Hard decoding algorithm for optimizing thresholds under general Markovian noise

Christopher Chamberland, Joel Wallman, Stefanie Beale, and Raymond Laflamme

Hard decoding algorithm for optimizing thresholds under general Markovian noise

Christopher Chamberland,[1, *] Joel Wallman,[1, †] Stefanie Beale,[1, ‡] and Raymond Laflamme[1, 2, 3]

[1]*Institute for Quantum Computing and Department of Physics and Astronomy,*
University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
[2]*Perimeter Institute, Waterloo, Ontario, N2L 2Y5, Canada*
[3]*Canadian Institute For Advanced Research, Toronto, Ontario, M5G 1Z8, Canada*

Quantum error correction is instrumental in protecting quantum systems from noise in quantum computing and communication settings. Pauli channels can be efficiently simulated and threshold values for Pauli error rates under a variety of error-correcting codes have been obtained. However, realistic quantum systems can undergo noise processes that differ significantly from Pauli noise. In this paper, we present an efficient hard decoding algorithm for optimizing thresholds and lowering failure rates of an error-correcting code under general completely positive and trace-preserving (i.e., Markovian) noise. We use our hard decoding algorithm to study the performance of several error-correcting codes under various non-Pauli noise models by computing threshold values and failure rates for these codes. We compare the performance of our hard decoding algorithm to decoders optimized for depolarizing noise and show improvements in thresholds and reductions in failure rates by several orders of magnitude. Our hard decoding algorithm can also be adapted to take advantage of a code's non-Pauli transversal gates to further suppress noise. For example, we show that using the transversal gates of the 5-qubit code allows arbitrary rotations around certain axes to be perfectly corrected. Furthermore, we show that Pauli twirling can increase or decrease the threshold depending upon the code properties. Lastly, we show that even if the physical noise model differs slightly from the hypothesized noise model used to determine an optimized decoder, failure rates can still be reduced by applying our hard decoding algorithm.

I. INTRODUCTION

Idealized quantum computers are capable of efficiently factoring very large numbers and simulating quantum systems [1, 2]. However, realistic quantum computers are very sensitive to noise, making their output unreliable. To overcome the effects of noise, methods for error correction and fault-tolerant quantum computation have been developed that allow error rates below some threshold value to be arbitrarily suppressed with polylogarithmic overhead [3–6].

In quantum error-correction schemes, ancilla qubits are entangled with the set of data qubits that we want to protect. Measuring the ancilla qubits produces a (measurement) syndrome that specifies a set of possible errors. A recovery operation is then performed in order to correct the error(s) most likely to have occurred. A decoding algorithm is an algorithm for determining a good recovery operation for an observed syndrome [7]. Note that decoding is elsewhere used to refer to the different process of transferring information from logical to physical qubits.

By the Gottesman-Knill theorem, Pauli channels can be efficiently simulated on a classical computer when the underlying quantum circuits contain only gates from the Clifford group, qubits prepared in computational basis states and measurements that are performed in the computational basis [9]. Simulating non-Pauli channels in fault-tolerant architectures is computationally demanding and has been done only for small codes [10–13]. Assuming perfect error correction, that is, perfect preparations of encoded states and syndrome measurements, Rhan *et al.* introduced a technique to obtain the effective noise channel after performing error correction [14]. The technique, based on the process matrix formalism, is applicable to general completely positive and trace-preserving (CPTP) noise. Rhan *et al.* also showed how to efficiently compute threshold values for concatenated codes under a fixed decoder when each qubit is afflicted by CPTP noise. However, the recovery protocols were suboptimal, that is, they did not achieve the best error suppression.

Concatenated codes are formed by encoding each physical qubit of an error correcting code into another code, and the procedure can be repeated recursively. One could obtain the optimal recovery operator by measuring the error syndrome of the full concatenated code. For a code encoding one logical qubit into n physical qubits, the number of syndroms grows as $2^{cn^{l-1}}$ for l levels (c is a constant that depends on the code) making it computationally unfeasible to keep track of all of them [15].

In order to find optimal recovery operators without having to measure the syndromes of the full concatenated code, soft-decoding algorithms were implemented in Refs. [15, 16] under the perfect error correction assumptions. In Ref. [15], the entire list of probabilities for all possible recoveries conditioned on the observed syn-

drome were retained and passed on to the next level of concatenation in order to implement the optimal recovery operation. The method was applied to study thresholds for depolarizing noise. In a message passing simulation, the total number of syndromes that must be retained grows exponentially with increasing concatenation levels. Therefore, keeping track of all syndromes is inefficient.

In Ref. [17], again using the perfect error correction assumptions, Darmawan and Poulin developed a tensor-network algorithm to compute threshold values for the surface code under arbitrary local noise. The algorithm allowed for the simulation of higher-distance surface codes compared to work done in Refs. [13] and resulted in competitive threshold values for the studied noise models. However, the algorithm does not use non-Pauli transversal gates to its advantage.

In both the soft-decoding and tensor network approaches, the number of syndromes grows exponentially when increasing the codes distance. Therefore, rather than considering all syndrome values, syndromes are sampled from a distribution, leading to statistical fluctuations in the reported thresholds. It is possible that certain unsampled syndromes could change the behavior of the effective noise at the next level in a significant way.

Hard decoding algorithms apply recovery operations independently at each concatenation level based on the measured syndrome (see section II E). Syndrome information from previous levels are not used to update the recovery maps. This will generally result in a suboptimal recovery protocol. However, hard decoding has the advantage of being constant in the code's distance even when considering *all* syndrome measurements, meaning that the required computational resources to compute the recovery operation remain constant even as the code distance increases exponentially with the number of concatenation levels.

In this paper we develop a hard decoding algorithm capable of optimizing threshold values and lowering error rates of an error correcting code compared to traditional hard decoding schemes. If the code has non-Pauli single-qubit transversal gates, our algorithm can lead to even further improvements in the computed threshold values and error rates. By single-qubit transversal gates, we refer to gates that can be implemented by applying single-qubit gates to the qubits in a code block. We assume that error correction can be done perfectly so that additional errors are not introduced during the encoding and decoding protocols.

Our hard decoding algorithm is implemented using the process matrix formalism and can be applied to noise models described by general CPTP maps. For noise models which are not depolarizing, the noise behaviour can change between different concatenation levels. Therefore, for a particular syndrome measurement, the best choice of a recovery operator can differ from level to level.

We show that our hard decoding algorithm can still lead to reduced error rates even when applied to noise that differs slightly from the noise used to optimize the

recovery maps. This indicates that our decoding scheme is robust to perturbative deviations from the assumed noise model.

For codes with transversal Clifford gates, we show that applying a Pauli twirl to a coherent noise channel results in lower threshold values and higher error rates than those obtained when applying our hard decoding algorithm to the original channel. However, if we only optimize over Pauli recovery maps, the Pauli twirl improves the threshold.

The manuscript is structured as follows. We review some preliminary concepts in section II, such as the process matrix formalism (section II A), stabilizer codes (section II B), logical noise resulting from independent and correlated physical noise (section II C and section II D respectively), logical noise in concatenated codes (section II E) and threshold hypersurfaces for general noise models (section II F). In section III we describe our hard decoding algorithm for optimizing threshold values of Markovian noise models. In section IV we describe how to numerically calculate threshold hypersurfaces for both symmetric decoders and decoders obtained from our hard decoding algorithm.

We then present the results of numerical simulations of the 5-qubit code, Steane's 7-qubit code, Shor's 9-qubit code and the surface-17 code. For each code (excluding the surface-17 code), thresholds and infidelities using our hard decoding optimization protocol are computed for amplitude-phase damping noise (section V) and coherent noise (section VI). The concept of infidelity is defined later in the manuscript in eq. (35). For the same noise models, we consider level-1 infidelities of the surface-17 code. We also consider thresholds and infidelities for the 7-qubit code where the noise model was described by two-qubit correlated dephasing noise (section VII). In section VIII, we compute thresholds of the Steane code for a coherent error noise channel and its Pauli twirled counterpart (using both logical Clifford corrections and Pauli only corrections). Lastly, we study the robustness of our decoding algorithm to small unknown perturbations of a noise channel (section IX).

The amplitude-phase damping threshold and infidelity plots can be found in fig. 4. Applying our optimized hard decoding algorithm can more than double thresholds and reduce infidelities by more than two orders of magnitude.

For coherent error noise, threshold plots are given in fig. 5 and infidelity plots are given in fig. 6. For certain rotation axes, our optimized hard decoding algorithm results in errors that are correctable for all rotation angles. In some regimes, infidelities can be reduced by several orders of magnitude. For all the aforementioned noise models, the 5-qubit code consistently achieves higher thresholds and lower error rates compared to the 7 and 9-qubit codes. There is one exception where the Hadamard transform of the 9-qubit code outperforms the 5-qubit code for amplitude-phase damping noise in a small regime. For most studied noise models, the 7-qubit code achieves higher threshold values and lower error rates than the

9-qubit code, with the exception of rotations near the y-axis due to the asymmetries in the Shor codes stabilizer generators.

The threshold plot comparing a coherent error noise channel to its Pauli twirled counterpart is shown in fig. 8. By performing Clifford corrections, the coherent noise channel outperforms its Pauli twirled counterpart for all sampled rotation axes. Lastly, plots showing the robustness of our decoding algorithm to small unknown perturbations are shown in fig. 9. In certain regimes, applying our decoding algorithm results in lower logical failures rates even if the noise model is not perfectly known.

II. STABILIZER CODES AND THE PROCESS MATRIX FORMALISM

We begin by outlining the formalism we use to simulate the performance of concatenated codes under general CPTP noise. We review the process matrix formalism for CPTP maps in section II A and general stabilizer codes (an important class of quantum error-correcting codes) in section II B. We derive expressions for the process matrix conditioned on observing a specific measurement syndrome for independent single-qubit noise in section II C, and for two-qubit correlated noise in section II D. We then define thresholds for a noise model in section II F and define some fixed decoders in section II G.

For clarity, we always use Roman font for operators acting on $\mathbb{C}^d$ (e.g., a unitary U), calligraphic font for a channel acting on the operator space (i.e., a superoperator, e.g., $\mathcal{U}(\rho) = U\rho U^\dagger$) and bold calligraphic font for the matrix representation of a channel (e.g., $\boldsymbol{\mathcal{U}}$).

A. Process matrix formalism for noise at the physical level

A CPTP noise channel $\mathcal{N}$ acting on a state ρ can be written in terms of its Kraus operator decomposition

$$\mathcal{N}(\rho) = \sum_j A_j \rho A_j^\dagger, \tag{1}$$

where $\sum_j A_j^\dagger A_j = I$ for trace-preserving channels [18, 19]. Alternatively, eq. (1) can be rewritten as a matrix product using the process matrix formalism. To do so, note that any matrix $M \in \mathbb{C}^{d \times d}$ can be expanded as

$$M = \sum_i M_i B_i, \tag{2}$$

where $\boldsymbol{B} = \{B_i\}$ is a trace-orthonormal basis for the space of density matrices, that is, $\text{Tr}(B_i^\dagger B_j) = \delta_{i,j}$, and $M_i = \text{Tr}(B_i^\dagger M)$. We exclusively study multi-qubit channels and so set the B_i to be the normalized Pauli matrices, $\boldsymbol{B} = \boldsymbol{\sigma} = (I, X, Y, Z)/\sqrt{2}$ for a single qubit, and $\boldsymbol{B} = \boldsymbol{\sigma}^{\otimes n}$ for n qubits.

We then define a map $|.\rangle\!\rangle : \mathbb{C}^{d \times d} \to \mathbb{C}^{d^2}$ by setting $|B_j\rangle\!\rangle = \mathbf{e}_j$, where $\{\mathbf{e}_j\}$ is the canonical unit basis of $\mathbb{C}^{d^2}$, and extend the map linearly so that

$$|M\rangle\!\rangle = \sum_j M_j |B_j\rangle\!\rangle = \begin{pmatrix} \text{Tr}[B_1^\dagger M] \\ \vdots \\ \text{Tr}[B_{d^2}^\dagger M] \end{pmatrix}. \tag{3}$$

Defining $\langle\!\langle M| = |M\rangle\!\rangle^\dagger$, we have $\langle\!\langle M|N\rangle\!\rangle = \text{Tr}(M^\dagger N)$.

Because quantum channels are linear,

$$\begin{aligned} |\mathcal{N}(\rho)\rangle\!\rangle &= \sum_j \rho_j |\mathcal{N}(B_j)\rangle\!\rangle \\ &= \sum_{i,j} \rho_i |\mathcal{N}(B_j)\rangle\!\rangle\langle\!\langle B_j|B_i\rangle\!\rangle \\ &= \Big(\sum_j |\mathcal{N}(B_j)\rangle\!\rangle\langle\!\langle B_j|\Big)\Big(\sum_i \rho_i |B_i\rangle\!\rangle\Big) \\ &= \boldsymbol{\mathcal{N}}|\rho\rangle\!\rangle \end{aligned} \tag{4}$$

where $\rho_j = \langle\!\langle B_j|\rho\rangle\!\rangle$ are the expansion coefficients of ρ and we used $\langle\!\langle B_j|B_i\rangle\!\rangle = \text{Tr}(B_j^\dagger B_i) = \delta_{j,i}$. We implicitly defined the matrix representation $\boldsymbol{\mathcal{N}}$ of the quantum channel $\mathcal{N}$.

