*The authors contributed equally to the work.
†pavithran.sridhar@gmail.com

[15]. There are several inequivalent ways to define the strength of noise modelled by a CPTP map $\mathcal{E}$. Of these, the two most widely used to study fault tolerance are the *average gate infidelity* [16–18]: $r(\mathcal{E}) = 1 - \int d\psi \,\mathrm{tr}[|\psi\rangle\langle\psi|\,\mathcal{E}(|\psi\rangle\langle\psi|)]$, and the *diamond distance* [19–23]: $||\mathcal{E} - \mathcal{I}||_\diamond = \max_\rho ||(\mathcal{E}\otimes\mathcal{I})\rho - \rho||_1$. While the average gate infidelity can be efficiently estimated using randomized benchmarking [24–27], the diamond distance satisfies mathematical properties that are needed in FT proofs [28–30]. Other standard error metrics include the 2-norm [31], Bures distance [32], Uhlmann fidelity [33], unitarity [34], channel entropy [35,36] and the adversarial error probability [7]. None of these reflect the logical performance of a code [10,37].

The net effect of a physical noise process $\mathcal{E}_0$ together with a quantum error correction (QEC) routine using an $[[n, k]]$ stabilizer code [38] is captured by the *effective logical channel* $\mathcal{E}_1^s$ [39] acting on an encoded state $\overline{\rho}$ as

$$\mathcal{E}_1^s(\overline{\rho}) = R_s\,\Pi_s\,\mathcal{E}_0(\overline{\rho})\,\Pi_s\,R_s^\dagger/\mathsf{Pr}(s), \tag{1}$$

where $\mathsf{Pr}(s)$ is the probability of measuring the syndrome outcome s, Π_s is the syndrome projector and R_s is the corresponding recovery. The average logical channel $\overline{\mathcal{E}}_1$ is given by [10,40] $\overline{\mathcal{E}}_1(\overline{\rho}) = \sum_s \mathsf{Pr}(s)\mathcal{E}_1^s(\overline{\rho})$.

Concatenated quantum codes are a popular family of codes of increasing sizes [41] often used to guarantee error suppression in fault tolerance proofs [7,42]. Physical qubits of a code $\mathcal{C}_{\ell+1}$ are encoded using a code $\mathcal{C}_\ell$, for $1 \leqslant \ell \leqslant L-1$, yielding a *level-L* concatenated code. The recursive encoding structure is represented by a tree where the ith node at a depth $(L-\ell)$ denotes a quantum error correcting code block $\mathcal{C}_{\ell,i}$. The subtree of the node is itself a concatenated code, denoted by $\mathcal{C}^\star_{\ell,i}$, consisting of $(n^\ell-1)/(n-1)$ code blocks. There are $n-1$ independent stabilizer measurements corresponding to each of the code blocks of $\mathcal{C}^\star_{\ell,i}$. The resulting error syndrome $s(\mathcal{C}^\star_{\ell,i})$ has $(n^\ell-1)$ bits, which can be grouped into subsets of $n-1$ bits that are identified by the code blocks. We identify the subset of syndrome bits obtained by measurements on a code block $\mathcal{C}_{\ell,j}$ by $s(\mathcal{C}_{\ell,j})$.

We consider the following iterative routine for QEC in concatenated codes. For each level $\ell = 1, \ldots, L$: (i) syndromes are extracted for each code block $\mathcal{C}_{\ell,1}, \ldots, \mathcal{C}_{\ell,n}$, and (ii) a minimum-weight correction [43] is applied in each case. Although we assume the popular choice of minimum-weight decoder in (ii), the methods prescribed in this work can be adapted to any lookup table decoder [44]. The correction applied at any level depends on the syndrome history of the code blocks in the lower levels.

The effective channel for a level-ℓ concatenated code can also be computed in a recursive fashion using Eq. (1) where $\mathcal{E}_0$ is replaced by effective channel on the level-$(\ell-1)$ code blocks, i.e., $\mathcal{E}^s_{\ell-1,1}\otimes\cdots\otimes\mathcal{E}^s_{\ell-1,n}$ [39,45]. The performance of the level-ℓ concatenated code can be quantified [10] by the infidelity $r(\overline{\mathcal{E}}_\ell) = \sum_s \mathsf{Pr}(s) r(\mathcal{E}^s_\ell)$ of the average logical channel $\overline{\mathcal{E}}_\ell$. As the number of syndromes grow exponentially with the number of levels, Monte Carlo sampling techniques described in Appendix G can be used to estimate this average.

III. METHODS

While the special setting of Pauli errors drastically simplifies the predictability problem, realistic noise processes are nonetheless poorly described by Pauli error models. To circumvent this problem, we recall a straightforward application of RC [11] to FT circuits, that allows us to model the effect of complex noise processes by simple Pauli errors. In other words, RC ensures that there is no effect on the LER from parameters of the physical channel other than the Pauli error probabilities. The physical twirling gates required to do RC can be absorbed into the logical gadgets of FT circuits at no additional cost in overhead.

A. Quantum error correction with RC

We now show how randomized compiling (RC) can be performed in fault tolerant circuits. Note that a Pauli error P can be decomposed with reference to a stabilizer code: $P = \overline{P}\,S_P\,E_P$, where S_P is an element of the stabilizer group $\mathcal{S}$, $\overline{P}$ is a logical Pauli operator in $\mathcal{L} = \mathcal{N}(\mathcal{S})/\mathcal{S}$, and E_P is an element of $\mathcal{N}(\mathcal{L})/\mathcal{S}$, usually called a pure error [45,46]. Unlike pure errors, stabilizers and logical operators commute with QEC routines. A Pauli error P can be compiled into QEC resulting in a new quantum error correction routine $QEC(P)$ in which the input to the decoder corresponding to a syndrome outcome s is $s \oplus s(P)$ [47,48].

In fault tolerant circuits, each logical gate $\overline{G}$ is sandwiched between QEC routines. Following the prescription in Ref. [11], we divide logical gates into two sets: $\mathcal{S}_1$ and $\mathcal{S}_2$, calling them easy and hard gates, respectively. A crucial requirement for $\mathcal{S}_1$ and $\mathcal{S}_2$ is

$$\overline{G}\,T\,\overline{G}^\dagger\,QEC = QEC(T)\,\overline{C}, \tag{2}$$

for all easy logical gates $\overline{C} \in \mathcal{S}_1$, n-qubit Pauli gates T, and hard gates $\overline{G}$. Recall that $QEC(T)$ refers to the compilation of the Pauli gate T in the QEC routine, discussed in the background section. The previous requirement follows from

$$\overline{GT\overline{G}^\dagger}\,QEC = \overline{GT\overline{G}^\dagger}\,S_T E_T\,QEC \tag{3}$$

$$= \overline{GT\overline{G}^\dagger}\,QEC(E_T), \tag{4}$$

where in Eq. (3) we have used the decomposition of Pauli gates with reference to a stabilizer code. Note that the expression $\overline{G}\,\overline{T}\,\overline{G}^\dagger$ in Eq. (4) is guaranteed to be an easy gate for a choice of easy and hard gate sets in Ref. [11].

Figure 1(a) shows a canonical presentation of a quantum circuit, where the kth clock cycle is composed of an easy gate $\overline{C}_k$ and a hard gate $\overline{G}_k$, sandwiched between QEC routines. Noise processes affecting easy and hard gates are denoted by $\mathcal{E}_{1,k}$ and $\mathcal{E}_{2,k}$, respectively. These complex processes can be tailored to Pauli errors by inserting Pauli gates $T_{1,k}, T_{1,k}^\dagger, T_{2,k}, T_{2,k}^\dagger$. However, to guarantee that they be applied in a noiseless fashion, we compile them into the existing gates in the fault tolerant circuit. This is achieved in two steps. First, $T_{1,k}^\dagger$ and $T_{2,k}$ are compiled into QEC following $\mathcal{E}_{1,k}$, resulting in $QEC(T_{1,k}^\dagger T_{2,k})$. Second, $T_{2,k}^\dagger$ is propagated across $\overline{G}_k$, and compiled with $QEC\,\overline{C}_{k+1}T_{k+1}$, resulting in a *dressed gate* $\overline{C^D_{k+1}} = \overline{G}_k\,T_k\,\overline{G}_k^\dagger\;QEC\;\overline{C}_{k+1}\;T_{k+1}$. It follows from Eq. (2)

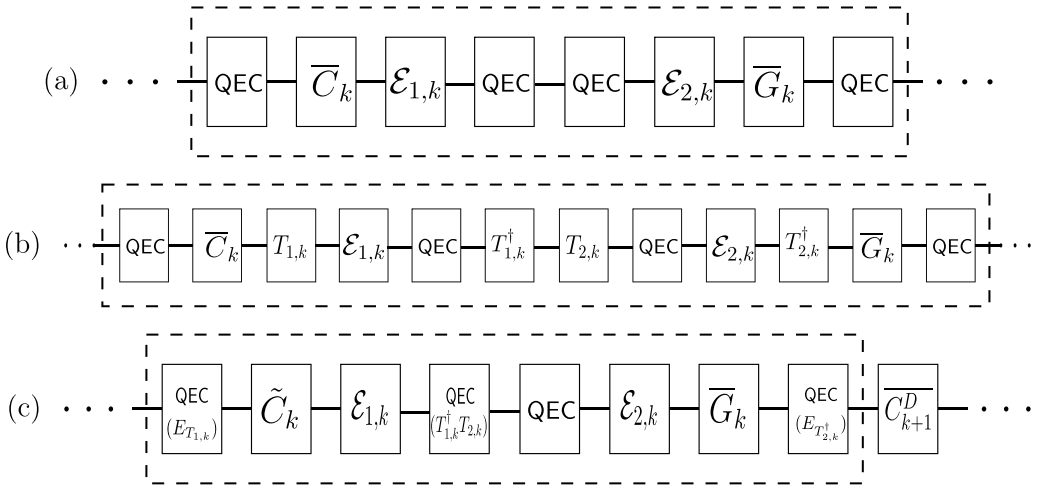


FIG. 1. Compiling twirling (random physical Pauli) gates into fault tolerant gadgets. Panel (a) shows the noisy gates in the kth clock cycle of a fault tolerant quantum algorithm presented in the standard form prescribed in Ref. [11]. Twirling gates are inserted in panel (b) to tailor the noise processes to Pauli errors. These gates are compiled into existing gates by replacing easy gates by their dressed versions in panel (c).

that $\overline{C^D_{k+1}}$ is equivalent to quantum error correction followed by an easy gate.

Figure 1(c) shows the result of compiling all of the twirling gates into the easy gates and quantum error correction routines. Note that the compiled circuit is logically equivalent to the original circuit in the absence of noise. However, in the presence of noise, the average output of the circuit is dictated by the performance of $QEC(T)$ averaged over the different choices of Pauli gates T. This is what we refer to as QEC in the *RC setting*. In practice, this average performance can be achieved by repeating every iteration (shot) of the algorithm with a different Pauli operation compiled into the constituent QEC routines. For the purpose of numerical simulations in this paper, we have used the performance of the QEC routine under the twirled noise process, as a proxy to the performance of QEC in the RC setting.