B. Stabilizer codes

We now review stabilizer codes [21]. An $[\![n, k, d]\!]$ stabilizer code C corresponds to the unique subspace $\mathcal{H}_C$ of the n-qubit Hilbert space $\mathcal{H}$ which is the $+1$ eigenspace of an Abelian subgroup $\mathcal{S}$ ($-I \notin \mathcal{S}$) of the n-qubit Pauli group. The stabilizer group $\mathcal{S}$ is generated by a set of $n - k$ mutually-commuting n-qubit Pauli operators $\{g_1, g_2, \ldots, g_{n-k}\}$. Non-identity Pauli operators which commute with all elements of $\mathcal{S}$ act non-trivially (i.e. differs from the identity) on at least d qubits. Defining $N(\mathcal{S})$ as the set of all Pauli operators that commute with $\mathcal{S}$, any Pauli in $N(\mathcal{S}) \setminus \mathcal{S}$ acts as a logical Pauli operator on encoded states. For this paper, we only consider the stabilizer codes in table I, and so set $k = 1$ for the remainder of this section.

We assume that states in $\mathcal{H}_2$ (the Hilbert space of unencoded states) can be encoded in $\mathcal{H}_C$ by an encoding map $\mathcal{E}$ and decoded back to $\mathcal{H}_2$ by the adjoint map $\mathcal{E}^\dagger$. We consider the case where the encoding and decoding protocols can be done perfectly without introducing additional errors, so that $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is encoded to $|\overline{\psi}\rangle = \alpha|\overline{0}\rangle + \beta|\overline{1}\rangle \in \mathcal{H}_C$ and

$$\mathcal{E}(\rho_{in}) = B\rho_{in}B^\dagger \tag{5}$$

with $B = |\overline{0}\rangle\langle 0| + |\overline{1}\rangle\langle 1|$ and some abuse of notation. Since $(1/|\mathcal{S}|)\sum_k \mathcal{S}_k$ acts as the projector onto the codespace ($|\mathcal{S}|$ is the total number of elements in the stabilizer group) and representing $\overline{\tau}$ as the logical version of τ (

$\tau \in \boldsymbol{\sigma}$), we define

$$E_\tau = \mathcal{E}(\tau) = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} S\overline{\tau}, \tag{6}$$

so that E_τ implements τ in $\mathcal{H}_C$ and vanishes elsewhere.

We also assume that syndrome measurements and recovery maps $\mathcal{R}$ are perfect, so that the only errors are due to a noise process $\mathcal{N}$ acting on $\mathcal{H}$. More details about fault-tolerant encoding and measurements can be found in, for example, Refs. [8, 21–24].

Suppose a physical Pauli error E occurs on a system in the encoded state $|\overline{\psi}\rangle \in \mathcal{H}_C$. Measuring the stabilizer generators yields the syndrome $l = l_1 l_2 \ldots l_{n-k}$ where

$$l_i = \begin{cases} 0 & \text{if } [E, g_i] = 0 \\ 1 & \text{if } \{E, g_i\} = 0, \end{cases} \tag{7}$$

$[A, B] = AB - BA$ and $\{A, B\} = AB + BA$. Let Q_l be the set of physical Pauli errors that give the syndrome l, which are all of size $|Q_l| = 2^{2n}/2^{n-k} = 2^{n+k}$. When the syndrome l is measured, a recovery operator $R_l \in Q_l$ is chosen and applied to the state $E|\overline{\psi}\rangle$, returning it to the code space. If $R_l E \in \mathcal{S}$, then the correct state is recovered and the error is removed. Otherwise, $R_l E|\overline{\psi}\rangle$ will differ from $|\overline{\psi}\rangle$ by a logical Pauli operator [21]. The desired outcome of decoding is to find a set of recovery operators which result in the highest probability of recovering the original input state under a given noise model.

As an example, the stabilizer generators for the 3-qubit code protecting against bit-flip errors are $S_1 = Z_1 Z_2$ and $S_2 = Z_2 Z_3$. It can be verified that the errors X_1 and $X_2 X_3$ produce the syndrome $l = 10$ so that $Q_{10} = \{X_1, X_2 X_3\}$. Therefore, if the measured syndrome is $l = 10$, one can either choose X_1 or $X_2 X_3$ to implement the recovery. The particular choice can influence the fidelity of the encoded qubit. For instance, for uncorrelated noise models where single-weight errors are more likely, the better choice for the recovery operator would be $R_{10} = X_1$.

C. Effective process matrix at the logical level

The process of encoding, applying the physical noise $\mathcal{N}$ to the encoded state, implementing the appropriate recovery maps for the measured syndrome l and decoding yields the effective single-qubit channel

$$\mathcal{G}(\mathcal{N}, R_l) = \mathcal{E}^\dagger \circ \mathcal{R}_l \circ \mathcal{N} \circ \mathcal{E}, \tag{8}$$

where $\mathcal{R}_l$ includes the measurement update and the recovery map $R_l \in Q_l$. We now outline how this effective channel can be represented in the process matrix formalism, mostly following Ref. [14] with a straightforward generalization to consider individual syndromes. The states before encoding and after decoding, ρ_{in} and ρ_{out} respectively, are related by

$$|\rho_{\text{out}}\rangle\!\rangle = \boldsymbol{\mathcal{G}}(\mathcal{N}, R_l)|\rho_{\text{in}}\rangle\!\rangle, \tag{9}$$

5-qubit code	Steane code	Z-Shor code	Surface-17 code
XZZXI	IIIZZZZ	ZZIIIIIII	ZIIZIIIII
IXZZX	IZZIIZZ	ZIZIIIIII	IZZIZZIII
XIXZZ	ZIZIZIZ	IIIZZIIII	IIIZZIZZI
ZXIXZ	IIIXXXX	IIIZIZIII	IIIIZIIZ
	IXXIIXX	IIIIIIZZI	XXIXXIIII
	XIXIXIX	IIIIIZIZ	IXXIIIIII
		XXXXXXIII	IIIIXXIXX
		IIIXXXXXX	IIIIIXXI
$\langle C_{\pi/3}, X, Z\rangle$	$\langle H, S\rangle$	$\langle X, Z\rangle$	$\langle X, Z\rangle$

TABLE I: Stabilizer generators (top) and the group $\mathscr{L}$ of single-qubit transversal logical operations (bottom) for the 5-qubit code [25], Steane's 7-qubit code [26], Shor's 9-qubit code [3], and the surface-17 code [13], where H and S are the Hadamard and phase gates respectively, $C_{\pi/3} = \exp[i\pi(X + Y + Z)/3\sqrt{3}] \propto SH$, and $\langle . \rangle$ denotes the group generated by the argument. We also consider the X-Shor code, obtained from the Z-Shor code by mapping $X \leftrightarrow Z$. For each code, the logical operators are $X_L = X^{\otimes n}$ and $Z_L = Z^{\otimes n}$. We only consider the surface-17 code at the first level, as surface codes are not scaled up by concatenation. The surface-17 code is so named as it consists of 9 data qubits and 8 ancilla measurement qubits, and is equivalent to the other 2-D configuration with 9 data qubits in Ref. [13] under the assumption of perfect measurements.

where the process matrix representation of $\boldsymbol{\mathcal{G}}(R_l)$ is

$$\boldsymbol{\mathcal{G}}(\mathcal{N}, R_l) = \sum_{\sigma \in \boldsymbol{\sigma}} |\mathcal{G}(\mathcal{N}, R_l)(\sigma)\rangle\!\rangle\langle\!\langle\sigma| \tag{10}$$

by eq. (4). The entries of the process matrix are

$$\begin{aligned} \boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}, R_l) &= \langle\!\langle\sigma|\mathcal{G}(\mathcal{N}, R_l)(\tau)\rangle\!\rangle \\ &= \langle\!\langle\mathcal{E}(\sigma)|\mathcal{R}_l \circ \mathcal{N}(E_\tau)\rangle\!\rangle \\ &= \langle\!\langle E_\sigma|\mathcal{R}_l \circ \mathcal{N}(E_\tau)\rangle\!\rangle. \end{aligned} \tag{11}$$

By the Born rule, the probability of the syndrome l occurring is $p(l) = \text{Tr}(\mathscr{P}_l\mathcal{N}(\rho_{\text{in}}))$ where the projection operator for the syndrome l is

$$\mathscr{P}_l = \prod_{j=1}^{n-k} \frac{1}{2}(I + (-1)^l g_j). \tag{12}$$

Note that from eq. (6), $E_\tau = \mathscr{P}_0\overline{\tau}$. With the corresponding recovery operator R_l, the transformation on the process matrix can be obtained by implementing the von Neumann-Lüders update rule [27] resulting in

$$\begin{aligned} \boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}, R_l) &= \frac{1}{p(l)}\langle\!\langle E_\sigma|R_l\mathscr{P}_l\mathcal{N}(E_\tau)\mathscr{P}_l^\dagger R_l^\dagger\rangle\!\rangle \\ &= \frac{1}{p(l)}\langle\!\langle\mathscr{P}_l^\dagger R_l^\dagger E_\sigma R_l\mathscr{P}_l|\mathcal{N}(E_\tau)\rangle\!\rangle \end{aligned} \tag{13}$$

Following Ref. [14], eq. (13) can be further simplified by noting that $R_l^\dagger E_\sigma R_l$ is a map from the space projected by $\mathscr{P}_l$ to itself and vanishes elsewhere so that $\mathscr{P}_l^\dagger R_l^\dagger E_\sigma R_l \mathscr{P}_l = R_l^\dagger E_\sigma R_l$. Defining

$$D_\sigma^{(l)} \equiv R_l^\dagger E_\sigma R_l, \tag{14}$$

we arrive at

$$\boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}, R_l) = \frac{1}{p(l)} \langle\!\langle D_\sigma^{(l)} | \mathcal{N}(E_\tau) \rangle\!\rangle. \tag{15}$$

In the remainder of this section we will assume that the noise is uncorrelated, so that it takes the form $\mathcal{N} = \mathcal{N}^{(1)} \otimes \ldots \otimes \mathcal{N}^{(n)}$ where $\mathcal{N}^{(i)}$ is the process matrix for the physical noise acting on qubit i. As in eq. (2), we can expand E_τ and $D_\sigma^{(l)}$ as

$$E_\tau = \sum_{\mu_i \in \tilde{\boldsymbol{B}}} \alpha^\tau_{\{\mu_i\}} \mu_1 \otimes \ldots \otimes \mu_n, \tag{16}$$

$$D_\sigma^{(l)} = \sum_{\nu_i \in \tilde{\boldsymbol{B}}} \beta^\sigma_{\{\nu_i\}}(R_l) \nu_1 \otimes \ldots \otimes \nu_n, \tag{17}$$

where $\tilde{\boldsymbol{B}} \subset \boldsymbol{B}$ only has support over products of the stabilizer group and logical operators from eq. (6). For an operator of the form $U = \pm\mu_1 \otimes \ldots \otimes \mu_n$ and using the notation of Ref. [14], we define the function $\phi(U) = \mu_1 \otimes \ldots \otimes \mu_n$ and $a(U) \in \{0, 1\}$ such that $U = (-1)^{a(U)} \phi(U)$. Substituting eq. (6) into eq. (16) gives

$$\alpha^\tau_{\phi(S\bar\tau)} = \frac{1}{2^{\frac{n}{2}-1}} (-1)^{a(S\bar\tau)}. \tag{18}$$

The α coefficient takes into account the overall sign of the product between elements in the stabilizer group and logical operators (for example, the code with XX and ZZ stabilizers also has $(XX)(ZZ) = -YY$ as a stabilizer). The factor of $\frac{1}{2^{\frac{n}{2}-1}}$ comes from choosing a trace-orthonormal basis.

The β coefficient can be obtained by substituting eq. (6) into eq. (14), commuting R_l to the left, using $R_l^\dagger R_l = I$ and setting the result equal to eq. (17). Defining $\eta(A, B) = \pm 1$ for $AB = \pm BA$, we obtain

$$\beta^\sigma_{\phi(S_k\bar\sigma)}(R_l) = \alpha^\sigma_{\phi(S_k\bar\sigma)} \eta(R_l, S_k) \eta(R_l, \bar\sigma). \tag{19}$$

Therefore, for a particular error syndrome l, picking different recovery operators from the set $Q_l^\dagger$ will, in general, yield different values for the coefficient β. This will, in turn, result in different effective noise dynamics. Closed form expressions for α and β are given in appendix A.

Substituting eq. (17) into eq. (15), we obtain

$$\boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}, R_l) = \frac{1}{p(l)} \sum_{\{\mu_i\},\{\nu_i\}} \beta^\sigma_{\{\nu_i\}}(R_l) \alpha^\tau_{\{\mu_i\}} \prod_{i=1}^n \boldsymbol{\mathcal{N}}^{(i)}_{\nu_i\mu_i}, \tag{20}$$

where the sum is over all elements in the stabilizer group and $\boldsymbol{\mathcal{N}}^{(i)}_{\nu_i\mu_i} = \text{Tr}[\nu_i \mathcal{N}^{(i)}(\mu_i)]$.