B. Logical estimator for concatenated codes

With a noise model described by Pauli errors, we first develop the background needed to define notion of a *logical estimator* that can accurately predict the LER. A stabilizer code and a decoder pair is designed to correct a target set of errors, called *correctable errors* [49,50] $\mathcal{E}_C$. For an $[[n, k]]$ code, $\mathcal{E}_C$ can be partitioned into 2^{n-k} disjoint subsets $\mathcal{E}_{C,1}, \ldots, \mathcal{E}_{C,2^{n-k}}$, each of which can be identified with a unique syndrome measurement outcome. The construction of the set $\mathcal{E}_{C,s}$ closely depends on the choice of a decoder. Recall that the output of a decoder on input syndrome s is a Pauli recovery operator R_s, i.e., $R_s \in \mathcal{E}_{C,s}$. A key observation to construct elements in $\mathcal{E}_{C,s}$ besides R_s is that any error of the form R_sS where S is an element of the stabilizer group is also correctable, so, $\mathcal{E}_{C,s} = \{R_sS : S \in \mathcal{S}\}$. Uncorrectable errors cause the quantum error correction scheme to fail. We adopt the notation p_c to denote the total probability of correctable errors,

$$p_c = \sum_{E \in \mathcal{E}_C} \chi_{E,E}, \tag{5}$$

and p_u to denote the total probability of uncorrectable errors, $p_u = 1 - p_c$. It is easy to note that p_u is an upper bound to the standard infidelity metric which is measured by randomized benchmarking, i.e., $r = 1 - \chi_{0,0}$:

$$p_u = r - \sum_{E \in \mathcal{E}_C E \neq \mathbb{I}} \chi_{E,E}. \tag{6}$$

In particular, for Pauli noise processes the following equations show that p_u is exactly the average *logical* infidelity $\bar{r}$:

$$\bar{r} = 1 - \sum_{\substack{E,E' \in \mathcal{E}_C \\ s(E)=s(E'),\, \overline{E}=\overline{E}'}} \chi_{E,E'} \tag{7}$$

$$= r - \sum_{\substack{E,E' \in \mathcal{E}_C,\, E,E' \neq \mathbb{I} \\ s(E)=s(E'),\, \overline{E}=\overline{E}'}} \chi_{E,E'} \tag{8}$$

$$= p_u - \sum_{\substack{E,E' \in \mathcal{E}_C,\, E \neq E' \\ s(E)=s(E'),\, \overline{E}=\overline{E}'}} \chi_{E,E'}. \tag{9}$$

A detailed derivation of Eq. (7) is presented in Appendix A. The expressions in Eqs. (8) and (9) point out a conceptual difference between infidelity and the uncorrectable error probability. While on the one hand, r accounts for the effect of only the trivial correctable error $\mathbb{I}$, p_u on the other hand captures many more degrees of freedom—including all other correctable errors in $\mathcal{E}_C$. Hence, we expect r to be a worse predictor of the logical infidelity than p_u.

It is generally infeasible to enumerate all the $\mathcal{O}(4^{n-k})$ correctable errors for an $[[n, k]]$ stabilizer code, to compute p_u exactly. Our logical estimator is the result of an efficient heuristic to approximate p_u, particularly for concatenated code families. In particular, we use a coarse grained estimate of the probability of a syndrome outcome—a joint probability distribution over $\mathcal{O}(n^\ell)$ syndrome bits—calculated as a product of marginal probability distributions over the n code blocks at level $(\ell - 1)$. This procedure is recursed through the

ℓ levels of the concatenated code. Furthermore, its accuracy is provably high for uncorrelated Pauli error models.

While for concatenated codes, the number of physical qubits itself grows exponentially in the size of a code block n, we can exploit its encoding structure to simplify the complexity of computing p_u. However, it turns out that despite this simplification we cannot exactly compute p_u efficiently, i.e., in time that scales polynomially in the number of physical qubits. This leads us to resort to a heuristic method for a reasonable approximation of p_u for concatenated codes. Here we present a method to measure and compute an approximation, denoted by $\widetilde{p}_u(\mathcal{C}_\ell^\star)$, to the probability of uncorrectable errors for a concatenated code $\mathcal{C}_\ell^\star$: $p_u(\mathcal{C}_\ell^\star)$. For ease of notation we also define the quantities $p_c(\mathcal{C}_\ell^\star) = 1 - p_u(\mathcal{C}_\ell^\star)$ and $\widetilde{p}_c(\mathcal{C}_\ell^\star) = 1 - \widetilde{p}_u(\mathcal{C}_\ell^\star)$.

An error E_ℓ for the level-ℓ concatenated code $\mathcal{C}_\ell^\star$ can be expressed as a tensor product of Pauli errors $E_{\ell-1,i}$ for the level-$\ell-1$ codes $\mathcal{C}_{\ell-1,i}^\star$:

$$E_\ell = \bigotimes_{i=1}^{n} E_{\ell-1,i}. \tag{10}$$

Let us define E_ℓ to be a correctable pattern if the above tensor product corresponds to an encoded version of a correctable error for the code block $\mathcal{C}_\ell$. For example, $E_2 = \overline{X} \otimes \overline{\mathbb{I}}^{\otimes 6}$ is a correctable pattern for the $\ell = 2$ concatenated Steane code since $X \otimes \mathbb{I}^{\otimes 6}$ is a correctable error for the Steane code block.

A correctable error E_ℓ for the concatenated code $\mathcal{C}_\ell^\star$ is either (i) corrected within the lower-level code blocks $\mathcal{C}_{\ell-1,1}^\star, \ldots, \mathcal{C}_{\ell-1,n}^\star$, or (ii) has a nontrivial correction applied by the decoder of the level-ℓ code block $\mathcal{C}_{\ell,1}$. Let us denote the contribution to $p_c(\mathcal{C}_\ell^\star) = 1 - p_u(\mathcal{C}_\ell^\star)$ from case (i) by Λ, while that from case (ii) by Γ; so that

$$p_c(\mathcal{C}_\ell^\star) = \Lambda(\mathcal{C}_\ell^\star) + \Gamma(\mathcal{C}_\ell^\star). \tag{11}$$

Case (i) implies that each of the errors $E_{\ell-1,i}$ are correctable errors for the codes $\mathcal{C}_{\ell-1,i}^\star$. Therefore, the total probability of correctable errors in case (i) admits a recursive definition:

$$\Lambda(\mathcal{C}_\ell^\star) = p_c(\mathcal{C}_{\ell-1,1}^\star) p_c(\mathcal{C}_{\ell-1,2}^\star) \ldots p_c(\mathcal{C}_{\ell-1,n}^\star). \tag{12}$$

Recall that case (ii) is the total probability of nontrivial correctable patterns for $\mathcal{C}_\ell^\star$, i.e.,

$$\Gamma(\mathcal{C}_\ell^\star) = \sum_{E \in \mathcal{E}_\mathcal{C} \setminus \mathbb{I}} \mathsf{Pr}(E_\ell), \tag{13}$$

$$= \sum_{E \in \mathcal{E}_\mathcal{C} \setminus \mathbb{I}} \mathsf{Pr}(E_{\ell-1,1} \otimes E_{\ell-1,2} \otimes \ldots \otimes E_{\ell-1,n}), \tag{14}$$

where we have used the fact that each correctable error corresponds to a pattern according to Eq. (10). A logical error $E_{\ell-1,i}$ occurs on the code block $\mathcal{C}_{\ell-1,i}$ whenever the decoder fails in correcting the physical errors in such a way that the residual effect of the physical noise process affecting the qubits of $\mathcal{C}_{\ell-1,i}^\star$ and the recovery operation applied by the decoder results in $E_{\ell-1,i}$. Let us denote the probability of the decoder for $\mathcal{C}_{\ell-1,i}^\star$ to leave a residual $E_{\ell-1,i}$, conditioned on the syndrome measurements by $\mathsf{Pr}_\mathcal{D}[E_{\ell-1,i} \,|\, s(\mathcal{C}_{\ell-1,i}^\star)]$. We can rewrite Eq. (14) as

$$\Gamma(\mathcal{C}_\ell^\star) = \sum_{E \in \mathcal{E}_\mathcal{C} \setminus \mathbb{I}} \sum_{s(\mathcal{C}_\ell^\star)} \mathsf{Pr}[s(\mathcal{C}_\ell) s(\mathcal{C}_{\ell-1,1}^\star) \ldots s(\mathcal{C}_{\ell-1,n}^\star)] \times \prod_{j=1}^{n} \mathsf{Pr}_\mathcal{D}[E_{\ell-1,j} | s(\mathcal{C}_{\ell-1,j}^\star)], \tag{15}$$

$$= \sum_{E \in \mathcal{E}_\mathcal{C} \setminus \mathbb{I}} \sum_{s(\mathcal{C}_\ell^\star)} \mathsf{Pr}[s(\mathcal{C}_\ell) | s(\mathcal{C}_{\ell-1,1}^\star) \ldots s(\mathcal{C}_{\ell-1,n}^\star)] \times \prod_{j=1}^{n} \mathsf{Pr}_\mathcal{D}[E_{\ell-1,j} | s(\mathcal{C}_{\ell-1,j}^\star)] \mathsf{Pr}[s(\mathcal{C}_{\ell-1,j}^\star)], \tag{16}$$

where $\mathsf{Pr}[s(\mathcal{C}_\ell) | s(\mathcal{C}_{\ell-1,1}^\star) \ldots s(\mathcal{C}_{\ell-1,n}^\star)]$ is the conditional probability of measuring the syndrome outcomes $s(\mathcal{C}_\ell)$ on the code block $\mathcal{C}_\ell$ when the outcomes on the lower-level code blocks $\mathcal{C}_{\ell-1,1}^\star, \ldots, \mathcal{C}_{\ell-1,n}^\star$ are $s(\mathcal{C}_{\ell-1,1}^\star), \ldots, s(\mathcal{C}_{\ell-1,n}^\star)$, respectively. Equivalently,

$$\mathsf{Pr}[s(\mathcal{C}_\ell) | s(\mathcal{C}_{\ell-1,1}^\star) \ldots s(\mathcal{C}_{\ell-1,n}^\star)] = \mathsf{Pr}\big[s(\mathcal{C}_\ell) \big| \mathcal{E}_{\ell-1,1}^{s(\mathcal{C}_{\ell-1,1}^\star)} \ldots \mathcal{E}_{\ell-1,n}^{s(\mathcal{C}_{\ell-1,n}^\star)}\big)\big]. \tag{17}$$

A major hurdle in computing Γ using Eq. (16) is the sum over an exponentially large set of syndrome outcomes for the concatenated code. To circumvent this difficultly, we will apply an efficient heuristic to approximate the probability in Eq. (17). In essence, we will replace the conditional channel $\mathcal{E}_{\ell-1,i}^{s(\mathcal{C}_{\ell-1,i}^\star)}$ by the average logical channel $\hat{\mathcal{E}}_{\ell-1,i}$, which is defined as

$$\hat{\mathcal{E}}_{\ell-1,i} = \sum_{s(\mathcal{C}_{\ell-1,i})} \mathsf{Pr}[s(\mathcal{C}_{\ell-1,i})] \mathcal{E}_{\ell-1,i}^{s(\mathcal{C}_{\ell-1,i})} [\hat{\mathcal{E}}_{\ell-2,1} \otimes \ldots \otimes \hat{\mathcal{E}}_{\ell-2,n}]. \tag{18}$$

Note that $\hat{\mathcal{E}}_{0,j}$ is the physical noise model while $\hat{\mathcal{E}}_{1,j}$ is the exact average logical channel $\overline{\mathcal{E}}_{1,j}$. However, in general for $\ell \geqslant 2$, $\hat{\mathcal{E}}_\ell$ is a coarse-grained approximation for the exact average logical channel $\overline{\mathcal{E}}_\ell$. In other words, $\hat{\mathcal{E}}_{\ell-1,i}$ is computed using the knowledge of the syndrome bits measured only at level $\ell - 1$, while assuming the noise model: $\hat{\mathcal{E}}_{\ell-2,1} \otimes \ldots \otimes \hat{\mathcal{E}}_{\ell-2,n}$, that accounts for the average effect of all syndrome measurements at lower levels.

Replacing the conditional channel $\mathcal{E}_{\ell-1,i}^{s(\mathcal{C}_{\ell-1,i}^\star)}$ in Eq. (16) by the average channel $\hat{\mathcal{E}}_{\ell-1,i}$ defined in Eq. (C11) allows us to approximate Γ by $\widetilde{\Gamma}$ defined as follows:

$$\widetilde{\Gamma}(\mathcal{C}_\ell^\star) = \sum_{E \in \mathcal{E}_\mathcal{C} \setminus \mathbb{I}} \sum_{s(\mathcal{C}_\ell)} \sum_{s(\mathcal{C}_{\ell-1,1}^\star)} \ldots \sum_{s(\mathcal{C}_{\ell-1,n}^\star)} \mathsf{Pr}[s(\mathcal{C}_\ell) | \hat{\mathcal{E}}_{\ell-1,1} \ldots \hat{\mathcal{E}}_{\ell-1,n}] \prod_{j=1}^{n} \mathsf{Pr}_\mathcal{D}(E_{\ell-1,j} \,|\, \hat{\mathcal{E}}_{\ell-1,j}) \mathsf{Pr}[s(\mathcal{C}_{\ell-1,j}^\star)], \tag{19}$$

$$= \sum_{E \in \mathcal{E}_\mathcal{C} \setminus \mathbb{I}} \prod_{j=1}^{n} \mathsf{Pr}_\mathcal{D}(E_{\ell-1,j} \,|\, \hat{\mathcal{E}}_{\ell-1,j}). \tag{20}$$

Denote $\mathcal{R}[s(\mathcal{C}_{\ell-1,i}), P]$ to be the set of n-qubit errors on which a lookup table decoder for the code block $\mathcal{C}_{\ell-1,i}$ leaves a residual logical error P when the error syndrome $s(\mathcal{C}_{\ell-1,i})$ is encountered. Now $\mathsf{Pr}_\mathcal{D}(E_{\ell-1,i} \,|\, \hat{\mathcal{E}}_{\ell-1,j})$ can be computed

recursively:

$$\mathsf{Pr}_{\mathcal{D}}(E_{\ell-1,i}\,|\,\hat{\mathcal{E}}_{\ell-1,i}) = \sum_{Q\in\mathcal{R}[s(\mathcal{C}_{\ell-1,i}),E_{\ell-1,i}]}\prod_{j=1}^{n}\mathsf{Pr}_{\mathcal{D}}(Q_{\ell-2,j}\,|\,\hat{\mathcal{E}}_{\ell-2,j}). \tag{21}$$

Note that the probability of leaving a residual error at level 0 is simply specified by the physical noise model, i.e., $\mathsf{Pr}_{\mathcal{D}}(P|\hat{\mathcal{E}}_{0,j})$ is the probability of the Pauli error P on the physical qubit j. This concludes the method to efficiently compute $\widetilde{\Gamma}$, an approximation to Γ.

Recall that the total probability of correctable errors is given by Eq. (11). An approximation to $p_c(\mathcal{C}^\star_\ell)$, is given by

$$\widetilde{p}_c(\mathcal{C}^\star_\ell) = \widetilde{\Lambda}(\mathcal{C}^\star_\ell) + \widetilde{\Gamma}(\mathcal{C}^\star_\ell), \tag{22}$$

where $\widetilde{\Gamma}$ is defined in Eq. (C8), while $\widetilde{\Lambda}$ is defined in a similar fashion to Eq. (12):

$$\widetilde{\Lambda}(\mathcal{C}^\star_\ell) = \widetilde{p}_c(\mathcal{C}^\star_{\ell-1,1})\widetilde{p}_c(\mathcal{C}^\star_{\ell-1,2})\ldots\widetilde{p}_c(\mathcal{C}^\star_{\ell-1,n}). \tag{23}$$

Using the approximation in Eq. (22), we can efficiently estimate the logical estimator $\widetilde{p}_u$ for concatenated codes.

For i.i.d. Pauli error models with sufficiently small single-qubit infidelity r_0, the quality of approximation is: $|\bar{r}_\ell - \widetilde{p}_u| \leqslant n_C^{\ell+1} r_0^{2+\lfloor (d_C+1)/2\rfloor}$. Here, d_C and n_C describe the distance and the size of a code block of a level-ℓ concatenated code. For instance, using an i.i.d. depolarizing error model with $r_0 = 10^{-3}$ and the level-2 concatenated Steane code, the above expression yields $|\bar{r}_2 - \widetilde{p}_u| \leqslant 5\times10^{-10}$. This is validated by numerics: $\widetilde{p}_u = 4.24\times10^{-9}$ and $\bar{r}_2 = 4.20\times10^{-9}$. A detailed derivation of quality of approximation is provided in Appendix C.

Notably, the time complexity of computing $\widetilde{p}_u$ for the concatenated code: $\mathcal{O}(4^{n_C+\ell}\, n^\ell)$, scales polynomially in the total number of physical qubits n^ℓ, whereas an exact computation of p_u would scale doubly exponentially in ℓ. We will now prove this statement.

Recall that $\widetilde{p}_u = 1 - \widetilde{p}_c(\mathcal{C}^\star_\ell)$, where $\widetilde{p}_c(\mathcal{C}^\star_\ell)$ is an approximation to the total probability of correctable errors. Note that $\widetilde{p}_c(\mathcal{C}^\star_\ell) = \widetilde{\Lambda} + \widetilde{\Gamma}$ where both $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are defined recursively. So, if computing $\widetilde{p}_c(\mathcal{C}^\star_\ell)$ takes time τ_ℓ and computing $\widetilde{\Gamma}$ takes time κ_ℓ, we have

$$\tau_\ell = n\,\tau_{\ell-1} + \kappa_\ell. \tag{24}$$

The recurrence relation in Eq. (C8) for computing $\widetilde{\Gamma}(\mathcal{C}^\star_\ell)$ implies

$$\kappa_\ell = 4n\,\kappa_{\ell-1} + \mathcal{O}(4^n), \tag{25}$$

$$= \mathcal{O}(4^{n+\ell}\, n^\ell). \tag{26}$$

Using the above solution in Eq. (24), we find that

$$\tau_\ell = \mathcal{O}(4^{n+\ell}\, n^\ell). \tag{27}$$

The last equation establishes that computing logical estimator is linear in the total number of qubits i.e., n^ℓ.

IV. RESULTS AND DISCUSSION

We provide numerical evidence to highlight the improvement offered by our methods developed for optimizing FT schemes. We begin with the task of accurately predicting the performance of concatenated Steane codes. We perform numerical simulations of QEC in the RC and non-RC settings under a large ensemble of random CPTP maps applied to the physical qubits. Following Ref. [10], we generate a single qubit CPTP map $\mathcal{E}$ from its Stinespring dilation: a random unitary matrix U of size (8×8), given by $U = e^{-iHt}$ for a complex Hermitian matrix H whose entries are sampled from a Gaussian distribution of unit variance, centered at 0. We vary the time parameter t between 0.001 and 0.1 to vary the noise strength.

Figure 2 shows that logical error rates can vary wildly across physical noise processes with fixed infidelity and diamond distance in agreement with Ref. [10]. The variation, captured by the amount of dispersion in the scatter plots, is quantified using the ratio of the maximum and the minimum

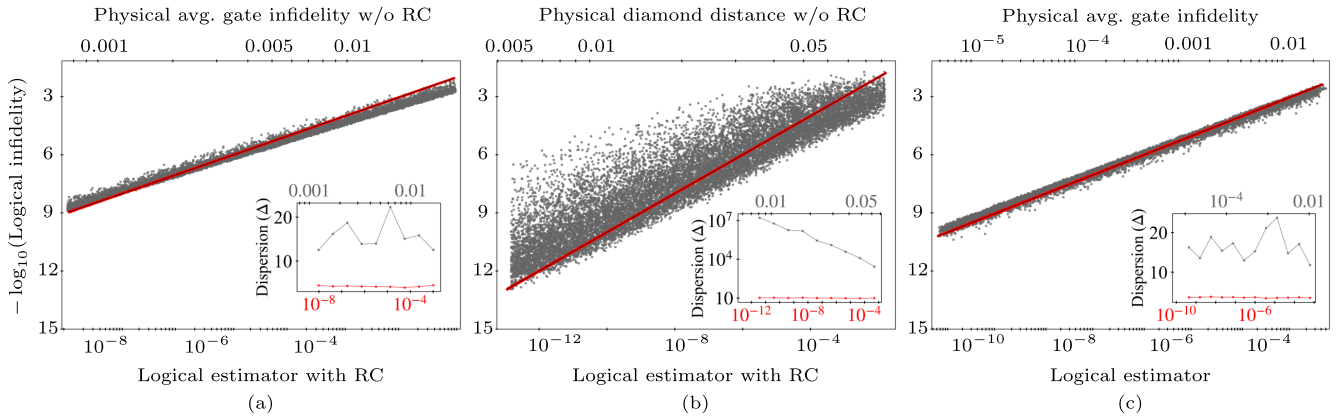


FIG. 2. Predictability of logical infidelity for level-2 concatenated Steane code. Figures 1(a) and 1(b) compare the predictive powers of our logical estimator (red) against two standard error metrics (gray): the average gate infidelity (a) and the diamond distance (b), under an ensemble of 18 000 CPTP maps. Each point $p = (x_p, y_p)$ corresponds to a physical noise process; x_p is its physical error metric and y_p, its LER. The dispersion of points, quantified as Δ in the insets, indicates the predictive power of the physical error metric. While LERs can vary over several orders of magnitude with respect to standard error metrics, our logical estimator is strongly correlated with the LER. (c) Correlated Pauli error model.

LERs across channels of similar physical error rate, denoted by Δ. In other words, we partition the range of physical error rates into bins b_i and use $\Delta(b_i)$ to quantify the amount of dispersion: $\Delta(b_i) = (1/|b_i|)\,(\max_{p\in b_i} y_p)/(\min_{p\in b_i} y_p)$, where $|b_i|$ is the number of channels in the bin b_i. The large fluctuations in the LERs can be attributed to two extreme features of the error metrics. While infidelity controls only one parameter out of the many that specify a noise process, diamond distance suffers from being sensitive to the details of a noise process that are irrelevant to the LER. In addition, standard error metrics can only reveal intrinsic properties of the underlying noise process that are agnostic to the choice of an error correcting code.

Logical estimator with RC, in contrast, is highly correlated with the LER. This improvement can be attributed to two features. First, RC provides a drastic reduction from $\mathcal{O}(12^n)$ parameters that specify an n-qubit Markovian noise process to $\mathcal{O}(4^n)$ Pauli error probabilities. Second, unlike standard error metrics, $\widetilde{p}_u$ carefully accounts for Pauli error probabilities that contribute to the LER. Numerical evidence for drastic gains in predictability using the logical estimator with RC for the class of coherent errors is presented in Appendix D.

The special setting of i.i.d. noise hides the drastic advantages provided by $\widetilde{p}_u$ in predicting logical infidelity because the dominant contribution to $\widetilde{p}_u$ comes from $\chi_{0,0}$, which is also well captured by r. However, for correlated error-models, given only $\chi_{0,0}$, the uncertainty on the LER ranges between extremities 0 and 1. While r is completely insensitive to either of these scenarios, $\widetilde{p}_u$ in contrast helps distinguish between them, thereby providing a far more accurate estimate of the LER.

We support the above argument by numerical studies of correlated Pauli error models generated from a convex combination of an i.i.d process of infidelity r_0 and multiqubit interactions. While the i.i.d. component $\mathcal{E}_{\mathsf{iid}}$ is specified by single qubit error probabilities, multiqubit interactions are specified by an arbitrary subset $\mathfrak{S}$, so, $\mathcal{E}_{\mathsf{cor}}(\rho) = \sum_{P\in\mathfrak{S}} \chi_{P,P} P\,\rho\,P$, where $\chi_{P,P}$ is sampled from the normal distribution with mean and variance $4^n r_0$. The combined Pauli error model is therefore given by $\mathcal{E}(\rho) = q\mathcal{E}_{\mathsf{iid}}(\rho) + (1-q)\mathcal{E}_{\mathsf{cor}}(\rho)$, where $0 \leqslant q \leqslant 1$. Explicitly setting $\chi_{0,0}$ followed by appropriate normalization, ensures that the infidelity of the above noise model is r_0. Figure 2(c) highlights the importance of the $\widetilde{p}_u$ over r for predicting the performance of the concatenated Steane code under correlated Pauli noise processes.