The effective noise channel can be obtained by averaging eq. (20) over all syndrome measurements. Defining $\beta^\sigma_{\{\nu_i\}} \equiv \sum_l \beta^\sigma_{\{\nu_i\}}(R_l)$, we have

$$\boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}) = \sum_l p(l) \boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}, R_l)$$

$$= \sum_{\{\mu_i\},\{\nu_i\}} \beta^\sigma_{\{\nu_i\}} \alpha^\tau_{\{\mu_i\}} \prod_{i=1}^n \boldsymbol{\mathcal{N}}^{(i)}_{\nu_i\mu_i}, \tag{21}$$

and we will refer to $\mathcal{G}$ as the effective process matrix for the noise channel $\mathcal{N}$. Note that the normalization factor $1/p(l)$ that appears when implementing the von Neumann-Lüders update rule gets cancelled when averaging over all syndrome measurements. For simplicity, and in the remaining sections of this paper, when referring to process matrices for individual syndrome measurements as in eq. (20), we will omit the normalization factor.

When considering concatenated codes in section II E, it will prove useful to define the coding map Ω^C for a code C as

$$\Omega^C : \mathcal{N} \to \mathcal{G}(\mathcal{N}) = \mathcal{E}^\dagger \circ \mathcal{R} \circ \mathcal{N} \circ \mathcal{E}, \tag{22}$$

where the matrix representation of $\mathcal{G}$ is obtained from eq. (21). Note that in eq. (22), $\mathcal{R}$ includes the measurement update and recovery map averaged over all syndrome measurements. The coding map relates the effective noise dynamics at the logical level resulting from the error correction protocol to the noise dynamics occurring at the physical level.

D. Process matrix for two-qubit correlated noise

In this paper we also consider noise models where nearest-neighbor two-qubit correlations occur. More specifically, we will consider a noise channel of the form

$$\mathcal{N}(\rho^{\otimes n}) = \mathcal{N}^{(1)}(\rho)^{\otimes n} +$$
$$p_2 \Big(\sum_{j=1}^{n-1} \mathcal{N}^{(2)}_{j,j+1}(\rho^{\otimes n}) + \mathcal{N}^{(2)}_{1,n}(\rho^{\otimes n}) \Big), \tag{23}$$

where $\mathcal{N}^{(1)}$ corresponds to local uncorrelated noise and with probability p_2, $\mathcal{N}^{(2)}_{j,j+1}(\rho^{\otimes n}) = Z_j Z_{j+1} \rho^{\otimes n} Z_{j+1} Z_j$ applies phase-flip operators to qubits j and $j + 1$. For noise models of this form, the process matrix describing the effective noise is given by

$$\boldsymbol{\mathcal{G}}_{\sigma\tau}(\mathcal{N}) = \boldsymbol{\mathcal{G}}_{\sigma\tau}([\mathcal{N}^{(1)}]^{\otimes n}) +$$
$$p_2 \sum_{\{\nu_i\}} \beta^\sigma_{\nu_i} \alpha^\tau_{\nu_i} \Big(\sum_{j=1}^{n-1} \boldsymbol{\mathcal{Z}}^{(2)}_{\nu_j\nu_j} \boldsymbol{\mathcal{Z}}^{(2)}_{\nu_{j+1}\nu_{j+1}} + \boldsymbol{\mathcal{Z}}^{(2)}_{\nu_1\nu_1} \boldsymbol{\mathcal{Z}}^{(2)}_{\nu_n\nu_n} \Big), \tag{24}$$

where $\mathcal{Z}(\rho) = Z\rho Z$ in keeping with our standard notation for channels. The contribution from correlated noise appears in the second term of eq. (24).

E. Effective noise channels for concatenated codes

Concatenation is the process of encoding each of the n physical qubits encoded in an inner code C_1 into an outer code C_2. One can go to arbitrary levels of concatenation by recursively applying this procedure.

More formally, we consider an m-qubit code C^{out} with encoding map $\mathcal{E}^{out}$ which will form the outer code, and an n-qubit code C^{in} with encoding map $\mathcal{E}^{in}$ which will form the inner code. The logical qubit ρ_0 is first encoded using C^{out}, and afterwards each of the m qubits are encoded using the code C^{in}. The composite encoding map is given by

$$\tilde{\mathcal{E}} = (\mathcal{E}^{in})^{\otimes m} \circ \mathcal{E}^{out}. \tag{25}$$

Throughout this paper we will implement a hard decoding scheme, which applies a recovery operation independently at each concatenation level [21]. Each code block is thereby corrected based on the inner code. The entire register is then corrected based on the outer code. We denote the mn-qubit code with the effective encoding map $\tilde{\mathcal{E}}$ by $C^{out}(C^{in})$. The procedure for choosing a decoding map for a given noise model described by a CPTP map will be addressed in section III.

Let $\mathcal{G}$ describe the effective dynamics of C^{in} where the physical noise dynamics are described by $\mathcal{N}$. To obtain the effective noise dynamics of $\tilde{\mathcal{G}}$ for the code $C^{out}(C^{in})$, we assume that all n-qubit blocks evolve according to $\mathcal{N}$ so that the mn-qubit code evolves according to

$$\tilde{\mathcal{N}} = \mathcal{N}^{\otimes m}. \tag{26}$$

For convenience, we define $\mathcal{E}_{\mathcal{R}}^{\dagger} \equiv \mathcal{E}^{\dagger} \circ \mathcal{R}$ so that $\mathcal{E}_{\mathcal{R}}^{\dagger}$ includes both the recovery and decoding step. In Ref. [14], it was shown that with the above assumptions $\tilde{\mathcal{G}}$ is given by

$$\tilde{\mathcal{G}} = (\mathcal{E}^{\dagger})_{\mathcal{R}}^{out} \circ \mathcal{G}^{\otimes m} \circ \mathcal{E}^{out}. \tag{27}$$

From eq. (22), the above equation can be written as

$$\tilde{\mathcal{G}} = \Omega^{C^{out}}(\mathcal{G}) = \Omega^{C^{out}}(\Omega^{C^{in}}(\mathcal{N})). \tag{28}$$

For uncorrelated noise, we conclude that the effective channel for the code $C^{out}(C^{in})$ can be computed in the same way that lead to eq. (21) by replacing $\mathcal{N}$ with $\mathcal{G}$ for the code C^{out}. The concatenated code $C^{out}(C^{in})$ can then be described by the composition of maps

$$\Omega^{C^{out}(C^{in})} = \Omega^{C^{out}} \circ \Omega^{C^{in}}. \tag{29}$$

The above equation can be easily generalized to the concatenation of codes in the set $\{C_1, C_2, \ldots, C_n\}$ yielding the map

$$\Omega^{C_1(C_2(\ldots C_n))} = \Omega^{C_1} \circ \Omega^{C_2} \circ \ldots \circ \Omega^{C_n}. \tag{30}$$

For the particular case where the same code $C_i = C$ ($i \in \{1, 2, \ldots, t\}$) is used at t levels of concatenation, we define

$$\boldsymbol{\mathcal{G}}^{(t)}(\mathcal{N}) = \Omega^{C_1(C_2(\ldots C_t))}(\mathcal{N}). \tag{31}$$

For correlated noise as in section II D, we cannot in general write the map for the code $C^{out}(C^{in})$ as a composition of maps for the code C^{out} and C^{in}. However, in this paper we will assume that when the code is concatenated, no correlations occur between different code blocks. Only qubits within each code block undergo correlated noise described by eq. (23). The noise dynamics for each code block of the code C^{out} will thus be described by the effective noise dynamics of eq. (24) and the analysis leading to eq. (30) will also apply in this case. This situation could be realized if the physical qubits in each lowest-level code are contained in individual nodes of a distributed quantum computer and is a good approximation if correlations decay exponentially with the separation between physical qubits.

F. Thresholds for noise models

A fixed noise process $\mathcal{N}$ is correctable by a concatenated code C if successive levels of concatenation eventually remove the error completely for arbitrary input states, that is, if

$$\lim_{t \to \infty} \boldsymbol{\mathcal{G}}^{(t)}(\mathcal{N}) = I_4, \tag{32}$$

where $\boldsymbol{\mathcal{G}}^{(t)}(\mathcal{N})$ is as defined in eq. (31) and I_4 is the 4×4 identity matrix. (Formally, we could also require the error rate to decrease doubly-exponentially when quantified by an appropriate metric [19], however, we do not verify this requirement.)

A threshold for a code is defined relative to an m-parameter noise model, that is, a family $\mathcal{N} = \{\mathcal{N}_p : p \in [0,1]^m\}$ of noise processes such that $\mathcal{N}_0 = \mathcal{I}$. The $\mathcal{N}$-threshold for a code C is the hypersurface of the largest volume in $[0,1]^m$ containing only correctable noise processes and the origin, with the faces of $[0,1]^m$ removed.

The typical behavior of the diagonal components of the process matrix for a 1-parameter noise model is illustrated in fig. 1. The diagonal components converge to one (zero) below (above) threshold, while the off-diagonal components converge to zero.

G. Specific decoders

The effective noise acting on a logical qubit is highly dependent upon both the physical noise processes and the choice of recovery operators for each syndrome (cf. eq. (21)).

One decoder that will be very useful is the symmetric decoder; this decoder associates the measured syndrome
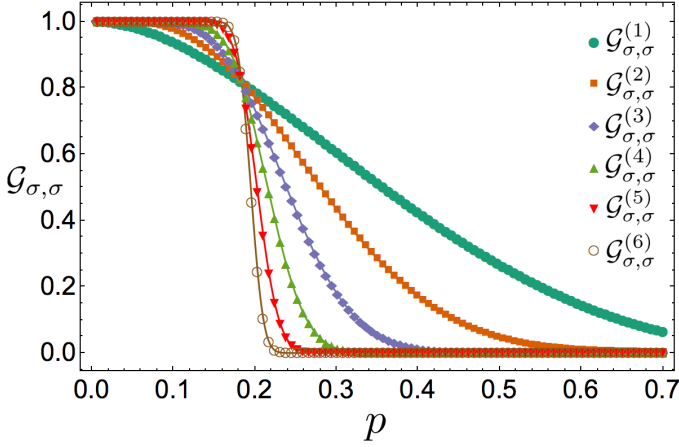
FIG. 1: Representative plot of the smallest diagonal component $\mathcal{G}_{\sigma,\sigma}$ of the process matrix for a noise model parametrized by p. As functions of p, the diagonal components of the process matrix approach a step-function as the number of concatenation levels approaches infinity. The threshold is the smallest value p_{th} such that $\lim_{t\to\infty} \mathcal{G}^{(t)}(\mathcal{N}_p) = I_4$ for all $p \leq p_{th}$.

with the error that acts on the fewest number of qubits and is consistent with the syndrome. If multiple errors acting on equal numbers of qubits are consistent with a syndrome, one is chosen arbitrarily and used each time that syndrome occurs. However the particular choice could affect the threshold value.

The symmetric decoder for the $[\![5,1,3]\!]$ code, for example, associates each syndrome to a unique weight-one Pauli operator. Therefore, all weight-one Pauli operators are corrected. However, one could choose a different decoder for the 5-qubit code. If we consider a noise model where only X-errors occur, a decoder could be chosen which corrects all weight-one and weight-two Pauli X errors. However, this decoder would not be able to correct any Y or Z type Pauli errors. More details will be provided in sections V and VI.

III. HARD DECODING ALGORITHM FOR OPTIMIZING ERROR-CORRECTING CODES

We now present our optimized hard decoding algorithm that determines the choice of recovery operators at each level of concatenation. The goal of the algorithm is to correct the effective noise, that is, to map it to the identity channel $\mathcal{I}$ as quickly as possible. More formally, let ϵ be a pre-metric on the space of CPTP maps, that is, a function such that $\epsilon(\mathcal{N},\mathcal{M}) \geq 0$ with equality if and only if $\mathcal{N} = \mathcal{M}$ and $\epsilon(\mathcal{N},\mathcal{M}) = \epsilon(\mathcal{M},\mathcal{N})$ for all CPTP maps $\mathcal{N}$ and $\mathcal{M}$ (this is a pre-metric as ϵ does not have to satisfy the triangle inequality). The function $\epsilon(\mathcal{N}) := \epsilon(\mathcal{N},\mathcal{I})$ defines an 'error rate'.

We will set $1 - \epsilon(\mathcal{N})$ to be the average gate fidelity to the identity, defined in section III A. This choice signif-

icantly reduces the amount of computational resources required to find the optimal recovery maps. The choice of ϵ may affect the performance of the decoder, however, we defer an investigation of this to future work.