A. Limited noise reconstruction data

Even in the absence of correlations across the n-qubit code blocks of a concatenated code, we require $\mathcal{O}(4^n)$ Pauli error rates from NR to compute $\widetilde{p}_u$. Extracting this exponential sized NR dataset is a challenge for experimentalists. References [14,51] describe how to extract the leading K Pauli error probabilities in a noise process, where $K \ll 4^n$. We want to combine a handful of leading Pauli error rates extracted by NR with a simple method to extrapolate the remaining ones. For a Pauli error Q that is not given in the NR dataset we set

$$\mathsf{Pr}(Q) = (1-r_0)^{n-\mathsf{wt}(Q)}(r_0/3)^{\mathsf{wt}(Q)}, \tag{28}$$

where $\mathsf{wt}(Q)$ is the Hamming weight of Q, and r_0 is derived from the infidelity of the noise process: $r = 1-(1-r_0)^n$. We construct an adversarial error model where the above extrapolation is unlikely to perform well by setting some multiqubit error probabilities that violate Eq. (28). Furthermore, when errors are sampled uniformly from the set of correctable and uncorrectable errors, we observe maximum fluctuations in the LER. However, Fig. 3 presents strong numerical evidence indicating that the simple extrapolation works well in practice even for the adversarial example.

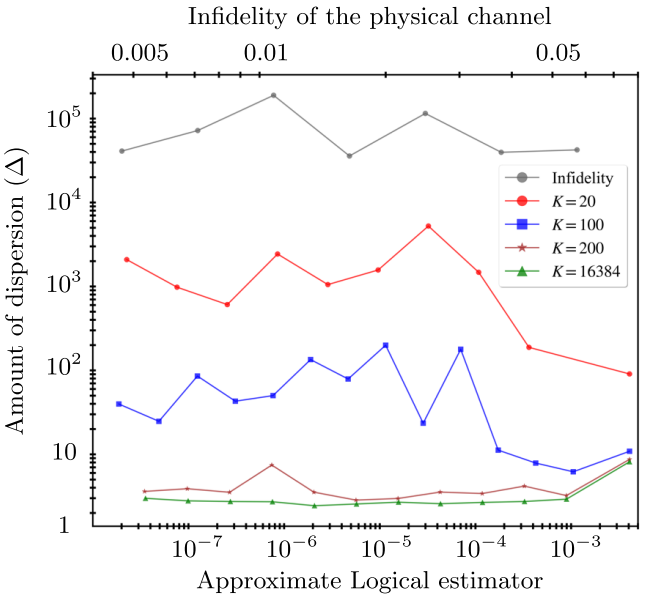


FIG. 3. Accuracy of the logical estimator based on limited NR data, using a level-2 concatenated Steane code for an ensemble of about 15 000 random correlated Pauli channels. The accuracy, quantified by Δ, improves sharply with the number of Pauli error rates (K) extracted using NR. We observe that for $K = 200$, which is about 1.2% of all Pauli error rates on the Steane code block, the accuracy closely matches the logical estimator computed using all NR data, i.e., $K = 4^7$.

B. Code selection

Selecting a quantum error correcting code that has the smallest LER under an existing physical noise process is a crucial step in optimizing resources for fault tolerance. To demonstrate the efficacy of the logical estimator for this problem, we consider an example of an error model and two different error-correcting codes: (i) concatenated Steane code and (ii) concatenated version of a [[7,1,3]] code used in Ref. [2] that we refer to as a *Cyclic code*. The error model is obtained from a Pauli twirl on the i.i.d. application of the CPTP map $\mathcal{E}: \rho \mapsto p_I\rho + \sum_{Q\in\{X,Y,Z\}} p_Q e^{-i\theta Q}\rho e^{i\theta Q}$, where $p_X = r_X(1-r_Z)$, $p_Z = r_Z(1-r_X)$, $p_Y = r_X r_Z$, $p_I = 1-p_X-p_Y-p_Z$ and set a bias specified by $\eta = r_Z/r_X$. Based on Ref. [2], we expect the Steane code to outperform the Cyclic code in one noise regime, and the converse in a different regime. Our tool is successful if it produces a lower value of $\widetilde{p}_u$ for the code with lower logical infidelity, for any noise rate. Last, to compute the logical estimator as well as the LER estimates, we use a bias-adapted minimum-weight decoder that assigns weights η, η, and 1 to each Pauli error of type X, Y, and Z, respectively.

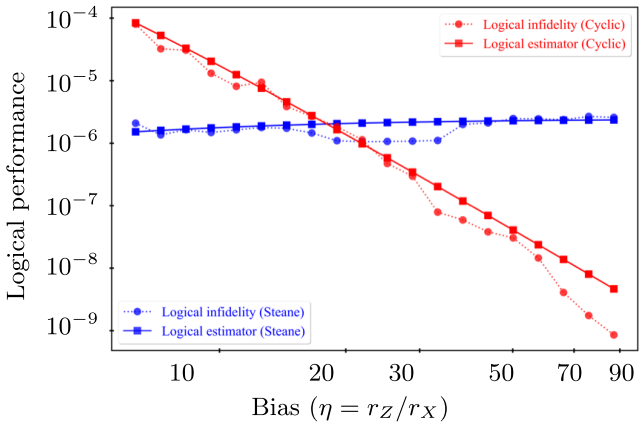


FIG. 4. Using the logical estimator to select an optimal code. The figure demonstrates the use of our tool in selecting an optimal error correcting code under a biased Pauli error model. The choices of codes include level-3 concatenated versions of the Steane code and the cyclic code. While the solid lines depict the values of the logical estimator, the dashed lines correspond to LERs estimated using numerical simulations. We observe that $\widetilde{p}_u$ accurately selects the optimal code for all noise rates.

Figure 4 shows that the logical estimator correctly identifies the optimal code for all values of the physical error rate (bias). It also replicates the functional form of the LER, showing that the performance gain from the Cyclic code over the Steane code increases with the bias.

V. CONCLUSION

We have shown how experimental data from NR, even limited data, can be used to successfully predict the logical performance of FT architectures based on concatenated codes. It can be used to precisely and efficiently estimate the resource overhead required to achieve a target logical error rate [2,52,53] for implementing quantum algorithms. Along with informing the choice of an optimal code for an underlying physical noise process, the logical estimator provides directives for other components in a FT scheme, such as a decoder. Different lookup table decoders can be compared using our logical estimator, similar to the work in Refs. [54–56].

Our scheme relies on RC to yield a Pauli error model, and although in theory this requires twirling with the full Pauli group, it has been observed that a handful of random compilations of the original circuit are sufficient in practice [57,58]. A natural question that follows is whether RC also mitigates the impact of physical noise on the logical qubit. There is no persistent trend across the general class of Markovian noise processes, and in some cases, RC degrades the performance of the code. Developing noise tailoring techniques that guarantee an improvement to the performance as well as predictability is an interesting problem for future research.

Although the methods and techniques presented in the paper address generic noise processes, there are a number of roadblocks in broadening the scope of this study beyond concatenated codes, where the complexity of computing the logical estimator grows exponentially with the size of the code. We have proposed an application for surface codes in Appendix E. While these results are preliminary, they demonstrate that our method may find broader application beyond concatenated codes. Also, further research is needed to extend these ideas to the context of multiple logical qubits.

ACKNOWLEDGMENTS

We thank Daniel Gottesman, Joel Wallman and Stefanie Beale for their valuable inputs. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund. Research was partially sponsored by the ARO and was accomplished under Grant No. W911NF-21-1-0007. S.D.B. acknowledges support from the Australian Research Council (ARC) via the Centre of Excellence in Engineered Quantum Systems (EQuS) Project No. CE170100009.

APPENDIX

The purpose of the Appendix is to aid and enhance the understanding of the main text. It is organized as follows. Appendix A presents a derivation of an expression for the logical fidelity for a generic stabilizer code, as a function of the code properties as well as the parameters of the underlying physical noise process. In Appendix B, we review key ideas from randomized compiling (RC) and noise reconstruction (NR) for understanding our results. Appendix C derives the accuracy of the logical estimator in estimating the logical error rate. In Appendix D, we show how our tool improves the predictability of logical performance under coherent errors. While the abovementioned studies focus on the family of concatenated codes, Appendix E discusses how our studies can be applied to predict the performance of surface codes. Appendices F and G describe techniques used for numerical simulations, including importance sampling to yield accurate estimates of average logical error rates with a reasonable number of syndrome samples.

APPENDIX A: LOGICAL FIDELITY AND CORRECTABLE ERRORS

The average logical channel $\overline{\mathcal{E}}_1$, defined in the main text, summarizes the effect of quantum error correction on a physical noise process $\mathcal{E}_0$ affecting an encoded state $\overline{\rho}$. In this section, we derive a closed form expression for the average logical channel in terms of the physical channel and the error correcting code parameters. Similar derivations have appeared in Refs. [39,40,59]; however, we present ours for the sake of completeness.

The action of the average logical channel on the logical state is

$$\begin{aligned}\overline{\mathcal{E}}_1(\overline{\rho}) &= \sum_s \mathsf{Pr}(s)\mathcal{E}_1^s(\overline{\rho}),\\ &= \sum_s R_s \Pi_s \mathcal{E}_0(\overline{\rho})\Pi_s R_s\\ &= \sum_s \sum_{i,j} \chi_{i,j} R_s \Pi_s P_i \overline{\rho} P_j \Pi_s R_s\\ &= \sum_s \sum_{\substack{i,j\\ s(P_i)=s(P_j):=s}} \chi_{i,j} R_s P_i \overline{\rho} P_j R_s, \end{aligned} \tag{A1}$$

where in the last line we used the fact that $\Pi_s P_i = P_i \Pi_{s\oplus s(P_i)}$. In other words, whenever $s \neq s(P_i)$, the corresponding projector $\Pi_{s\oplus s(P_i)}$ annihilates the encoded state $\overline{\rho}$.

The χ matrix $\overline{\chi}$ of the effective logical channel defined by

$$\mathcal{E}_1(\overline{\rho}) = \sum_{l,m} \overline{\chi}_{lm} \overline{P}_l \overline{\rho} \overline{P}_m, \tag{A2}$$

where $\overline{P}_l$ and $\overline{P}_m$ are logical operators of the code; can be extracted from Eq. (A1).

The total probability of errors successfully corrected by the decoder: $\overline{\chi}_{00}$, can be estimated from the following observation. An error whose syndrome is s is corrected if the net effect of applying the error along with a recovery prescribed by the decoder results in an effective action of a stabilizer. In other words, all the terms in Eq. (A1) where $R_s P_i$ and $P_j R_s$ are stabilizers contribute to $\overline{\chi}_{00}$. So,

$$\overline{\chi}_{0,0} = \sum_{\substack{E,E' \in \mathcal{E}_C \\ s(E)=s(E'),\, \overline{E}=\overline{E}'}} \phi(E)\, \phi^\star(E')\, \chi_{E,E'}, \tag{A3}$$

where $\overline{E}$ is the logical component in the decomposition of E with respect to the Stabilizer group and $\phi(E)$ is specified by $R_{s(E)}E = \phi(E)\, S$ for any Pauli error E and some stabilizer S. The average logical infidelity $\bar{r}$ is then given by $1 - \overline{\chi}_{00}$.

When a Pauli error is not correctable, the effect of applying a recovery yields a logical operator. Hence, in general,

$$\overline{\chi}_{l,m} = \sum_{\substack{E,E' \in \mathcal{E}_C \\ s(E)=s(E'),\, \overline{E}=\overline{E}'}} \phi(E,l)\, \phi^\star(E',m)\, \chi_{E\overline{P}_l, \overline{P}_m E'}, \tag{A4}$$

where $R_{s(E)}\, |E\, \overline{P}_l| = \phi(E,l)\, S\, |\overline{P}_l|$, for $l \in \{0,1,2,3\}$, any Pauli error E and some stabilizer S. Here $|P|$ stands for the bare Pauli without any associated global phase.

APPENDIX B: INTRODUCTION TO RANDOMIZED COMPILING AND NOISE RECONSTRUCTION

Randomized compiling [11] is a noise-tailoring technique that transforms coherent errors into stochastic errors with little to no overhead. The basic idea is to insert Pauli randomizing gates into the layers (called cycles) of a target circuit. To ensure that the circuit depth remains the same, these gates are compiled into the existing ones. Figure 5 illustrates the key steps of Randomized compiling.

It can be shown that averaging over a number of different compilations nullifies the effect of the coherent parts of the noise on the results. Since coherent errors grow faster than stochastic errors, this technique helps mitigate the effect of noise on the output of the circuit.

Noise reconstruction (NR) is a technique introduced in Refs. [12–14] to estimate Pauli error probabilities in a physical noise process. In particular, for a n-qubit physical error channel $\mathcal{E}$, NR is designed to first estimate the Pauli decay rates, followed by post processing to obtain Pauli error probabilities.

The diagonal elements in the Pauli Transfer Matrix [60] of $\mathcal{E}$ are denoted by $\Gamma(\mathcal{E})_{i,i} = \mathrm{tr}[P_i \cdot \mathcal{E}(P_i)]$ for some $1 \leqslant i \leqslant 4^n$. The input P_i can be written in its eigenbasis, i.e., $P_i = \sum_\lambda c_\lambda |\lambda\rangle\langle\lambda|$. In its simplest form, the experimental protocol

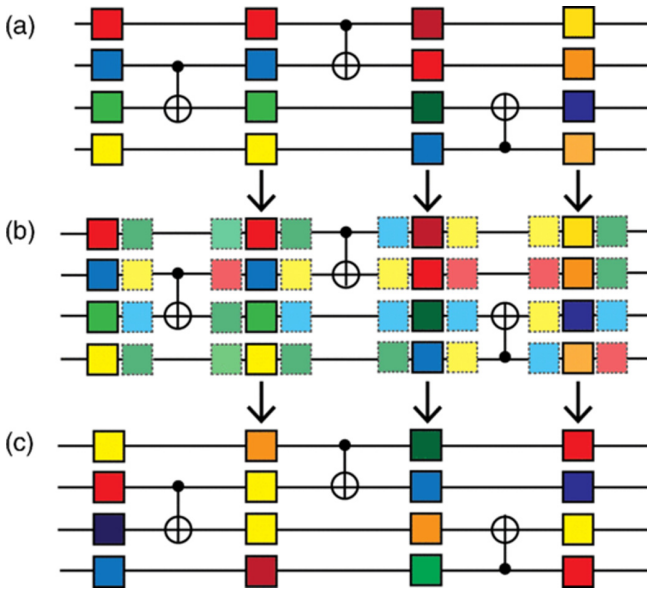


FIG. 5. Randomized compiling. The top panel shows a bare circuit with alternating easy and hard cycles. The middle panel shows insertion of random Pauli gates in between easy and hard cycles. The bottom panel shows that the extra randomization gates are compiled into the existing gates resulting in a random compilation of the bare circuit [11].

for NR can be summarized as follows. For each eigen value c_λ of the Pauli matrix P_i, we prepare an input in $|\lambda\rangle$, apply two sequences, one of m_1 random Pauli gates $Q_1, \ldots, Q_{m_1}$, and likewise, another of m_2 such that their product is the identity operation. At the end of the sequence, we measure the output in the eigenbasis of P_i, implying that the survival probability of the input state is solely affected by a unique Pauli error, and it decays exponentially with the sequence length. In summary, we find two exponential decay laws, one where the survival probability of the input state is given by $A f_i^{m_1} + B$ and other, $A f_i^{m_2} + B$, where $f_i = \mathrm{tr}[P_i \mathcal{E}(P_i)]$ is the decay rate of the Pauli whose eigenstate matches the input, and A is the error associated with faulty state preparation and measurement, a.k.a., SPAM. Solving the above equations for unknowns yields the quantity of interest: f_i. This decay parameter can be used to obtain the probability p_i of error P_i using the Walsh–Hadamard transform.

APPENDIX C: APPROXIMATION QUALITY FOR THE UNCORRECTABLE ERROR PROBABILITY

In this section, we will quantity the accuracy of the approximating the uncorrectable error probability using $\widetilde{p}_u$ for concatenated codes. For simplicity, we will assume that the code blocks in the concatenated code are all identical, and equal to a $[[n, 1, d]]$ quantum error correcting code, with $d \geqslant 3$. Recall that the distance of a level-ℓ concatenated code scales as d^ℓ. We will use $t_\ell = \lfloor (d^\ell + 1)/2 \rfloor$ to denote the Hamming weight of the smallest uncorrectable error. Recall that $\widetilde{p}_u$ is defined recursively as the sum of two quantities: $\widetilde{Q}_1$ and $\widetilde{Q}_2$. We will use δ_ℓ to denote the inaccuracy in computing

p_u for a level-ℓ concatenated code,

$$\delta_\ell = |p_u(\mathcal{C}^\star_{\ell,1}) - \widetilde{p}_u(\mathcal{C}^\star_{\ell,1})|, \tag{C1}$$

and γ_ℓ to denote the inaccuracy in computing Γ,

$$\gamma_\ell = |\widetilde{\Gamma}(\mathcal{C}^\star_\ell) - \Gamma(\mathcal{C}^\star_\ell)|. \tag{C2}$$

Then it follows that

$$\delta_\ell \leqslant n\delta_{\ell-1} + \gamma_\ell. \tag{C3}$$

The most important ingredient in computing δ_ℓ is γ_ℓ, defined in Eq. (C2). For simplicity we will compute γ_ℓ for the i.i.d. depolarizing error model. However, for generic i.i.d. Pauli error models, we can replace the depolarizing rate p in our analysis by the physical infidelity of the single qubit error model, r_0. The extension to correlated Pauli error models remains unclear.

An i.i.d. application of the depolarizing channel on n-qubits can be described by

$$\mathcal{E}(\rho) = \sum_{P\in\mathcal{P}_n} \chi_{P,P} P\,\rho\,P, \quad \text{such that } \chi_{P,P} = (1-p)^{n-|P|}\left(\frac{p}{3}\right)^{|P|}, \tag{C4}$$

where $\mathcal{P}_n$ is the n-qubit Pauli group, $0 \leqslant p \leqslant 1$ is the depolarizing rate, and $|P|$ is the Hamming weight of the Pauli error P. In this case, we will show that

$$\gamma_\ell = \mathcal{O}(n^{\ell-1}p^{t_{\ell-1}+2}), \tag{C5}$$

for a level-ℓ concatenated code.

Combining Eq. (C5) with Eq. (C3), we arrive at an expression for δ_ℓ:

$$\delta_\ell = \mathcal{O}(n^{\ell-1}p^{2+\lfloor (d+1)/2\rfloor}), \tag{C6}$$

where d is the distance of a code block.

In the rest of this section, we will derive Eq. (C5). Recall the following Eq. (C8) that outlines the approximation made by the heuristic to compute $\Gamma(\mathcal{C}^\star_\ell)$:

$$\widetilde{\Gamma}(\mathcal{C}^\star_\ell) = \sum_{E\in\mathcal{E}_\mathcal{C}\setminus\mathbb{I}} \sum_{s(\mathcal{C}_\ell)} \sum_{s(\mathcal{C}^\star_{\ell-1,1})} \cdots \sum_{s(\mathcal{C}^\star_{\ell-1,n})} \mathsf{Pr}[s(\mathcal{C}_\ell)|\hat{\mathcal{E}}_{\ell-1,1}\ldots\hat{\mathcal{E}}_{\ell-1,n}] \prod_{j=1}^n \mathsf{Pr}_\mathcal{D}(E_{\ell-1,j}\,|\,\hat{\mathcal{E}}_{\ell-1,j})\mathsf{Pr}[s(\mathcal{C}^\star_{\ell-1,j})], \tag{C7}$$

$$= \sum_{E\in\mathcal{E}_\mathcal{C}\setminus\mathbb{I}} \prod_{j=1}^n \mathsf{Pr}_\mathcal{D}(E_{\ell-1,j}\,|\,\hat{\mathcal{E}}_{\ell-1,j}). \tag{C8}$$

It involves replacing the knowledge of conditional channels $\mathcal{E}^s_{\ell-1,j}$ by the average channel, $\hat{\mathcal{E}}_{\ell-1,j}$. We will prove the scaling in Eq. (C5) two steps. First, is an observation that

$$\prod_{j=1}^n \mathsf{Pr}_\mathcal{D}(E_{\ell-1,j}|\hat{\mathcal{E}}_{\ell-1,j}) = \mathcal{O}(p^{t_{\ell-1}}). \tag{C9}$$

This follows from the fact that at least one of the errors $E_{\ell-1,j}$ in the error pattern $E_{\ell-1,1}\otimes\ldots\otimes E_{\ell-1,n}$ must be nonidentity. Note that a nonidentity logical error is left as a residual when the decoder for the subsequent lower level fails. Such an event will not occur for errors whose weight is below $t_{\ell-1}$.

Second, by showing that

$$\mathsf{Pr}\big[s(\mathcal{C}_{\ell,i})|\mathcal{E}^{s(\mathcal{C}^\star_{\ell-1,1})}_{\ell-1,1}\ldots\mathcal{E}^{s(\mathcal{C}^\star_{\ell-1,n})}_{\ell-1,n}\big]\prod_{i=1}^n \mathsf{Pr}[s(\mathcal{C}^\star_{\ell-1,j})] = \mathsf{Pr}[s(\mathcal{C}_{\ell,i})|\hat{\mathcal{E}}_{\ell-1,1}\ldots\hat{\mathcal{E}}_{\ell-1,n}]\prod_{i=1}^n \mathsf{Pr}[s(\mathcal{C}^\star_{\ell-1,j})] + \mathcal{O}(n^{\ell-1}p^2). \tag{C10}$$

Recall from the following equation that the average channel $\hat{\mathcal{E}}_{\ell,i}$ is defined recursively in terms of $\hat{\mathcal{E}}_{\ell-1,j}$:

$$\hat{\mathcal{E}}_{\ell-1,i} = \sum_{s(\mathcal{C}_{\ell-1,i})} \mathsf{Pr}[s(\mathcal{C}_{\ell-1,i})]\mathcal{E}^{s(\mathcal{C}_{\ell-1,i})}_{\ell-1,i}[\hat{\mathcal{E}}_{\ell-2,1}\otimes\ldots\otimes\hat{\mathcal{E}}_{\ell-2,n}]. \tag{C11}$$

While the term corresponding to $s(\mathcal{C}_{\ell,i}) = 0$ describes the effect of stabilizers on the input state, the other terms include the effect of nontrivial errors. Note that the nontrivial error E_ℓ has weight at least $t_{\ell-1}$, equal to the weight of the smallest uncorrectable error of the concatenated code $\mathcal{C}^\star_{\ell-1,j}$. Carrying this idea from level $\ell-1$ to level 1, we find

$$\hat{\mathcal{E}}_{\ell,i} = \mathcal{E}^{s(\mathcal{C}_{\ell,i})=0}_{\ell,i}[\hat{\mathcal{E}}_{\ell-1,1}\otimes\ldots\otimes\hat{\mathcal{E}}_{\ell-1,n}] + \mathcal{O}(p^{t_{\ell-1}}), \tag{C12}$$

$$= (\hat{\mathcal{E}}_{\ell-1,1}\otimes\ldots\otimes\hat{\mathcal{E}}_{\ell-1,n}) + \mathcal{O}(p^{t_{\ell-2}}), \tag{C13}$$

$$= (\hat{\mathcal{E}}_{1,1}\otimes\ldots\otimes\hat{\mathcal{E}}_{1,n^{\ell-1}}) + \mathcal{O}(p^{t_1}), \tag{C14}$$

where in Eq. (C13) we have used the fact that the leading contribution to the conditional channel for the trivial syndrome, is the physical channel itself. Equation (C14) describes the recursion until level $\ell = 1$, where $\hat{\mathcal{E}}_{1,j} = \overline{\mathcal{E}}_{1,j}$.