Our hard decoding optimization algorithm selects recovery operations with the goal of minimizing the logical error rate after the recovery operations have been applied. The flowchart given in fig. 2 applies the hard decoding algorithm to a channel $\mathcal{N}$ and determines whether the effective noise will converge to the identity with concatenation. In fig. 2, $\mathcal{M}$ is a general CPTP map, $\mathcal{R}_{sym}$ is a (not necessarily unique) set of recovery operators for symmetric decoding (see section II G), $\mathcal{G}(\mathcal{M},k)$ are the distinct elements of $\{\mathcal{G}(\mathcal{M},R) : R \in \mathcal{R}_{sym}\}$, $m(k)$ is the number of instances of $\mathcal{G}(\mathcal{M},k)$, and L_g is a set of transversal logical operators. The optimized physical recovery maps are

$$R \to T(L_g^\dagger)R \qquad (33)$$

or all $R \in \mathcal{R}_{sym}$, where $g = m(k)\mathcal{G}(\mathcal{M},R)$ and $T(L_g^\dagger)$ denotes the transversal implementation of $L_g^\dagger$. As we discuss in section IV, the choice for $L_g^\dagger$ may not be unique. The action of eq. (33) is equivalent to finding the set $\mathcal{L}_g = \{L_g : g \in \mathscr{G}(\mathcal{M})\}$ of transversal logical operators that minimize

$$\epsilon\left(\sum_{g \in \mathscr{G}(\mathcal{M})} \mathcal{L}_g g\right) \qquad (34)$$

There are 2^{n-1} syndromes, however, step 2 produces only 4, 7, 12, and 67 distinct process matrices for the $[\![5,1,3]\!]$, Steane, Shor, and surface-17 codes respectively, *independently* of the physical noise model. Therefore considering only the distinct $\mathcal{G}(\mathcal{M},k)$ in step 2 reduces the memory and computational requirements by a factor between 4 and 20 for the codes considered in this paper. We could also improve performance by setting the off-diagonal terms to zero when they are sufficiently small and recalculating $\mathscr{G}(\mathcal{M})$ for Pauli channels $\mathcal{M}$. Removing the off-diagonal terms corresponds to performing a Pauli twirl by applying a uniformly-random Pauli operator P to each physical qubit before the noise acts and then applying a logical P at the tth concatenation level. However, this step complicates the algorithm and was not necessary to obtain the results of this paper (it typically sped up computations by a factor between 2 and 10).

As we will discuss in section VIII, the use of a code's non-Pauli transversal gates (note that for any stabilizer code the logical Pauli operators are always transversal) can significantly increase performance. This improvement is obtained when a syndrome measurement results in a logical non-Pauli error with high probability, which can occur even when significantly below threshold. This suggests that using highly symmetric stabilizer codes may provide better performance even at low error rates (in addition to also making non-trivial fault-tolerant computations more viable).
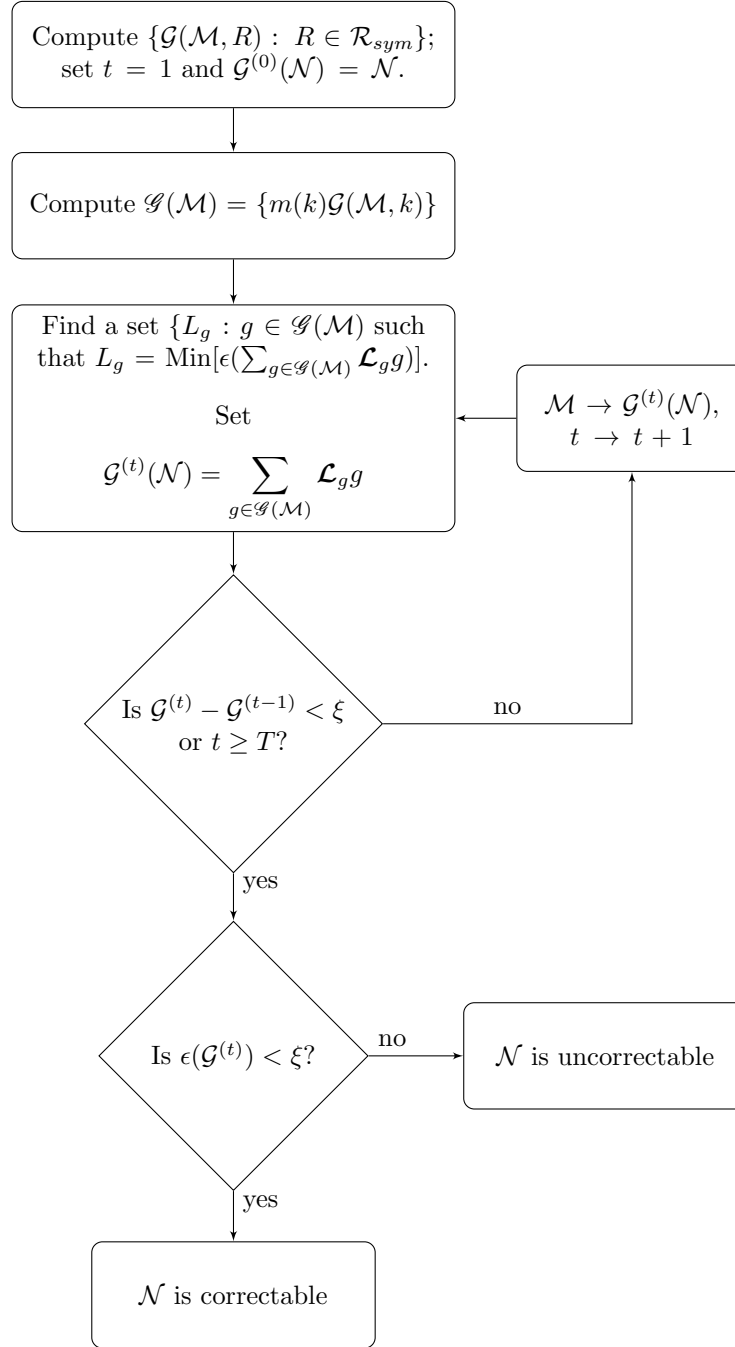
FIG. 2: Method for selection of recovery operations for a fixed code C and noise channel $\mathcal{N}$. The iterative step that calculates the process matrix at level t, $\mathcal{G}^{(t)}(\mathcal{N})$, is equivalent to setting the recovery maps to $R_{opt} = T(L_g^\dagger)R$ for all $R \in \mathcal{R}_{sym}$, where $m(k)\boldsymbol{\mathcal{G}}(\mathcal{M}, R) = g$ and $T(L_g^\dagger)$ denotes the transversal implementation of $L_g^\dagger$. This set may not be unique; see section III B for specific examples of when this can occur. We use the l_∞ norm, that is, the maximum of the absolute values of the entries of a matrix, to test whether the process matrix has converged. However, any matrix norm can be used instead.

The resources required to find the set of recovery maps which optimally correct a noise model $\mathcal{N}$ are efficient in the number of concatenation levels required because our algorithm is independent of the observed syndromes from previous concatenation levels. The largest contribution to the complexity of our scheme comes from computing the β matrix for each syndrome measurement. From eq. (19), there are $3 \times 2^{n-1}$ operations required to compute a β matrix for a particular syndrome value. The factor of 3 comes from computing the commutation re-

lations (encoded by η) between R_l and the code's logical Pauli operators (R_l always commutes with the identity) and the factor of 2^{n-1} comes from verifying the commutation relations between R_l and all elements in the stabilizer group. As there are 2^{n-1} possible syndrome values, $3 \times 2^{2(n-1)}$ operations are required to compute all the β matrices.

A. Infidelity-optimized decoding

The average gate infidelity to the identity (hereafter simply the infidelity),

$$r(\mathcal{N}) = 1 - \int \mathrm{d}\psi \langle \psi | \mathcal{N}(|\psi\rangle\langle\psi|) | \psi \rangle, \qquad (35)$$

is a commonly-used error pre-metric on the space of CPTP maps where the integral is over all pure states according to the unitarily-invariant Fubini-Study metric. The infidelity can be written as

$$r(\mathcal{N}) = \frac{4 - \mathrm{Tr}\,\mathcal{N}}{6} \qquad (36)$$

in the process matrix formalism for a single qubit [28]. The infidelity is particularly convenient for our algorithm because the trace is a linear function of the channel. Consequently, to find a set $\{L_g : g \in \mathscr{G}(\mathcal{M})\}$ that minimizes $\epsilon\left(\sum_{g \in \mathscr{G}(\mathcal{M})} \mathcal{L}_g g\right)$ it is sufficient to maximize

$$\mathrm{Tr}\,\mathcal{L}g \qquad (37)$$

independently for each $g \in \mathscr{G}(\mathcal{M})$, rather than considering all $|\mathscr{G}(\mathcal{M})|^{|\mathscr{L}|}$ possibilities.

B. Resolving ties

There is one important caveat in the implementation of our hard decoding algorithm, namely, there may be multiple sets $\{L_g : g \in \mathscr{G}(\mathcal{M})\}$ that minimize the error in $\epsilon\left(\sum_{g \in \mathscr{G}(\mathcal{M})} \mathcal{L}_g g\right)$.

For example, consider the Steane code with $\mathcal{U}_\theta(\rho) = U_\theta \rho U_\theta^\dagger$ and $U_\theta = \cos\theta I_2 + i\sin\theta X$. Then the only two matrices from step 2 of our algorithm for any value of $\theta \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ are

$$\mathcal{R}_{z1}(\theta) = \frac{7\cos(8\theta) + 25}{32} \mathcal{U}_{\phi/2}$$
$$\mathcal{R}_{z2}(\theta) = \frac{7\sin^2(4\theta)}{16} \mathcal{U}_{-3\theta} \qquad (38)$$

for the trivial syndrome and the syndromes that detect X errors respectively, where

$$\phi = \arctan\left(\frac{(3\cos(4\theta) + \cos(8\theta) + 10)\tan^3(2\theta)}{-3\cos(4\theta) + \cos(8\theta) + 10}\right)$$
$$= 14(\theta^3 + \theta^5) + O(\theta^7). \qquad (39)$$

Similar expressions hold for other values of θ with different signs.

For this example, using all transversal gates significantly improves the recovery, as, for example, $U_{\pm\pi/4}$ (the phase gate around the X-axis) is a transversal gate and so $\mathcal{R}_{z2}(\pi/12)$ can be perfectly recovered. Furthermore, there are two logical gates, namely U_0 and $U_{\pi/4}$, that maximize eq. (37) for $g = \mathcal{R}_{z2}(\pi/24)$, and so the choice is ambiguous. When confronted with such ambiguities, we choose the first logical operator that maximizes eq. (37) (in particular, the identity if it is one of the options). As we will discuss further in section V, this ambiguity due to the ordering of logical operators does arise in practical examples without "fine-tuning" any parameters and it can impact performance.

IV. NUMERICALLY CALCULATING THRESHOLD HYPERSURFACES

We now describe our numerical method for calculating threshold hypersurfaces under symmetric (section II G) and infidelity-optimized decoders (section III). For convenience, we regard a noise channel $\mathcal{N}$ as correctable if there exists some level of concatenation t such that $\mathrm{Tr}\,\mathcal{G}^{(t)} \geq 4 - \xi$, or, equivalently, the infidelity of $\mathcal{G}^{(t)}$ is at most $\xi/6$. The value of ξ was set to 0.01.

A. Symmetric threshold hypersurfaces

The subset of correctable errors for a given noise model is not generically connected. For this reason, a binary search between $p = 0$ and $p = 1$, where p is the noise parameter, is insufficient when calculating a threshold value because this method may miss some uncorrectable regimes. To calculate $p_{sym,thr}(q)$, a threshold value of p (with q fixed) when a symmetric decoder is applied at each level of concatenation, we initialized $p = 0$ and incremented by 0.05 until the noise with $p = p_u$ was uncorrectable, then implemented a binary search between $p = p_u - 0.05$ and $p = p_u$. To find a threshold hypersurface for a noise model with multiple parameters, we iteratively apply this procedure while varying q over a dense mesh.

B. Threshold hypersurfaces for our infidelity-optimized decoder

To calculate threshold values of a code C afflicted by a general CPTP map using our hard decoding algorithm, we follow the procedure illustrated in fig. 3. Here $t(p)$ is the minimum number of concatenation levels required to correct $\mathcal{N}_{pq}$.
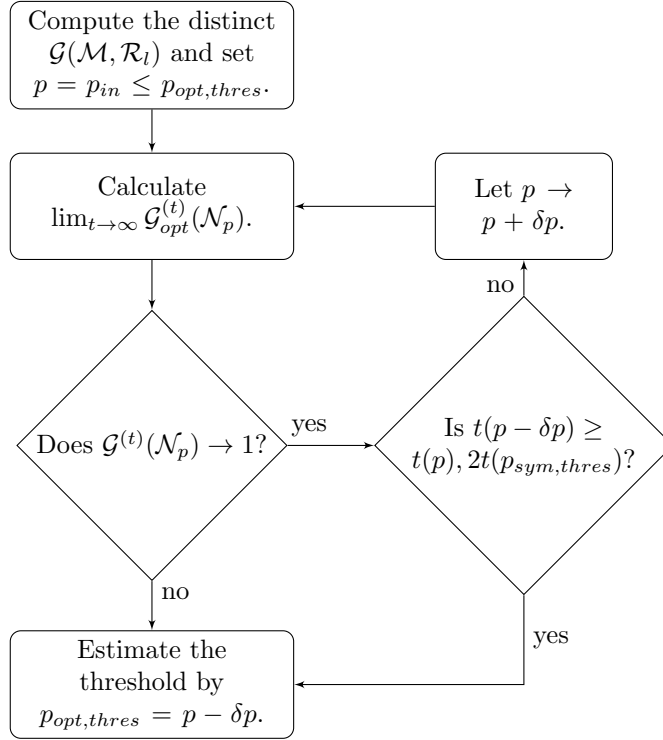
FIG. 3: Method for lower-bounding the threshold for a one-parameter noise model $\mathcal{N}_p$, where $t(p)$ is the minimum number of concatenation levels required to correct $\mathcal{N}_p$. We begin by setting $p_{in} = p_{sym,thres}$ and $\delta p = 0.01$, then repeating with the new lower bound and $\delta p = 0.001$ and finally $\delta p = 0.0001$. To find the threshold hypersurface for an m-parameter noise model, repeat this procedure for $\mathcal{N}_{pq}$ while iterating through a mesh of points, q.