Recall that the conditional channel for an error-syndrome $s(\mathcal{C}^{\star}_{\ell-1,i})$,

$$\mathcal{E}^{s(\mathcal{C}^{\star}_{\ell-1,i})}_{\ell-1,i} = \mathcal{E}^{s(\mathcal{C}_{\ell-1,i})s(\mathcal{C}_{\ell-2,1})\ldots s(\mathcal{C}_{\ell-2,n})\ldots s(\mathcal{C}_{1,1})\ldots s(\mathcal{C}_{1,n^{\ell-1}})}_{\ell-1,i}, \tag{C15}$$

is defined by applying quantum error correction routines corresponding to the syndrome outcomes in the respective code blocks of $\mathcal{C}^{\star}_{\ell-1,i}$. Note that an error is detected (by means of a nontrivial syndrome outcome) in a code block at level ℓ when the decoder operating on the code block at level $\ell-1$ leaves a nontrivial residue. Hence, for a leading order analysis, we will consider conditional channels that correspond to trivial syndromes in all the code blocks except for those at level 1, i.e., $s(\mathcal{C}_{\ell,i}) = 0$ for all $\ell > 1$ in Eq. (C15). In other words, we will consider errors that are corrected within the code blocks in level 1:

$$\mathcal{E}^{s(\mathcal{C}_{\ell-1,i})=0,\, s(\mathcal{C}_{\ell-2,1})=0,\, \ldots,\, s(\mathcal{C}_{\ell-2,n})=0,\, \ldots,\, s(\mathcal{C}_{1,1})\ldots s(\mathcal{C}_{1,n^{\ell-1}})}_{\ell-1,i} = \mathcal{E}^{s(\mathcal{C}_{1,1})}_{1,1} \otimes \ldots \otimes \mathcal{E}^{s(\mathcal{C}_{1,n^{\ell-1}})}_{1,n^{\ell-1}} + \mathcal{O}(p^{t_1}). \tag{C16}$$

Using Eqs. (C14) and (C16), we note that the quality of the approximation in Eq. (C10) can be bounded as follows:

$$\left\{\mathsf{Pr}\left[s(\mathcal{C}_\ell)\middle|\mathcal{E}^{s(\mathcal{C}_{1,1})}_1 \ldots \mathcal{E}^{s(\mathcal{C}_{1,n^{\ell-1}})}_1\right] - \mathsf{Pr}[s(\mathcal{C}_\ell)|\hat{\mathcal{E}}_{1,1} \ldots \hat{\mathcal{E}}_{1,n^{\ell-1}}]\right\} \prod_{j=1}^{n^{\ell-1}} \mathsf{Pr}[s(\mathcal{C}_{1,j})]$$

$$= \mathsf{tr}\left\{\Pi_{s(\mathcal{C}_\ell)} \cdot \left[\left(\mathcal{E}^{s(\mathcal{C}_{1,1})}_1 \otimes \ldots \otimes \mathcal{E}^{s(\mathcal{C}_{1,n^{\ell-1}})}_1\right)(\rho) - (\hat{\mathcal{E}}_{1,1} \otimes \ldots \otimes \hat{\mathcal{E}}_{1,n^{\ell-1}})(\rho)\right]\right\} \prod_{j=1}^{n^{\ell-1}} \mathsf{Pr}[s(\mathcal{C}_{1,j})], \tag{C17}$$

$$= \sum_i \left[\left(\chi^{s(\mathcal{C}_{1,1})}_{1,1} \otimes \ldots \otimes \chi^{s(\mathcal{C}_{1,n^{\ell-1}})}_{1,n^{\ell-1}}\right)_{i,i} - (\hat{\chi}_{1,1} \otimes \ldots \otimes \hat{\chi}_{1,n^{\ell-1}})_{i,i}\right] \mathsf{tr}[\Pi_{s(\mathcal{C}_\ell)} \cdot P_i \rho P_i] \prod_{j=1}^{n^{\ell-1}} \mathsf{Pr}[s(\mathcal{C}_{1,j})], \tag{C18}$$

$$\leqslant n^{\ell-1} \max_{s \in \mathbb{Z}_2^{n-k}} \left|\left|\chi^s_1 - \hat{\chi}_1\right|\right|_\infty \mathsf{Pr}(s), \tag{C19}$$

where χ^s_1 refers to the χ matrix of the conditional channel $\mathcal{E}^s_1$ while $\hat{\chi}_1$ refers to the χ matrix of the average channel $\hat{\mathcal{E}}_1$. In Eq. (C19), we have used the matrix norm $||A||_\infty$ to refer to the maximum absolute value in the matrix.

To establish the scaling in Eq. (C10) it remains to show that

$$\max_{s \in \mathbb{Z}_2^{n-k}} \left|\left|\chi^s_1 - \hat{\chi}_1\right|\right|_\infty \mathsf{Pr}(s) = \mathcal{O}(p^2). \tag{C20}$$

Recall that the effective channel for a given syndrome s: $\mathcal{E}^s_1$, describes the composite effect of the physical noise process and quantum error correction conditioned on the measurement outcome s. Comparing Eq. (A1) to the general form in Eq. (A2), we find an expression similar to Eq. (A4):

$$\left[\chi^s_1\right]_{i,i} = \frac{1}{\mathsf{Pr}(s)} \sum_{\substack{E \in \mathcal{E}_\mathcal{C} \\ s(E)=s}} \chi_{\overline{P}_i E, \overline{P}_i E}. \tag{C21}$$

For the specific case of the depolarizing channel in Eq. (C4) we can express $[\chi^s_1]_{i,i}$, $\mathsf{Pr}(s)$ and $\hat{\chi}_{i,i}$ as polynomials in the depolarizing rate p:

$$\left[\chi^s_1\right]_{i,i} = \frac{1}{\mathsf{Pr}(s)} \sum_{w=1}^{n} A^s_{i,w} (1-p)^{n-w} \left(\frac{p}{3}\right)^w, \tag{C22}$$

$$\mathsf{Pr}(s) = \sum_i \sum_{w=1}^{n} A^s_{i,w} (1-p)^{n-w} \left(\frac{p}{3}\right)^w, \tag{C23}$$

$$\hat{\chi}_{i,i} = \sum_s \sum_{w=0}^{n} A^s_{i,w} (1-p)^{n-w} \left(\frac{p}{3}\right)^w, \tag{C24}$$

where $A^s_{i,w}$ is the number of Pauli errors Q of Hamming weight w on which the action of the decoder leaves a residual logical error $\overline{P}_i$. In other words, $Q = \overline{P}_i R_s S$ where R_s is the recovery operation prescribed by the decoder for the error-syndrome s and S is any stabilizer. We can use two simple facts about errors to simplify the coefficients $A^s_{i,w}$. First, since the only error of Hamming weight zero is the identity which has $s = 0$, we find $A^s_{i,0} = \delta_{s,0}\delta_{i,0}$. Second, since all errors of Hamming weight up to $\lfloor (d-1)/2 \rfloor$ are correctable, we find $A^s_{i,w} = \delta_{i,0} A^s_{0,w}$ for all $w \leqslant \lfloor (d-1)/2 \rfloor$. Using these simplifications,

$$\left[\chi^s_1\right]_{i,i} = \frac{1}{\mathsf{Pr}(s)} \left[(1-p)^{n-1} \left(\frac{p}{3}\right) A^s_{0,1} + \mathcal{O}(n^2 p^2)\right], \tag{C25}$$

$$\mathsf{Pr}(s) = A^s_{0,1} (1-p)^{n-w} \left(\frac{p}{3}\right) + \mathcal{O}(n^2 p^2), \tag{C26}$$

$$\hat{\chi}_{i,i} = \delta_{i,0}(1-p)^n + 3n(1-p)^{n-1} \left(\frac{p}{3}\right) \delta_{i,0} + \mathcal{O}(n^2 p^2). \tag{C27}$$

It is now straightforward to see that Eq. (C20) follows from the above set of equations.

In summary, this section establishes that the approximation used by the heuristic to compute $\widetilde{p}_u(\mathcal{C}^{\star}_{\ell,1})$, is accurate to $\mathcal{O}(n^{\ell+1} p^{2+\lfloor (d+1)/2 \rfloor})$ for the i.i.d. depolarizing physical error model with error rate p. To get a sense for this approximation quality, we can plug in relevant numbers for an i.i.d. Pauli error model and level-2 concatenated Steane code: $p = 10^{-3}, n = 7, \ell = 2, d = 3$. Numerical simulations of quantum error correction yield an estimate of the logical infidelity given by 4.2×10^{-9}. The analytical bound suggests that the logical estimator derived from the our heuristic method agrees with the logical infidelity up to $\mathcal{O}(10^{-11})$. However, the scaling suggests that the heuristic may not be not accurate for large codes in the high noise regime. Nonethe-

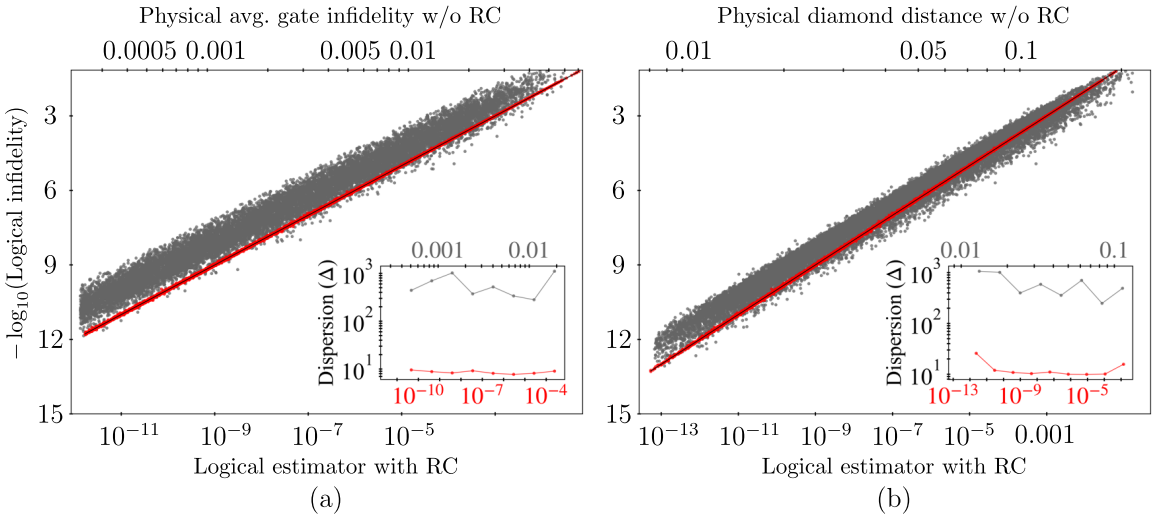


FIG. 6. Predicting the performance of level-2 concatenated Steane code under unitary errors. Panels (a) and (b) compare the predictive powers of our tool (red) with the standard error metrics: infidelity and the diamond distance, respectively, under an ensemble of 16 000 random unitary channels. These are similar to Figs. 2(a) and 2(b) in the main text. The dispersion in the scatter corresponding to a metric (Δ in the insets) is indicative of its predictive power. The gains in predictability offered by our tool is drastic for the above case of unitary errors when compared to CPTP maps.

less, we have strong numerical evidence to support that the logical estimator predicts the functional form of logical infidelity.

APPENDIX D: PREDICTABILITY RESULTS FOR CONCATENATED CODES UNDER COHERENT ERRORS

Numerical results presented in the results and discussion section highlight the predictive power of the tools developed in this work with respect to the standard error metrics, under random CPTP maps. Although CPTP maps encompass a wide range of physical noise processes, our method of generating random CPTP maps does not draw attention to an important class of noise processes—coherent errors—a special case of CPTP maps under which the evolution of a qubit is described by a unitary matrix. They occur due to imperfect control quantum devices and calibration errors [61,62]. Various methods such as dynamical decoupling [63,64], designing pulses using optimal control theory [65], and machine learning approaches [66] are used to mitigate these errors. However, each of these methods have their shortcomings and unitary errors continue to form a major part of the total error budget [67–69]. The methods presented in this paper will be particularly advantageous in these cases.

In this section we highlight the predictive power of our tool, over standard error metrics, under different coherent noise processes. We choose a simple class of coherent errors modeled by an unknown unitary U_i on each physical qubit i, of the form $U = e^{-i\frac{\pi}{2}\delta\hat{n}\cdot\vec{\sigma}}$, where δ is the angle of rotation about an axis $\hat{n}$ on the Bloch sphere. With a slight loss of generality, we will consider n-qubit unitary errors of the form $\otimes_{i=1}^{n} U_i$. We control the noise strength by rotation angles δ_i drawn from a normal distribution of mean and variance equal to μ_δ where $10^{-3} \leqslant \mu_\delta \leqslant 10^{-1}$.

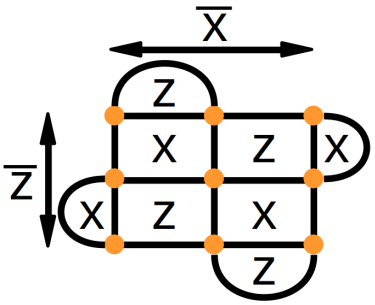
Figure 6 shows that logical error rates vary over several orders of magnitudes across coherent errors with noise strength as measured by standard error metrics such as infidelity and the diamond distance. In contrast, our tools provide an accurate prediction using the logical estimator. Moreover, we observe a drastic gain in in predictability using our tools for this case of unitary errors, when compared to CPTP maps in Fig. 2 of the main text.

APPENDIX E: PREDICTING THE PERFORMANCE OF SURFACE CODES

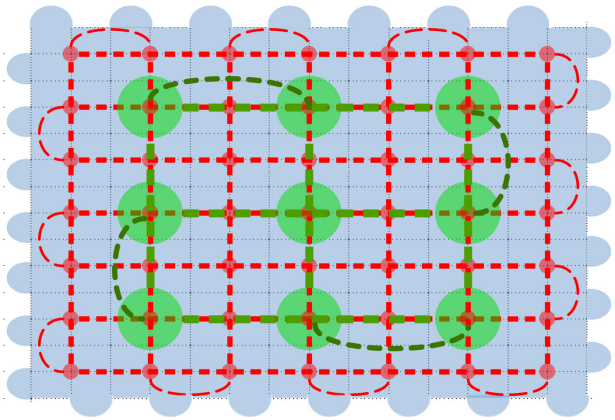
In this section we outline an extension of the techniques to predict the performance of concatenated codes using the logical estimator to surface codes. In summary, we make a crucial ansatz of a concatenated structure for surface codes. This assumption is motivated by a renormalization group-based decoding algorithm developed in Ref. [70] whose threshold is comparable to the optimal decoder. Hence, to define the logical estimator for surface codes, we must first specify a concatenated code structure for it.

For simplicity, we will illustrate the definition of the logical estimator using the square lattice rotated planar code in Ref. [71]. Let us consider the rotated planar code on a $3^\ell \times 3^\ell$ lattice for some integer $\ell > 0$, denoted by $\mathfrak{S}_{3^\ell \times 3^\ell}$. The $3^\ell \times 3^\ell$ square lattice has a self-similar structure in the bulk (ignoring boundaries) where a choice for the unit cell is the 3×3 lattice that specifies the smallest nontrivial code $\mathfrak{S}_{3\times 3}$ shown in Fig. 7(a). Based on this observation, we will construct a concatenated code shown in Fig. 7(b) that will serve as a proxy for the surface code to compute the logical estimator. The surface code on the 3×3 unit cell, $\mathfrak{S}_{3\times 3}$ forms the smallest code block of the concatenated code, and there are ℓ levels in total. Following the notation introduced in the background section, the resulting concatenated code is: $\mathfrak{S}_{3\times 3} \times \ldots \times \ldots \mathfrak{S}_{3\times 3}$.

It is important to iterate that the concatenated code $\mathfrak{S}_{3\times 3} \times \ldots \times \ldots \mathfrak{S}_{3\times 3}$ and the surface code $\mathfrak{S}_{3^\ell \times 3^\ell}$ have



(a) Stabilizer generators for a 3×3 rotated planar code.



(b) Rotated planar code as a concatenated code.

FIG. 7. Panel (a) shows the stabilizer generators for the 3×3 rotated planar code denoted by $\mathfrak{S}_{3\times3}$. Panel (b) depicts an enforced concatenated structure on a rotated planar code. The blue, red, and green lattice depict levels 0, 1, and 2, respectively, of the concatenated code corresponding to the underlying rotated planar code. For concatenated levels $\ell > 1$, the generators in panel (a) are replaced by the corresponding logicals at level $\ell - 1$.

fundamentally different encoding structures. Despite this difference, we use the former concatenated code to define the logical estimator for the latter surface code. Not to our surprise, we find that the logical estimator for the concatenated code $\mathfrak{S}_{3\times3} \times \ldots \times \ldots \mathfrak{S}_{3\times3}$ is significantly different from the average logical fidelity of the corresponding rotated planar surface code $\mathfrak{S}_{3^\ell\times3^\ell}$. In contrast, we find that our heuristic for computing logical estimator for the surface codes plays a crucial role in selecting an optimal code. Recall that in the code selection section, we discussed how the logical estimator is crucial for selecting an optimal concatenated code for an underlying error model. In what follows, we have a similar illustration comparing logical estimators computed for two different surface codes along with their logical infidelities estimated through numerical simulations [71]. The underlying error model is identical to the error model in Fig. 4—the twirl of a convex sum of rotations with a bias η between X and Z errors. We now consider two surface codes, one, $\mathfrak{S}_{9\times12}$—with the ability to correct more X than Z errors, and another, $\mathfrak{S}_{16\times9}$—which corrects more Z than X errors. The logical estimator verifies the expectation that the $\mathfrak{S}_{16\times9}$ performs better as the bias for the Z errors increases relative X errors. Our results are summarized in Fig. 8.

It is important to note that these results are preliminary and a first step toward efficiently and accurately estimating the performance of surface codes. We believe that some of the ideas presented here using the logical estimator would guide the future research in this direction.

APPENDIX F: NUMERICAL SIMULATION DETAILS

The key steps involved in the simulation of an error correcting circuit include encoding, syndrome detection and application of recovery. In our simulations we assume each of these steps to be perfect and model the noise as an explicit step after encoding. Since we deal with coherent errors, we perform a full density matrix simulation. After application of the noise $\mathcal{E}$ to the encoded state $\overline{\rho}$, a syndrome s is sampled with probability $\mathrm{tr}[\Pi_s\mathcal{E}(\overline{\rho})]$, where Π_s is the syndrome projector. The state after syndrome detection is given by

$$\mathcal{E}(\overline{\rho}) \mapsto \overline{\rho}_s = \frac{\Pi_s\mathcal{E}(\overline{\rho})\Pi_s}{\mathrm{tr}[\Pi_s\mathcal{E}(\overline{\rho})]}. \tag{F1}$$

Followed by this, we apply a recovery based on minimum weight decoding and pass the resulting channel to the next level of the concatenated code. At the last level ℓ, we calculate the infidelity of the average logical channel. We report the mean of the infidelity over a large number of syndrome samples. More details about this procedure can be found in Refs. [10,37]. Finally, we employ importance sampling for faster convergence detailed in Appendix G.

So far, we discussed the simulation details for concatenated codes. Simulating surface codes require a slightly different machinery due to the difference in code structure. For deriving the logical error rates in Appendix E, we used a software package called qecsim [71], which is also based on Monte

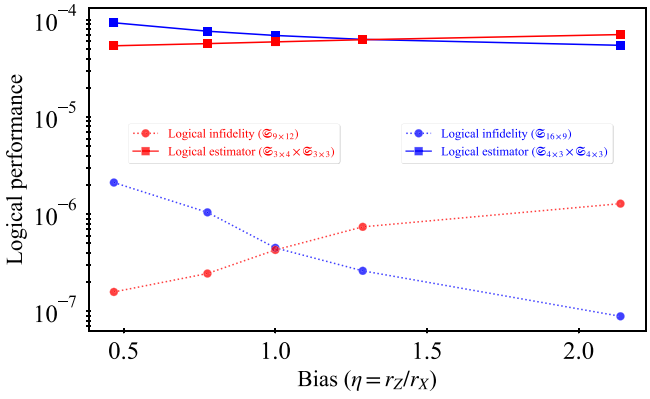


FIG. 8. Using the logical estimator to select optimal surface code. The figure demonstrates the use of our tool in selecting an optimal surface under a biased Pauli error model. The choices of codes include rotated planar code of dimensions 9×12 and 16×9. While the solid lines depict the values of the logical estimator, the dashed lines correspond to LERs estimated using numerical simulations. We observe that $\widetilde{p}_u$ helps select the optimal code for all noise rates.

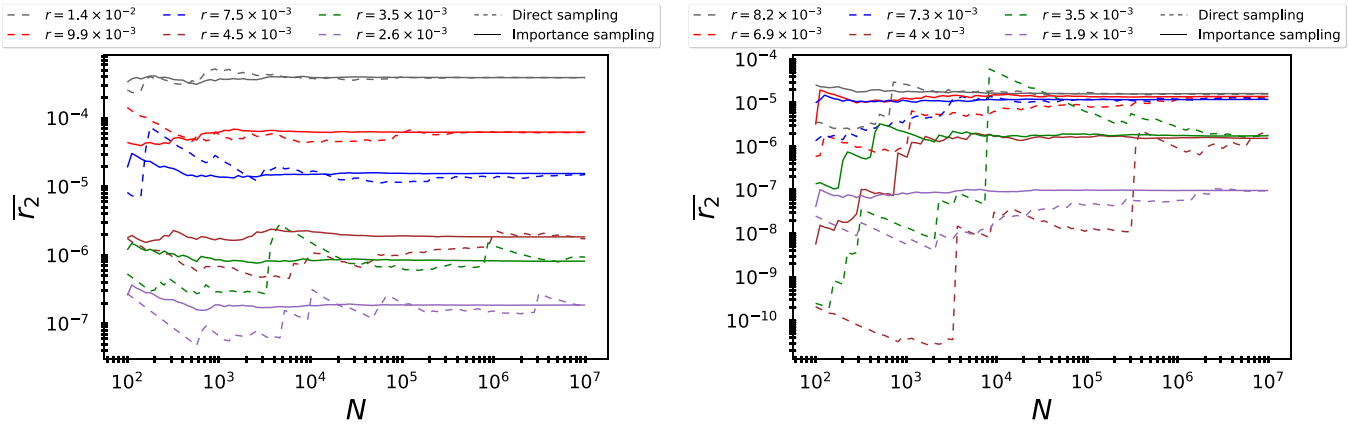


FIG. 9. The figures highlight the rapid convergence rate of the importance sampler as compared to the direct sampler, under CPTP noise processes in Fig. 6(a) and coherent errors in Fig. 6(b). Each trend line in the figures is associated to a physical noise rate. While different colors are used to identify different physical error rates, the solid and dashed lines are used to distinguish between the sampling techniques. Note that while the direct sampler takes a large number to syndrome samples to provide a reliable estimate of $r(\overline{\mathcal{E}}_\ell)$, the importance sampler achieves this task with far lesser syndrome samples. The speedup offered by importance sampling is quite drastic. The case for $r = 4 \times 10^{-3}$ in Fig. 6(b) is a good example. The direct sampler shows signs of convergence around 10^7 syndrome samples, whereas the importance sampler converges with just 10^4 samples. Notice however that with only 10^4 samples, the direct sampler underestimates $r(\overline{\mathcal{E}}_\ell)$ by almost two orders of magnitude.

Carlo simulation of error correcting circuits. The package also assumes perfect encoding, syndrome extraction and recovery application similar to the setting for concatenated codes. We used the minimum weight perfect matching (MWPM) decoder to obtain these results.

APPENDIX G: IMPORTANCE SAMPLING

A straightforward technique to estimate the logical error rate involves sampling syndrome outcomes according to the syndrome probability distribution for a quantum error correcting code and a physical noise process pair. However, there are serious drawbacks to this sampling method, due to the presence of rare syndromes—whose probability is typically less than the inverse number of syndrome samples. A detailed account of this can be found in Ref. [10] and in Sec. 3.3 of Ref. [37]. We briefly review the technique here for completeness.

In summary the average logical error rate is grossly underestimated unless an unreasonably large number of outcomes are sampled. We will resort to an importance sampling technique proposed in Ref. [10], to improve our estimate of the average logical error rate. Previously, similar techniques have also been discussed for Pauli noise processes in Refs. [72,73]. Instead of choosing to sample the syndrome probability distribution, we sample an alternate distribution $Q(s)$, which we will simply refer to as the *importance distribution*. The corresponding sampling methods with $\mathsf{Pr}(s)$ and $Q(s)$ will be referred to as *direct sampling* and *importance sampling*, respectively.