V. THRESHOLDS AND INFIDELITIES FOR AMPLITUDE-PHASE DAMPING

In the remainder of the paper, we show that our hard decoding algorithm leads to significant improvements in threshold values and, in some cases, decreases the infidelity by several orders of magnitude relative to the symmetric decoder. We will also show that the performance of our decoder is robust to perturbations in the noise, so that it can be implemented using the necessarily imperfect knowledge of the noise in an experiment.

In this section we consider a physical noise model consisting of both amplitude and phase damping processes. The amplitude damping channel acts on a two-level system at zero temperature. If the system is in the excited state, then a transition to the ground state occurs with probability p. If the system starts in the ground state, it will remain in the ground state indefinitely. A physical example of this scenario would be the spontaneous emission of a photon in a two-level atom. The Kraus operators for the amplitude damping channel are [19]

$$A_{AD}^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \; A_{AD}^{(1)} = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}. \quad (40)$$

We point out that eq. (40) can be generalized to take into account non-zero temperature effects. In this case,

when the system is in the ground state, there is a non-zero probability of making a transition to the excited state. In [20], the performance of the 5-qubit code, Steane code and non-additive quantum codes was estimated for the generalized amplitude damping channel. However, the methods used did not allow for an exact analysis. In the remainder of this manuscript we will only consider the amplitude-damping channel at zero temperature.

Phase damping arises when a phase kick $\exp(i\theta Z)$ is applied to a qubit with a random angle θ. When θ is sampled from a Gaussian distribution, then the Kraus operators are

$$A_{PD}^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix}, \; A_{PD}^{(1)} = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{pmatrix}, \quad (41)$$

where λ characterizes the width of the distribution of θ. The phase damping channel is also equivalent to the phase-flip channel, that is, applying a Z with probability $\alpha = (1 + \sqrt{1-\lambda})/2$.

Combining the amplitude and phase damping channel, we consider the amplitude-phase damping channel given by

$$\mathcal{N}_{APD}(\rho) = \mathcal{N}_{PD}(\mathcal{N}_{AD}(\rho)) = \mathcal{N}_{AD}(\mathcal{N}_{PD}(\rho)). \quad (42)$$

As the amplitude-phase damping channel contains two parameters (p and λ), the threshold hypersurface will be a curve below which the process matrix is correctable.
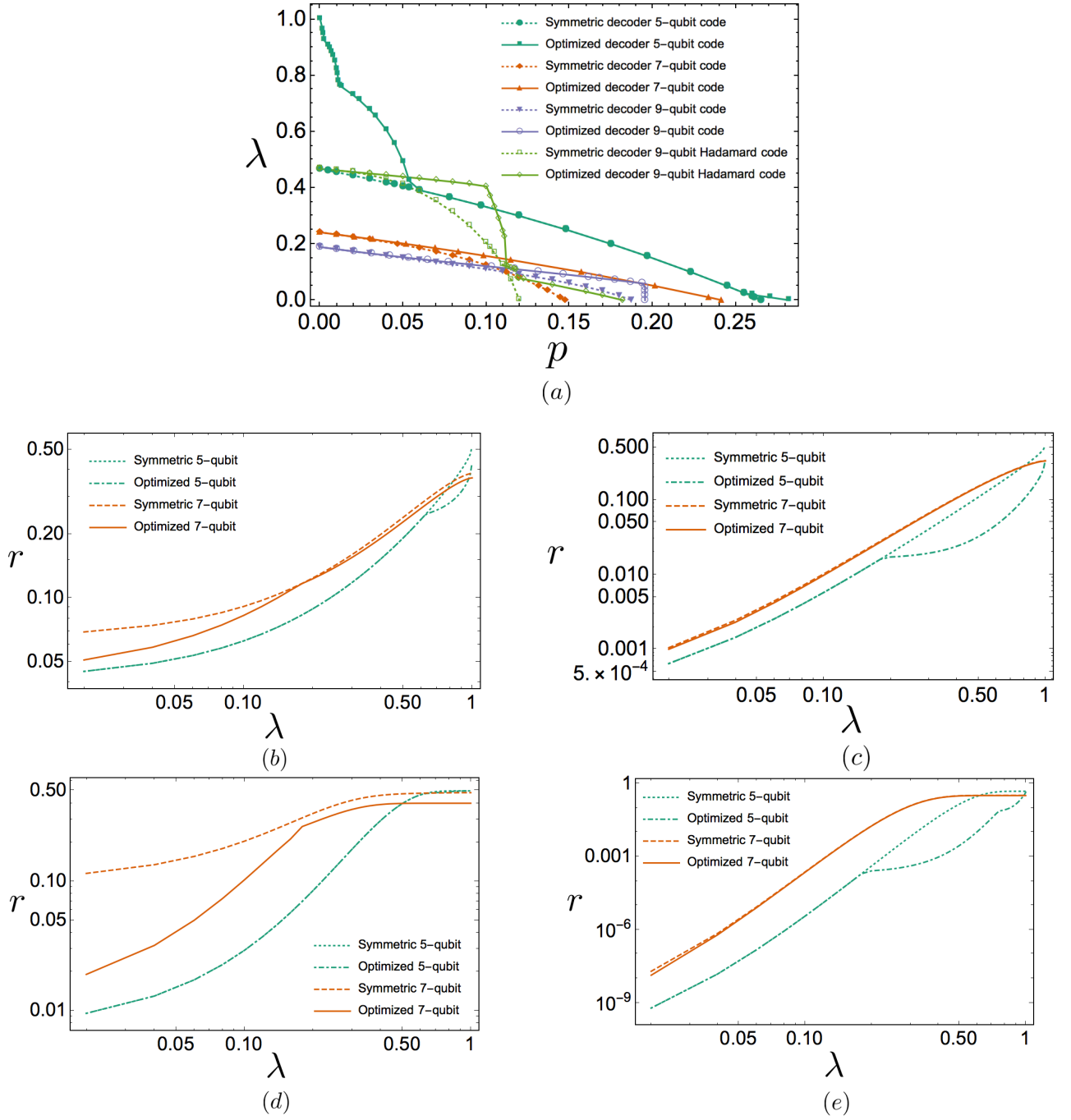
FIG. 4: Threshold curves (a) and infidelities r as functions of the dephasing parameter λ at the first (b), (c) and third (d), (e) concatenation level for the amplitude-phase damping channel under the $[\![5,1,3]\!]$ and Steane codes. The Shor and surface-17 codes behave similarly to the Steane code and so their infidelity curves are not displayed. Our numerics also show that for small p, the 5 and X-Shor codes have the highest thresholds due to lowest number of qubits and asymmetry of stabilizers matching the asymmetry of the noise. The amplitude damping rate is fixed at $p = 0.17$ (b), (d) and $p = 0.01$ (c), (e). The symmetrized and optimized curves overlap when the optimized decoder is the symmetric decoder, however, there are many regimes where the optimized decoder improves the threshold and infidelity. For the $[\![5,1,3]\!]$ code, thresholds using the optimized decoder increase by as much as a factor of 2.14 relative to the symmetric decoder. Infidelities are lowered by as much as 2 orders of magnitude. The curves for the fixed symmetric decoders are all smooth, whereas the curves for the optimized decoders have kinks corresponding to points where the decoder changes to exploit asymmetries in the noise. When λ is large compared to p, the noise is primarily dephasing. The $[\![5,1,3]\!]$ code can exploit this by correcting all single- and two-qubit Z errors for the first t concatenation levels, until the noise becomes unbiased (the kink in (b) corresponds to $t = 1$, and the two in (c) correspond to $t = 2,3$). The X-Shor code can also be biased to correct Pauli-Z errors as it has more X-type stabilizers. All codes also exhibit improved performance for large values of p relative to λ, that is, when the noise is primarily amplitude damping.

The threshold curves and infidelity at the first and third concatenation levels for the $[\![5,1,3]\!]$, Steane, and Shor codes are illustrated in fig. 4. The infidelities for the Shor and surface-17 codes are not shown since they behave similarly to the Steane code. The $[\![5,1,3]\!]$ code generally outperforms all other codes in terms of logical infidelity and thresholds under both optimized and symmetric decoders, except in an intermediate regime where the optimized decoder exploits the asymmetry in the stabilizers of the X-Shor code.

The optimized decoder coincides with the symmetric decoder for each code in some parameter regimes, although only when $p = 0$ for the Steane and Shor codes. However, the optimized decoder often differs significantly from the symmetric decoder, resulting in substantially improved logical infidelities and thresholds. The optimized decoder changes in different parameter regimes to exploit asymmetries in the noise, producing the kinks in the curves in fig. 4. The amplitude-phase damping channel is highly biased towards Z errors for small values of p relative to λ. The optimized decoder exploits this for the $[\![5,1,3]\!]$ code by only correcting Z errors for t levels of concatenation until the noise is approximately symmetric, and then switching to the symmetric decoder, with t increasing as p approaches zero. The X-Shor code also performs better in this regime as it has more X-type stabilizers that detect Z-type errors.

The optimized decoder also results in improved thresholds and logical infidelities for high amplitude damping rates for all codes. The noise is significantly different from Pauli noise in this regime and so decoders constructed under the assumption of Pauli noise will be less likely to identify the correct error compared to decoders optimized for amplitude-phase damping.

As discussed in section III B, multiple sets $\{L_g : g \in \mathscr{G}(\mathcal{M})\}$ maximize eq. (37) for the Steane code with amplitude-phase damping and large values of λ. For example, setting $\lambda = 0.1431$ and choosing the first recovery operator that maximizes eq. (37) gives a threshold of $p_{th} = 0.1032$. However, searching all tuples $\{L_g : g \in \mathscr{G}(\mathcal{M})\}$ that maximize eq. (37) (where the degeneracy only occurs at the first level) for the same value of λ gives a higher threshold of $p_{th} = 0.1150$.

VI. THRESHOLDS FOR COHERENT ERRORS

In this section we illustrate the behavior of coherent errors under error correction. We consider a coherent error noise model where every qubit undergoes a rotation by an unknown angle θ about an axis of rotation $\hat{n}$. The coherent noise channel can thus be written as

$$\mathcal{N}_{\theta,\phi,\gamma}(\rho) = e^{i\theta\hat{n}\cdot\vec{\sigma}} \rho e^{-i\theta\hat{n}\cdot\vec{\sigma}}, \qquad (43)$$

where $\hat{n} = (\sin\phi\cos\gamma, \sin\phi\sin\gamma, \cos\phi)$. We obtain the threshold hypersurface for the noise model

$$\mathscr{N} = \{\mathcal{N}_{\theta,\phi,\gamma} : \theta \in [0, 2\pi], \gamma, \phi \in [0, \pi]\}, \qquad (44)$$

by fixing γ and ϕ and obtaining the threshold for θ.

The threshold hypersurfaces for the $[\![5,1,3]\!]$, Steane, and Shor codes are illustrated as contour plots in fig. 5. The infidelities at the first and third concatenation levels for $[\![5,1,3]\!]$, Steane, Shor, and surface-17 (1st level only) codes are plotted in fig. 6.

Unlike with amplitude-phase damping noise, the optimized decoder strictly outperforms the symmetric decoder for all rotation axes. With the exception of the Shor code, the threshold hypersurfaces are relatively flat under symmetric decoding, that is, the threshold rotation angle is relatively independent of the rotation axis. The optimized decoder breaks this, giving larger threshold angles for different axes, particularly for rotations about an eigenbasis of the transversal Clifford gates listed in table I. In particular, the $[\![5,1,3]\!]$ code can correct any rotation about axes $\hat{n}$ that are close to $(\pm, \pm, \pm)/\sqrt{3}$. The Steane code can correct any rotation about the Pauli axes except for angles close to odd integer multiples of $\pi/4$. The performance of the Shor code is generally only modestly improved by the optimized decoder, largely because the Shor code has no transversal non-Pauli gates. However, the optimized decoder is able to exploit the asymmetries in the stabilizer generators to increase the threshold for rotations near the y axis by more than a factor of 3.