The expression for the average logical error rate estimated by the importance sampler takes the form

$$r(\hat{\mathcal{E}}_\ell) = \sum_{\hat{s}} r(\mathcal{E}_\ell^{\hat{s}}) \frac{\mathsf{Pr}(s)}{Q(s)}, \tag{G1}$$

where $\hat{s}$ is a random syndrome outcome drawn from the importance distribution $Q(s)$. The average estimated by importance sampling coincides with $r(\overline{\mathcal{E}}_\ell)$ which is estimated by the direct sampling technique. The crucial difference between the two sampling techniques is that the variance of the estimated average can be significantly lowered by an appropriate choice for the importance distribution $Q(s)$, which in our case, takes the form

$$Q(s) = \frac{P(s)^{1/k}}{Z}, \tag{G2}$$

where Z is a normalization factor

$$Z = \sum_s P(s)^{1/k}, \tag{G3}$$

and $k \in (0, 1]$ is chosen such that the total probability of nontrivial syndrome outcomes, $s \neq 00\ldots0$, is above a fixed threshold λ_0, i.e.,

$$\sum_{s \neq 00\ldots0} \frac{\mathsf{Pr}(s)^{1/k}}{Z} \geqslant \lambda_0. \tag{G4}$$

Figure 9 shows that our heuristic for the importance distribution provides a rapid convergence to $r(\overline{\mathcal{E}}_\ell)$, when compared to the direct sampling method. Note that the noise processes in these figures are the same as those used to compare the predictive powers of physical error metrics in Fig. 2 of the main text and Fig. 6. Hence, the employment of importance sampling is key to an honest comparison of the predictive powers of the physical error metrics.

[1] P. Aliferis and J. Preskill, Fault-tolerant quantum computation against biased noise, Phys. Rev. A **78**, 052331 (2008).

[2] A. Robertson, C. Granade, S. D. Bartlett, and S. T. Flammia, Tailored Codes for Small Quantum Memories, Phys. Rev. Appl. **8**, 064004 (2017).

[3] D. K. Tuckett, S. D. Bartlett, and S. T. Flammia, Ultrahigh Error Threshold for Surface Codes with Biased Noise, Phys. Rev. Lett. **120**, 050505 (2018).

[4] J. Guillaud and M. Mirrahimi, Repetition Cat Qubits for Fault-Tolerant Quantum Computation, Phys. Rev. X **9**, 041053 (2019).

[5] D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, Fault-Tolerant Thresholds for the Surface Code in Excess of 5% Under Biased Noise, Phys. Rev. Lett. **124**, 130501 (2020).

[6] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, The XZZX surface code, Nat. Commun. **12**, 2172 (2021).

[7] P. Aliferis, D. Gottesman, and J. Preskill, Accuracy threshold for postselected quantum computation, Quantum Info. Comput. **8**, 0181 (2008).

[8] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (London) **549**, 172 (2017).

[9] E. Magesan and P. Cappellaro, Experimentally efficient methods for estimating the performance of quantum measurements, Phys. Rev. A **88**, 022127 (2013).

[10] P. Iyer and D. Poulin, A small quantum computer is needed to optimize fault-tolerant protocols, Quant. Sci. Technol. **3**, 030504 (2018).

[11] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[12] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. **10**, 5347 (2019).

[13] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Trans. Quant. Comput. **1**, 1 (2020).

[14] A. Carignan-Dugas, I. Hincks, E. Ospadov, D. Dahlen, J. Skanes-Norman, S. J. Beale, S. Ferracin, J. Emerson, and J. J. Wallman, The learning and compiled calibration of cycle errors in quantum computing architectures (unpublished).

[15] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Alg. Appl. **10**, 285 (1975).

[16] M. A. Nielsen, The entanglement fidelity and quantum error correction, arXiv:quant-ph/9606012.

[17] B. Schumacher, Sending entanglement through noisy quantum channels, Phys. Rev. A **54**, 2614 (1996).

[18] M. Raginsky, A fidelity measure for quantum channels, Phys. Lett. A **290**, 11 (2001).

[19] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russian Math. Surveys **52**, 1191 (1997).

[20] A. Y. Kitaev, *Quantum Error Correction with Imperfect Gates* (Springer, Boston, MA, 1997), pp. 181–188.

[21] J. Watrous, Semidefinite programs for completely bounded norms, Theory Comput. **5**, 217 (2009).

[22] A. Y. Kitaev, A. Shen, and M. N. Vyalyi, *Classical and Quantum Computation*, Graduate studies in mathematics (American Mathematical Society, Providence, RI, 2002).

[23] G. Gutoski, On a measure of distance for quantum strategies, J. Math. Phys. **53**, 032202 (2012).

[24] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).

[25] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Phys. Rev. A **77**, 012307 (2008).

[26] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Phys. Rev. Lett. **106**, 180504 (2011).

[27] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A **85**, 042311 (2012).

[28] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (2008).

[29] K. M. Svore, D. P. DiVincenzo, and B. M. Terhal, Noise threshold for a fault-tolerant two-dimensional lattice architecture, Quant. Info. Comput. **7**, 297 (2007).

[30] P. Aliferis and J. Preskill, Fibonacci scheme for fault-tolerant quantum computation, Phys. Rev. A **79**, 012332 (2009).

[31] B. M. Terhal and G. Burkard, Fault-tolerant quantum computation for local non-Markovian noise, Phys. Rev. A **71**, 012336 (2005).

[32] D. Bures, An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite w*-algebras, Trans. Amer. Math. Soc. **135**, 199 (1969).

[33] A. Uhlmann, The "transition probability" in the state space of a *-algebra, Rep. Math. Phys. **9**, 273 (1976).

[34] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, New J. Phys. **17**, 113020 (2015).

[35] W. Roga, K. Życzkowski, and F. Mark, Entropic characterization of quantum operations, Int. J. Quant. Inform. **09**, 1031 (2011).

[36] L. Zhang, Entropy, stochastic matrices, and quantum operations, Linear Multilinear Alg. **62**, 396 (2014).

[37] P. Iyer, Une analyse critique de la correction d'erreurs quantiques pour du bruit rèaliste, Ph.D. thesis, Université de Sherbrooke, 2018.

[38] D. E. Gottesman, Stabilizer Codes and Quantum Error Correction, Ph.D. thesis, California Institute of Technology, 1997.

[39] B. Rahn, A. C. Doherty, and H. Mabuchi, Exact performance of concatenated quantum codes, Phys. Rev. A **66**, 032304 (2002).

[40] C. Chamberland, J. Wallman, S. Beale, and R. Laflamme, Hard decoding algorithm for optimizing thresholds under general Markovian noise, Phys. Rev. A **95**, 042332 (2017).

[41] E. Knill and R. Laflamme, Concatenated quantum codes, arXiv:quant-ph/9608012.

[42] T. Jochym-O'Connor and R. Laflamme, Using Concatenated Quantum Codes for Universal Fault-Tolerant Quantum Gates, Phys. Rev. Lett. **112**, 010505 (2014).

[43] M.-H. Hsieh and F. Le Gall, Np-hardness of decoding quantum error-correction codes, Phys. Rev. A **83**, 052331 (2011).

[44] Y. Tomita and K. M. Svore, Low-distance surface codes under realistic quantum noise, Phys. Rev. A **90**, 062320 (2014).

[45] D. Poulin, Optimal and efficient decoding of concatenated quantum block codes, Phys. Rev. A **74**, 052333 (2006).

[46] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, Cambridge, UK, 2013).

[47] D. P. DiVincenzo and P. Aliferis, Effective Fault-Tolerant Quantum Computation with Slow Measurements, Phys. Rev. Lett. **98**, 020501 (2007).

[48] C. Chamberland, P. Iyer, and D. Poulin, Fault-tolerant quantum computing in the Pauli or Clifford frame with slow error diagnostics, Quantum **2**, 43 (2018).

[49] A. M. Steane, A tutorial on quantum error correction, in *Quantum Computers, Algorithms and Chaos*, edited by D. L. Shepelyansky, G. Casati, and P. Zoller (IOS Press, Amsterdam, 2006), pp. 1–32.

[50] R. Raussendorf, Key ideas in quantum error correction, Philos. Trans. R. Soc. A **370**, 4541 (2012).

[51] R. Harper, W. Yu, and S. T. Flammia, Fast estimation of sparse quantum noise, PRX Quant. **2**, 010322 (2021).

[52] P. D. Johnson, J. Romero, J. Olson, Y. Cao, and A. Aspuru-Guzik, Qvector: An algorithm for device-tailored quantum error correction, arXiv:1711.02249.

[53] H. P. Nautrup, N. Delfosse, V. Dunjko, H. J. Briegel, and N. Friis, Optimizing quantum error correction codes with reinforcement learning, Quantum **3**, 215 (2019).

[54] C. Chamberland and P. Ronagh, Deep neural decoders for near term fault-tolerant experiments, Quant. Sci. Technol. **3**, 044002 (2018).

[55] M. Sheth, S. Z. Jafarzadeh, and V. Gheorghiu, Neural ensemble decoding for topological quantum error-correcting codes, Phys. Rev. A **101**, 032338 (2020).

[56] P. Das, C. A. Pattison, S. Manne, D. Carmean, Krysta Svore, M. Qureshi, and N. Delfosse, A scalable decoder micro-architecture for fault-tolerant quantum computing, arXiv:2001.06598.

[57] A. Hashim, R. K. Naik, A. Morvan, Jean-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, Randomized Compiling for Scalable Quantum Computing on a Noisy Superconducting Quantum Processor, Phys. Rev. X **11**, 041039 (2021).

[58] K. J. Satzinger, Y. Liu, A. Smith, C. Knapp *et al.*, Realizing topologically ordered states on a quantum processor, Science **374**, 1237 (2021).

[59] M. Gutiérrez and K. R. Brown, Comparison of a quantum error-correction threshold for exact and approximate errors, Phys. Rev. A **91**, 022335 (2015).

[60] C. J. Wood, J. D. Biamonte, and D. G. Cory, Tensor networks and graphical calculus for open quantum systems, Quant. Info. Comput. **15**, 0759 (2015).

[61] J. True Merrill and K. R. Brown, *Progress in Compensating Pulse Sequences for Quantum Computation*, in Advances in Chemical Physics (John Wiley & Sons, New York, NY, 2014), pp. 241–294.

[62] S. Majumder, L. Andreta de Castro, and K. R. Brown, Real-time calibration with spectator qubits, npj Quant. Info. **6**, 19 (2020).

[63] W. Yang, Z.-Y. Wang, and R.-B. Liu, Preserving qubit coherence by dynamical decoupling, Front. Phys. **6**, 2 (2010).

[64] G. A. Paz-Silva and D. A. Lidar, Optimally combining dynamical decoupling and quantum error correction, Sci. Rep. **3**, 1530 (2013).

[65] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms, J. Magn. Reson. **172**, 296 (2005).

[66] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quant. Info. **5**, 33 (2019).

[67] D. Greenbaum and Z. Dutton, Modeling coherent errors in quantum error correction, Quant. Sci. Technol. **3**, 015007 (2017).

[68] E. Huang, A. C. Doherty, and S. Flammia, Performance of quantum error correction with coherent errors, Phys. Rev. A **99**, 022313 (2019).

[69] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, npj Quant. Info. **4**, 55 (2018).

[70] G. Duclos-Cianci and D. Poulin, A renormalization group decoding algorithm for topological quantum codes, in *Proceedings of the IEEE Information Theory Workshop* (IEEE, Piscataway, NJ, 2010), pp. 1–5.

[71] D. K. Tuckett, Tailoring surface codes: Improvements in quantum error correction with biased noise, Ph.D. thesis, University of Sydney, 2020. qecsim: https://github.com/qecsim/qecsim.

[72] S. Bravyi and A. Vargo, Simulation of rare events in quantum error correction, Phys. Rev. A **88**, 062308 (2013).

[73] C. J. Trout, M. Li, M. Gutiérrez, Y. Wu, S.-T. Wang, L. Duan, and K. R. Brown, Simulating the performance of a distance-3 surface code in a linear ion trap, New J. Phys. **20**, 043038 (2018).