The improved threshold angles are reflected in the orders-of-magnitude reduction in the logical infidelities in fig. 6. The infidelities are periodic because transversal gates can be used to counteract the unitary noise. However, as discussed in section VIII, the transversal gates are useful even when the action of the noise on the codespace is far from a transversal gate. For the $[\![5,1,3]\!]$, Steane and Z-Shor codes, the infidelities are the infidelity at the first and third concatenation levels for rotations about the x axis, while for the surface-17 code the infidelities are at the first level for rotations about the x, y and $\frac{1}{\sqrt{3}}(1,1,1)$ axes. For the $[\![5,1,3]\!]$, Steane and Shor codes, the threshold values of θ correspond exactly to the cross-over points between the level-1 and level-3 curves. At the third level, the infidelity is greatly suppressed below threshold and increased above threshold. The optimized infidelity curves are lower (in some cases by several orders of magnitude) than the infidelities arising by applying the symmetric decoder at all levels.

For the $[\![5,1,3]\!]$ code, the only uncorrectable values of θ are odd integer multiples of $\pi/4$. The surface-17 code has a higher threshold against Y errors than against X (or Z) errors. The surface-17 code treats X and Z errors symmetrically. However, since the X and Z stabilizer generators have support on different qubits, error rates resulting from Y errors will differ from error rates resulting from X and Z errors.

The optimized decoding algorithm gives the greatest improvements for codes with transversal non-Pauli gates, namely, the $[\![5,1,3]\!]$ and Steane codes.
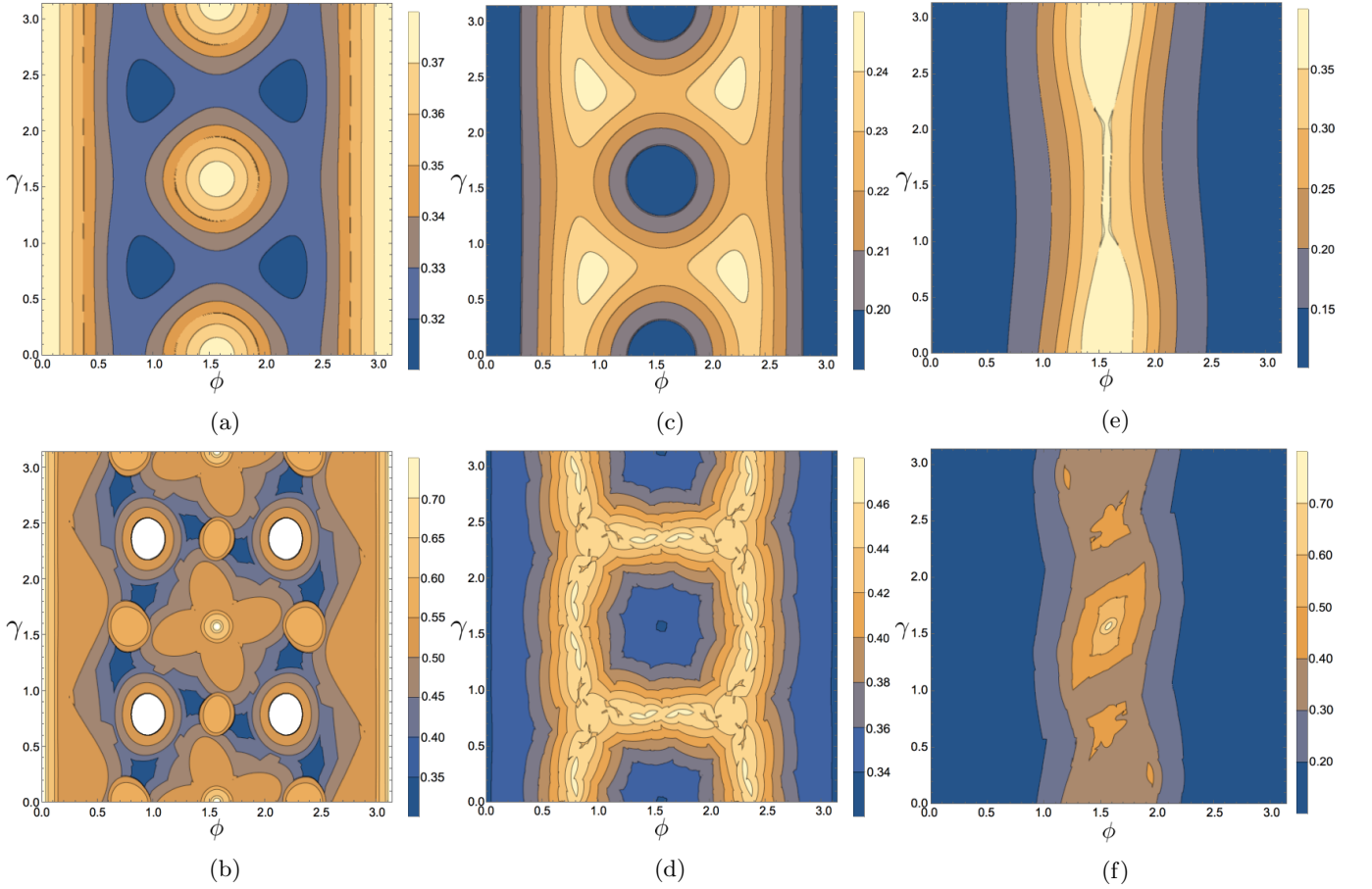
FIG. 5: Contour plots representing hypersurfaces of the threshold value of θ for rotations around the axis $\hat{n} = (\sin\phi\cos\gamma, \sin\phi\sin\gamma, \cos\phi)$ for (a) the $[\![5,1,3]\!]$ code, (c) the Steane code and (e) the Shor code using the symmetric decoder and (b) the $[\![5,1,3]\!]$ code, (d) the Steane code and (f) the Shor code using optimized decoding. The optimized decoder uses transversal gates to improve the threshold, particularly when the rotation is around an eigenbasis of a transversal Clifford gate (see table I). In particular, the $[\![5,1,3]\!]$ code has a transversal $\pi/3$ rotation around $\hat{n}_C = (1,1,1)/\sqrt{3}$ (i.e., $\gamma = \pi/4$, $\phi = \pi/3$), which enables the optimized decoder to correct *arbitrary* rotations around axes close to $\hat{n}_C$, illustrated by the white circular regions in (b). For Steane's code, the lightest colored regions in (d) corresponds to threshold angles $\theta_{th} \approx 0.46$ compared to $\theta_{th} \approx 0.24$ in (c), an improvement by almost a factor of 2. The Shor code has no transversal non-Pauli gates and so the improvements from the optimized decoder are not as substantial. However, for rotations near the y-axis, the Shor code outperforms the Steane code by a factor of at most 2.3.

VII. CORRELATED NOISE CHANNEL

In this section, we study the effect of correlated noise on the logical noise in Steane's code. The correlated noise we consider consists of local depolarizing noise and two-qubit correlated dephasing errors to all adjacent pairs. The composite noise channel maps an n-qubit state ρ

$$\mathcal{N}(\rho) = (1-q)\mathcal{D}_p^{\otimes n}(\rho) + \frac{q}{n}\sum_{j=1}^{n}\mathcal{Z}_{j,j+_n 1}^{(2)}(\rho), \qquad (45)$$

where $j +_n 1 = j + 1$ if $j < n$ and 1 otherwise (that is, we consider the qubits to be in a ring),

$$\mathcal{D}_p(\tau) = (1-p)\tau + \frac{p}{3}(X\tau X + Y\tau Y + Z\tau Z) \qquad (46)$$

is the depolarizing channel acting on a single-qubit state τ, and $\mathcal{Z}_{j,j+1}(\rho) = Z_j Z_{j+1}\rho Z_{j+1} Z_j$ applies phase-flip operators to qubits j and $j + 1$. The logical process matrix can be computed for the noise model in eq. (45) by eq. (24).

The threshold and infidelity at the first and third concatenation levels of Steane's code are illustrated in fig. 7. In the small p regime, the noise is dominated by the two-qubit correlated dephasing contribution. The optimized decoder corrects a larger amount of Z errors at the first
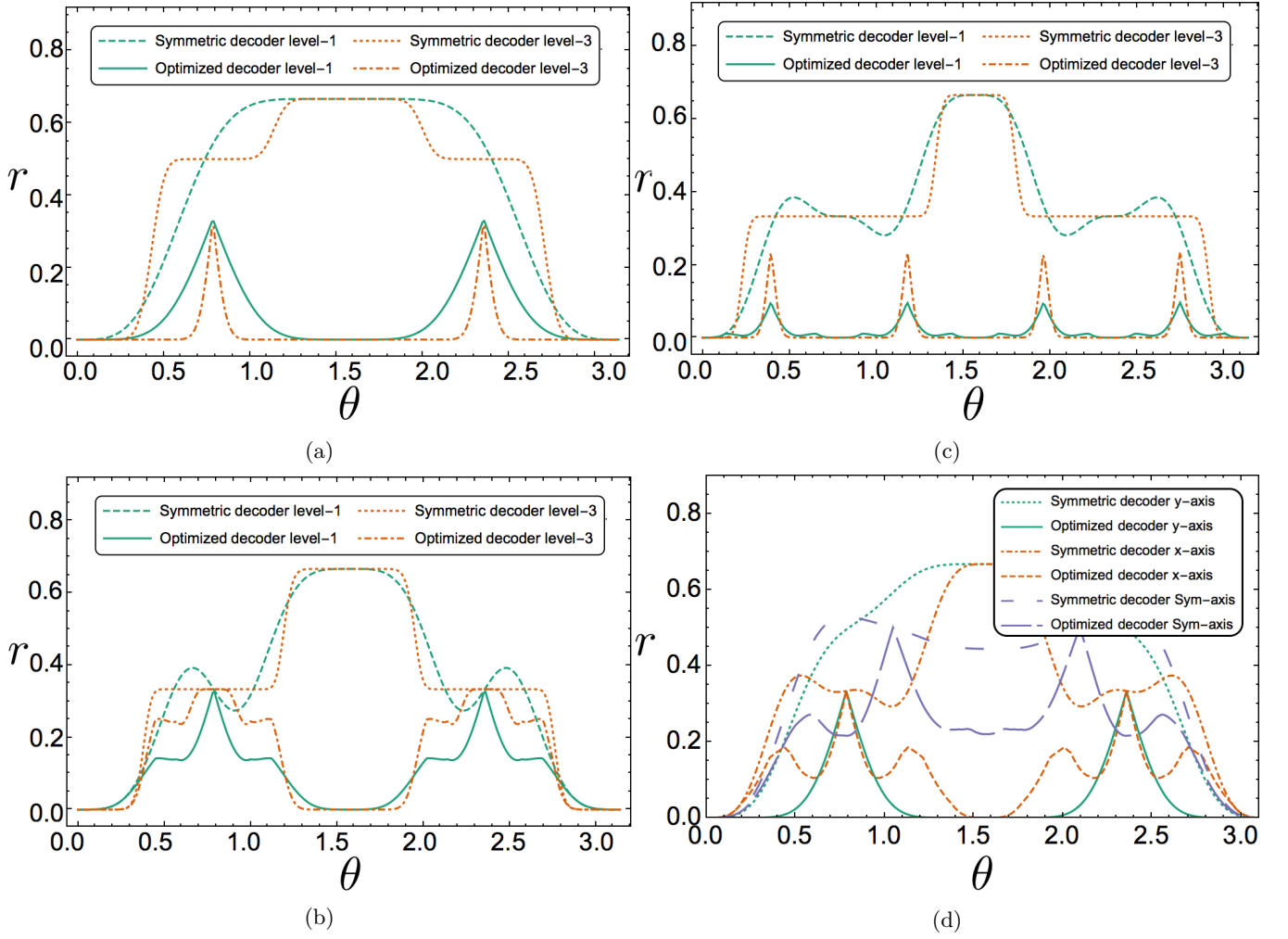
(a)

(c)

(b)

(d)

FIG. 6: In (a), (b) and (c) infidelities r of the $[\![5,1,3]\!]$ code, Shor code and Steane code are plotted at the first and third levels for a rotation about the x-axis. In (a), the optimized infidelity curves are peaked at the code's threshold value $\theta_{th} = \pi/4$. In (b), the peaks of the optimized infidelity curves are centred slightly above the code's threshold value $\theta_{th} = 0.3396$. However, the optimized level-3 infidelity curve intersects the optimized level-1 curve at the threshold value as expected. In (c), the optimized level-1 and level-3 infidelity curves intersect at the threshold value $\theta_{th} = 0.3692$. The peaks of the infidelity curves occur at $\theta = \pi/4$ due to the codes symmetry. In (d), infidelity plots of the surface-17 at the first concatenation level are shown for a rotation about the y-axis, x-axis and the $(1,1,1)/\sqrt{3}$ axis. It can be seen that the infidelity is lowest for rotations about the y-axis. In all 4 plots, it can be seen that applying our hard decoding algorithm reduces the infidelities by, in some cases, orders of magnitude compared to the symmetric decoder.

few levels by breaking the symmetry in the syndrome measurements. At higher levels, the decoder corrects in a more symmetric fashion in order to remove the remaining Pauli errors. This improved performance is also illustrated in the reduced optimized logical infidelities shown in fig. 7 (b) as a function of q with $p = 0.003$.

There is an intermediate regime where the local depolarizing noise contribution becomes more relevant, leading to a decrease in the threshold value for q. However, the optimized threshold is still noticeably larger than the symmetric decoder threshold. Finally, when the local depolarizing noise is the dominant source of noise, our

optimization algorithm chooses recovery maps consistent with the standard CSS decoder. The standard CSS decoder yields a slightly larger p threshold value compared to the symmetric decoder when $q = 0$.

VIII. THE EFFECT OF PAULI TWIRLING ON THRESHOLDS AND THE BENEFITS OF USING TRANSVERSAL OPERATIONS

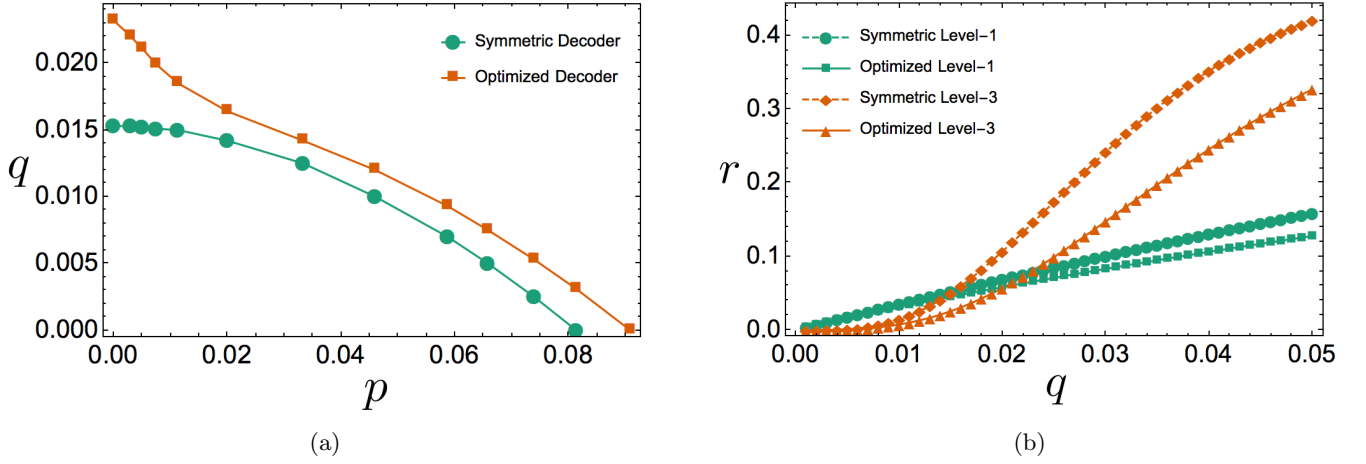In sections V to VII, we showed that our hard decoding optimization algorithm could improve threshold values

(a)   (b)

FIG. 7: (a) Threshold curves and (b) infidelities r at the first and third concatenation level (with fixed $p = 0.003$) for the $[\![7, 1, 3]\!]$ code. Two-qubit correlated dephasing occurs with probability q and depolarizing noise occurs with probability $1 - q$, with a depolarizing noise parameter p (see eqs. (45) and (46)). For small values of p, the Z errors arising from the two-qubit correlations dominate the noise. Applying our optimized hard decoding algorithm in this regime yields a threshold of $q_{th} = 0.0232$. The contribution from depolarizing noise increases with p until the noise is predominantly depolarizing. In this regime, the optimized decoder implements the standard CSS decoder at all levels. When $q = 0$, the noise is purely depolarizing and $p_{th} = 0.0908$. For all values of p, the threshold q_{th} obtained by implementing our optimized decoder is larger than the threshold obtained by implementing the symmetric decoder. The level-1 and level-3 infidelity curves intersect near the respective thresholds for $p = 0.003$, namely, $q_{th} = 0.0153$ and $q_{th} = 0.0220$ for the symmetric and optimized decoders respectively.

by more than a factor of 2 for amplitude-phase damping noise. For coherent noise there where certain rotation axes where the noise was correctable for arbitrary rotation angles. Infidelities were reduced by orders of magnitudes in certain regimes. The amplitude-phase damping and coherent noise models are both non-Pauli. Performing a Pauli twirl on a noise channel $\mathcal{N}$ (that is, conjugating it by a uniformly random Pauli channel) maps it to a channel $\mathcal{T}(\mathcal{N})$ that is a Pauli channel and so has a diagonal matrix representation with respect to the Pauli basis [12, 29]. In [11, 12], the effective noise at the first level for the amplitude damping channel was found to be in good agreement to a Pauli twirled approximation of the channel.

However, we now show that performing a Pauli twirl on coherent noise and using the Steane code can either reduce or increase threshold values, depending on the particular recovery protocol. We also illustrate the improvements obtained by using all transversal gates in the decoding algorithm, instead of just the Pauli gates. Threshold curves for rotations about $(\frac{1}{\sqrt{2}} \sin \phi, \frac{1}{\sqrt{2}} \sin \phi, \cos \phi)$ by an angle θ under three different decoding schemes for the Steane code are presented in fig. 8. The three schemes we consider are : 1) our optimized decoding algorithm applied to the twirled noise; 2) our optimized decoding algorithm applied to the bare noise using all transversal gates; and 3) our optimized decoding algorithm applied to the bare noise using only transversal Pauli gates. Using transversal Clifford gates in our recovery protocol gives the largest threshold values for all
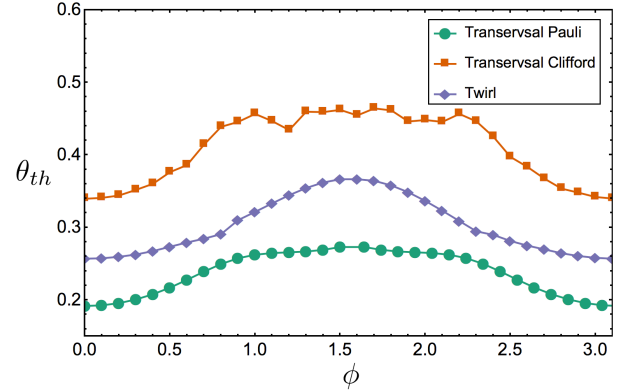

FIG. 8: Threshold curves for unitary rotations about $(\frac{1}{\sqrt{2}} \sin \phi, \frac{1}{\sqrt{2}} \sin \phi, \cos \phi)$ by an angle θ under three different decoding schemes for the Steane code. For all values of ϕ, an improvement by as much as a factor of 1.7 in the threshold θ_{th} obtained by using our algorithm optimizing over all transversal Clifford gates can be observed relative to optimizing over all Pauli gates. The Pauli-twirl reduces (increases) the threshold when optimizing over all transversal Clifford (Pauli) gates for all values of ϕ.

values of ϕ and so Pauli twirling reduces the threshold. However, if only transversal Pauli operators are used, Pauli twirling increases the threshold for all values of ϕ.

The curves in fig. 8 also demonstrate that the threshold

can increase by at most a factor of 1.7 when optimizing over all transversal gates for coherent noise compared to optimizing over all transversal gates for the twirled channel. This advantage arises for two reasons. First, for a known noise model, a transversal gate can be applied to map it to another noise model that may be closer to the identity. Second, syndrome measurements may map coherent errors closer to a non-Pauli unitary. However, both these benefits are lost when the noise is twirled because both the physical noise and the noise for each syndrome is Pauli noise, which is generally far from any non-Pauli unitary.

IX. SENSITIVITY AND ROBUSTNESS OF OUR HARD DECODING OPTIMIZATION ALGORITHM TO PERTURBATIONS OF THE NOISE MODEL

In section III, we presented a hard decoding algorithm for optimizing threshold values of an error correcting code for arbitrary CPTP maps. Our algorithm can therefore be applied to non-Pauli channels, including more realistic noise models that could be present in current experiments. However, the noise afflicting an experimental system is only ever approximately known. Nevertheless, we now demonstrate that applying the decoder obtained by our algorithm for a fixed noise channel $\mathcal{N}$ to a perturbed noise model $\mathcal{N}_p$ retains, in some cases, improvements in error suppression relative to the symmetric decoder.

To study perturbations about a noise channel, let $\mathcal{N} = \{\mathcal{N}_p : p \in [0,1]\}$ be a 1-parameter noise model and

$$\mathcal{N}_{U,p}(\rho) = [1 - f(p)]\mathcal{N}_p + f(p)U\rho U^\dagger \qquad (47)$$

where U is a random unitary and $f : [0,1] \to [0,1]$ is a function such that $f(p) \ll \|\mathcal{N}_p - \mathcal{I}\|$ for any suitable norm (e.g., the diamond norm). (The generalization to multi-parameter noise families is straightforward.)

We applied our algorithm to $\mathcal{N}_p$ and $\mathcal{N}_{U,p}$, giving the effective process matrices $\mathcal{G}(\mathcal{N}_p)$ and $\mathcal{G}(\mathcal{N}_{U,p})$ respectively. We then applied the symmetric decoder and the decoder optimized for $\mathcal{N}_p$ to the perturbed noise $\mathcal{N}_{U,p}$ to obtain the process matrices $\mathcal{G}_{U,sym}$ and $\tilde{\mathcal{G}}_U$ respectively. The infidelities of these process matrices (at the first concatenation level) are plotted in fig. 9 (a) for the $[\![5,1,3]\!]$ code with coherent noise and $f(\theta) = \sin^2 \theta / 10$ and in fig. 9 (b) for the Steane code with amplitude-phase damping, $p = 0.2$ and $f(\lambda) = \lambda/10$. For both plots we averaged the values over 100 uniformly random unitaries. These results demonstrate that the significant improvements obtained using the optimized decoder are, in most studied cases, robust to perturbations in the noise.

The one exception we observed is for coherent noise in the $[\![5,1,3]\!]$ code for $\theta \lesssim 0.185$, where the infidelity obtained using the optimized decoder for the unperturbed channel is larger than that obtained using the symmetric decoder by a factor of at most 7.

X. CONCLUSION

In this paper, we presented an optimized hard decoding algorithm for arbitrary local Markovian noise and numerical techniques to characterize thresholds for noise models. Block-wise two-qubit correlated noise was also considered. Using the analytical tools of section II, we provide numerical results in sections V to IX which shows substantial improvements obtained by our algorithms compared to a fixed decoder for a variety of noise models, including coherent errors, correlated dephasing and amplitude-phase damping, and codes, namely, the $[\![5,1,3]\!]$, Steane, Shor and surface-17 codes. For coherent noise, our optimized decoding algorithm allowed, in some cases, the noise to be corrected for all sampled rotation angles and reduced infidelities at a fixed concatenation level by orders of magnitude.

Our hard decoding algorithm is scalable and efficiently optimizes the recovery operations independently at each concatenation level while taking advantage of a code's transversal gates. At a given concatenation level, all syndrome measurements are considered rather than being sampled from a distribution, so that the performance is exactly characterized rather than containing statistical (and state-dependent) uncertainties.

In contrast to hard decoding, message-passing algorithms [15] can increase thresholds for Pauli noise, in some cases nearing the hashing bound subject to sampling uncertainties. Large codes can also be studied using tensor networks [17], although this requires a tensor-network description of the code and is exponential in the code distance. An interesting and important open problem is to combine the current techniques with those of Refs. [15, 17] to either reduce statistical uncertainties in message-passing algorithms by exploiting symmetries in the code or to treat larger, non-concatenated code families.

Further, we showed that performing a Pauli twirl can increase or decrease the threshold depending on the code and noise properties. In [10], the Pauli twirl was found to have little impact on the performance of amplitude damping, which is known to be "close" to Pauli noise (that is, exhibit similar worst-case errors) [30]. We conjecture that Pauli twirling will generally reduce thresholds for codes that have many transversal gates, but may improve performance for codes with fewer transversal gates.

Lastly, we considered the robustness of our hard decoding optimization algorithm to noise channels that were not perfectly known. We showed that by optimizing our decoder for a channel that was slightly perturbed by a random unitary operator from the actual channel acting on the qubits, it was still possible to obtain improved error rates over the symmetric decoder. However, there are some circumstances where the optimized decoder, while still being robust, is outperformed by the symmetric decoder. Determining the robustness of decoders is an open problem that will be especially relevant when
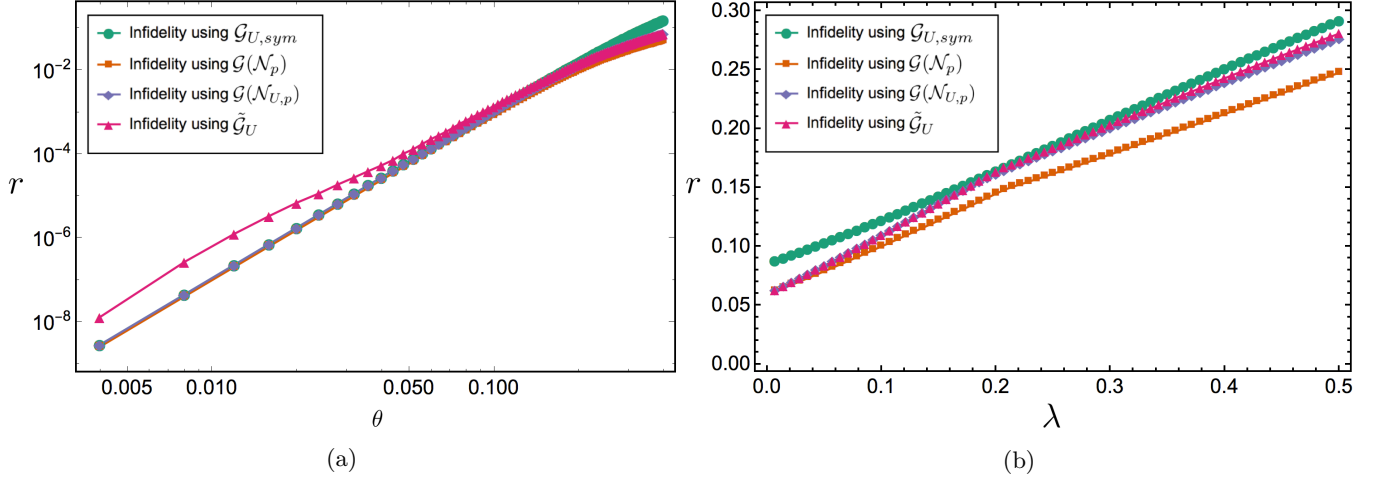
FIG. 9: Averaged infidelity plots over 100 random unitary operators U of the effective process matrices $\mathcal{G}(\mathcal{N}_p)$, $\mathcal{G}(\mathcal{N}_{U,p})$, $\mathcal{G}_{U,sym}$ and $\tilde{\mathcal{G}}_U$. The figure in (a) is obtained using the $[\![5,1,3]\!]$ code for coherent errors using random rotation axes for each random unitary. The perturbation was chosen to have the form $f(\theta) = \sin^2\theta/10$. The figure in (b) is obtained using the Steane code for the amplitude-phase damping channel. The perturbation was chosen to have the form $f(\lambda) = \lambda/10$. In (a), the inset plot shows all infidelities on a log-log scale in the regime where θ is small. As can be seen from the figure, in the regime where $\theta \gtrsim 0.185$, the optimized recovery maps for the unperturbed channel yield a lower infidelity when applied to the perturbed channel than that from applying the symmetric decoder. For smaller rotation angles, the infidelity from $\tilde{\mathcal{G}}_U$ is slightly larger than the infidelity arising from $\mathcal{G}_{U,sym}$. The two differ by at most a factor of 7 in the small θ limit. The infidelity obtained by applying the hard decoding optimization algorithm to the unperturbed channel is lowest for all sampled values of θ. In (b), it can be observed that applying the decoder chosen by our optimization algorithm for the unperturbed channel to the perturbed channel results in a lower infidelity than applying the symmetric decoder to the perturbed channel, for all sampled values of λ. This indicates that our decoding scheme is very robust to small perturbations of the amplitude-phase damping channel.

decoders are used for experimental systems with incompletely characterized noise.

In Refs. [10, 11], the process matrix formalism was used to obtain pseudo-thresholds for the Steane code using the standard CSS decoder. Measurement errors were taken into account, resulting in more accurate pseudo-threshold values. Our methods were developed assuming that the encoding and decoding operations were perfect. The next step in our work will be to generalize our results to include measurement and state-preparation errors.

XI. ACKNOWLEDGEMENTS

[1] Peter W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," Proc. 35th Annu. Symp. Found. Comput. Sci. , 124 (1994).

[2] Christof Zalka, "Simulating quantum systems on a quantum computer," Proc. R. Soc. A Math. Phys. Eng. Sci. **454**, 313 (1998).

[3] Peter W. Shor, "Scheme for reducing decoherence in quantum computer memory," Phys. Rev. A **52**, 2493 (1995).

[4] Dorit Aharonov and Michael Ben-Or, "Fault-Tolerant Quantum Computation With Constant Error Rate," SIAM J. Comput. **38**, 1207 (1999).

[5] John Preskill, "Reliable quantum computers," Proc. R. Soc. A Math. Phys. Eng. Sci. **454**, 385 (1998).

[6] Emanuel Knill, Raymond Laflamme, and Wojciech Hubert Zurek, "Resilient Quantum Computation," Science **279**, 342 (1998).

[7] Pavithran Iyer, and David Poulin, "Hardness of decoding quantum stabilizer codes" IEEE **61**, 5209-5223 (2015).

[8] Panos Aliferis, Daniel Gottesman, and John Preskill, "Accuracy threshold for postselected quantum computation," Quantum Inf. & Comput. **8**, 181 (2008).

[9] Scott Aaronson and Daniel Gottesman, "Improved simulation of stabilizer circuits," Phys. Rev. A **70**, 052328

(2004).

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

[11] Mauricio Gutiérrez, Conor Smith, Livia Lulushi, Smitha Janardan, and Kenneth R. Brown, "Errors and pseudothresholds for incoherent and coherent noise," Phys. Rev. A **94**, 042338 (2016).

[12] Michael R. Geller and Zhongyuan Zhou, "Efficient error models for fault-tolerant architectures and the Pauli twirling approximation," Phys. Rev. A **88**, 012314 (2013).

[13] Yu Tomita and Krysta M. Svore, "Low-distance surface codes under realistic quantum noise," Phys. Rev. A **90**, 062320 (2014).

[14] Benjamin Rahn, Andrew C. Doherty, and Hideo Mabuchi, "Exact performance of concatenated quantum codes," Phys. Rev. A **66**, 032304 (2002).

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

[16] Jesse Fern, "Correctable noise of quantum-error-correcting codes under adaptive concatenation." Phys. Rev. A **77**, 010301 (2008).

[17] Andrew S. Darmawan and David Poulin, "Tensor-network simulations of the surface code under realistic noise," (2016)arXiv:1607.06460.

[18] K. Kraus, *States, Effect, and Operations* (Springer-Verlag, 1983).

[19] Michael A. Nielsen and Isaac L. Chuang, *Quantum computation and quantum information*, 1st ed. (Cambridge University Press, New York, 2000).

[20] Carlo Cafaro and Peter van Loock, "Approximate quantum error correction for generalized amplitude damping errors," Phys. Rev. A **89**, 022316 (2014).

[21] Daniel Gottesman, "An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation," **68**, 13 (2010), arXiv:0904.2557.

[22] Adam Paetznick and Ben W. Reichardt, "Fault-tolerant ancilla preparation and noise threshold lower bounds for the 23-qubit Golay code," Quantum Inf. & Comput. **12**, 43 (2011).

[23] Christopher Chamberland, Tomas Jochym-O'Connor, and Raymond Laflamme, "Thresholds for Universal Concatenated Quantum Codes," Phys. Rev. Lett. **117**, 010501 (2016).

[24] Christopher Chamberland, Tomas Jochym-O'Connor, and Raymond Laflamme, "Overhead analysis of universal concatenated quantum codes," Phys. Rev. A. **95**, 022313 (2017).

[25] Raymond Laflamme, Cesar Miquel, Juan Pablo Paz, and Wojciech Hubert Zurek, "Perfect Quantum Error Correcting Code," Phys. Rev. Lett. **77**, 198 (1996).

[26] Andrew Steane, "Multiple-Particle Interference and Quantum Error Correction," Proc. R. Soc. A **452**, 2551 (1996).

[27] Veronika Baumann and Časlav Brukner, "Appearance of causality in process matrices when performing fixed-basis measurements for two parties," Phys. Rev. A **93**, 062324 (2016).

[28] Shelby Kimmel, Marcus P. da Silva, Colm A. Ryan, Blake R. Johnson, and Thomas Ohki, "Robust Extraction of Tomographic Information via Randomized Benchmarking," Phys. Rev. X **4**, 011050 (2014).

[29] Joseph Emerson, Marcus Silva, Osama Moussa, Colm A. Ryan, Martin Laforest, Jonathan Baugh, David G. Cory, and Raymond Laflamme, "Symmetrized characterization of noisy quantum processes." Science **317**, 1893 (2007).

[30] Richard Kueng, David M. Long, Andrew C. Doherty, and Steven T. Flammia, "Comparing Experiments to the Fault-Tolerance Threshold," Phys. Rev. Lett. **117**, 170502 (2016).

Appendix A: α and β coefficients in closed form

In this section we provide an alternative derivation of the α and β coefficients found in eq. (18) and eq. (19). The latter coefficients will be given in terms of the symplectic vector representation of Pauli operators. For the bit strings $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$, we will write

$$Z(a)X(b) = (Z^{a_1} \otimes \ldots \otimes Z^{a_n})(X^{b_1} \otimes \ldots \otimes X^{b_n}). \quad (A1)$$

Since the α coefficient is related to the overall sign of the operator $S_k\overline{\tau}$ (see eq. (6) and eq. (17)), the goal is to obtain an expression relating the overall sign of $S_k\overline{\tau}$ to its symplectic vector representation. An operator $S_k \in \mathcal{S}$ can always be written as a product of the codes stabilizer generators so that

$$S_k = g_{j_1} \ldots g_{j_k}, \quad (A2)$$

where

$$g_{j_i} = Z(a_{j_i})X(b_{j_i}). \quad (A3)$$

Defining

$$\overline{a} = a_{j_1} + \ldots + a_{j_k} \pmod 2 \quad (A4)$$
$$\overline{b} = b_{j_1} + \ldots + b_{j_k} \pmod 2, \quad (A5)$$

we commute all the Z operators in eq. (A2) to the left, allowing us to write S_k as in eq. (A1)

$$S_k = (-1)^{f(a_{j_1}, \ldots, a_{j_k}; b_{j_1}, \ldots, b_{j_k})} Z(\overline{a})X(\overline{b}). \quad (A6)$$

The overall sign can be obtained from the function f, which is given by

$$f(a_{j_1}, \ldots, a_{j_k}; b_{j_1}, \ldots, b_{j_k}) = \sum_{l=1}^{k-1} \sum_{t=l+1}^{k} b_{j_l} a_{j_t}. \quad (A7)$$

Writing the logical Pauli operator $\overline{\tau}$ as

$$\overline{\tau} = Z(\tau_z)X(\tau_x), \quad (A8)$$

$S_k\overline{\tau}$ can then be written as

$$S_k\overline{\tau} = (-1)^{f(a_{j_1}, \ldots, a_{j_k}; b_{j_1}, \ldots, b_{j_k}) + \overline{b} \cdot \tau_z} Z(\overline{a} + \tau_z)X(\overline{b} + \tau_x). \quad (A9)$$

For any $\mu_j \in \{I, X, Y, Z\}$, we can write μ_j in terms of X and Z Pauli operators:

$$\mu_j = (-i)^{a_j b_j} Z^{a_j} X^{b_j}, \quad (A10)$$

allowing us to write

$$Z(\overline{a} + \tau_z)X(\overline{b} + \tau_x) = i^{(\overline{a} + \tau_z) \cdot (\overline{b} + \tau_x)} \mu_1 \otimes \ldots \otimes \mu_n. \quad (A11)$$

It is important to note that the dot product in the factor of i is *not* added modulo 2.

Using eq. (A11), we have

$$S_k\overline{\tau} = (-1)^{f(a_{j_1}, \ldots, a_{j_k}; b_{j_1}, \ldots, b_{j_k}) + \overline{b} \cdot \tau_z} i^{(\overline{a} + \tau_z) \cdot (\overline{b} + \tau_x)} \mu_1 \otimes \ldots \otimes \mu_n. \quad (A12)$$

Since the α coefficient takes into account the overall sign of the product between elements in the stabilizer group and the logical operators, we have

$$\alpha^{\tau}_{\phi(S_k\overline{\tau})} = \frac{1}{2^{\frac{n}{2}-1}}(-1)^{f(a_{j_1}, \ldots, a_{j_k}; b_{j_1}, \ldots, b_{j_k}) + \overline{b} \cdot \tau_z} i^{(\overline{a} + \tau_z) \cdot (\overline{b} + \tau_x)}, \quad (A13)$$

where the normalization factor arises from choosing a trace orthonormal basis in the sum of E_τ.

Given a recovery map R_l for the syndrome measurement l, we can write it in terms of its symplectic vector representation as

$$R_l = Z(a_l)X(b_l). \quad (A14)$$

From eq. (15) and eq. (17), the β coefficient corresponding to the recovery map R_l can be obtained by commuting R_l to the left of $S_k\overline{\tau}$ and using $R_l^\dagger R_l = I$. Doing so, we find that

$$\beta^{\tau}_{\phi(S_k\overline{\tau})}(R_l) = \alpha^{\tau}_{\phi(S_k\overline{\tau})}(-1)^{a_l \cdot (\overline{b} + \tau_x) + b_l \cdot (\overline{a} + \tau_z)}. \quad (A15)$$