

## Performance of quantum error correction with coherent errors

Eric Huang,<sup>1</sup> Andrew C. Doherty,<sup>1</sup> and Steven Flammia<sup>1,2</sup>

<sup>1</sup>Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, NSW 2006, Australia

<sup>2</sup>Yale Quantum Institute, Yale University, New Haven, Connecticut 06520, USA



(Received 30 May 2018; published 12 February 2019)

We compare the performance of quantum error correcting codes when memory errors are unitary with the more familiar case of dephasing noise. For a wide range of codes, we analytically compute the effective logical channel that results when the error correction steps are performed noiselessly. Our examples include the entire family of repetition codes, the five-qubit, Steane, Shor, and surface codes. When errors are measured in terms of the diamond norm, we find that the error correction is typically much more effective for unitary errors than for dephasing. We observe this behavior for a wide range of codes after a single level of encoding, and in the thresholds of concatenated codes using hard decoders. We show that this holds with great generality by proving a bound on the performance of any stabilizer code when the noise at the physical level is unitary. By comparing the diamond norm error  $D'_\diamond$  of the logical qubit with the same quantity at the physical level  $D_\diamond$ , we show that  $D'_\diamond \leq cD_\diamond^d$  where  $d$  is the distance of the code and  $c$  is a constant that depends on the code but not on the error. This bound compares very favorably to the performance of error correction for dephasing noise and other Pauli channels, where an error correcting code of odd distance  $d$  will exhibit a scaling  $D'_\diamond \sim D_\diamond^{(d+1)/2}$ .

DOI: [10.1103/PhysRevA.99.022313](https://doi.org/10.1103/PhysRevA.99.022313)

### I. INTRODUCTION

Building a large-scale quantum computer will require substantial efforts to mitigate noise through the use of quantum error correction and fault tolerance. The fault tolerance threshold theorem [1–4] guarantees that as long as the errors are sufficiently rare and weakly correlated, an arbitrarily long quantum computation can proceed indefinitely and with low overhead. The exact numerical value of the threshold depends critically on the assumptions about the noise, and from the perspective of fault-tolerant quantum computing not all types of errors are equivalent. This is even true for uncorrelated noise since errors such as dephasing and depolarizing noise are purely stochastic, but control errors such as unitary over- or under-rotation can add coherently.

This distinction between stochastic and coherent errors was recognized quite early on to be important [5]. In particular, using our best theorems to date, the only known way to relate the threshold for stochastic errors to the threshold for coherent errors is to *square* the stochastic value of the threshold [6]. Thresholds quoted in the literature for stochastic-type noise range between values of about  $10^{-2}$  to  $10^{-4}$  depending on how generous the assumptions are on the stochastic noise and whether the threshold is a numerical estimate based on simulation [7] or a theorem based on rigorous proof [8] (or something intermediate). Not knowing if these values need to be squared therefore represents a rather large gap in our understanding of the threshold. The situation is complicated by the fact that the only large-scale numerical simulations that are tractable must necessarily deal with Pauli errors, for which it is known that the squaring is unnecessary [9].

The role that non-Pauli errors play in the fault-tolerance threshold is therefore quite poorly understood. Some recent works are beginning to develop our understanding, such as the use of the so-called honest Pauli approximation [10] or

recent full-scale simulations of coherent noise, using small codes [11], using tools such as tensor networks [12], or in some special cases via an exact solution [13–15].

The focus of this paper is on understanding the role of coherent errors in quantum error correction. In order to motivate our main results, it is helpful to focus the discussion on two meaningful quantities that we wish to study, the average gate infidelity  $r$  and the diamond distance  $D_\diamond$ , defined below. This will also help motivate the particular scaling behavior that we seek to quantify.

We would like to compare quantities that can actually be measured in experiments directly to the fault-tolerance threshold. Unfortunately, here we see another large gap between what we can measure and what we can infer about the threshold. For example, the average gate infidelity  $r(\mathcal{E})$ , defined as

$$r(\mathcal{E}) = 1 - \int d\psi \langle \psi | \mathcal{E}(\psi) | \psi \rangle, \quad (1)$$

is routinely measured to high precision in randomized benchmarking experiments [16]. However, all of our provable knowledge about thresholds uses a much stricter error metric, the diamond distance  $D_\diamond(\mathcal{E})$  to the identity, defined as

$$D_\diamond(\mathcal{E}) = \frac{1}{2} \|\mathcal{E} - \mathcal{I}\|_\diamond := \frac{1}{2} \sup_\rho \|\mathcal{E} \otimes \mathcal{I}(\rho) - \rho\|_1 \quad (2)$$

in terms of the Schatten 1-norm (the sum of the singular values). The gap between these two quantities in the regime of interest can be orders of magnitude in general, a point recently emphasized by Sanders *et al.* [17]. Unfortunately, there is no simple way to measure the diamond distance in general without doing complete process tomography [18].

Recent work has sought ways to quantify the worst-case behavior needed to prove a threshold theorem using measurement methods that are preferably scalable and avoid

tomography. Examples include gate set tomography [18] and the unitarity [19], though neither method is completely scalable in contrast to randomized benchmarking.

Complementing this line of research, and motivating the upper bound we prove below in Theorem 1, is the theoretical approach of [20,21], which seeks to classify physical noise sources in terms of “good” and “bad” noise scaling. A family  $\{\mathcal{E}_\gamma : \gamma \in [0, 1]\}$  of noise models with  $\mathcal{E}_0 = \mathcal{I}$  has “good” scaling if  $D_\diamond(\mathcal{E}_\gamma) \leq Cr(\mathcal{E}_\gamma)$  and “bad” scaling if  $D_\diamond(\mathcal{E}_\gamma) \geq C'\sqrt{r(\mathcal{E}_\gamma)}$  for all  $\gamma \in [0, 1]$  and some constants  $C, C' > 0$  that are independent of  $\gamma$ . Other scalings are also possible, and could be called “intermediate.” The purpose of this coarse distinction is that for a given noise channel  $\mathcal{E}$ , if  $D_\diamond(\mathcal{E}) \approx r(\mathcal{E})$ , then the proxy measure of  $r(\mathcal{E})$  that is easily obtained via benchmarking gives a good indication of how close one might be to the threshold, while if  $D_\diamond(\mathcal{E}) \approx \sqrt{r(\mathcal{E})}$ , then the proxy is highly misleading. These scaling limits are extremal [17,20–22]. Of course, the usefulness of this distinction depends implicitly on the constants  $C$  and  $C'$  being relatively civilized. Taking, say,  $C = 10^{15}$  shows that all noise is trivially good until  $r(\mathcal{E}_\gamma) \lesssim 10^{-30}$ . Similarly, we also need  $C'$  to be sufficiently large for this distinction to be meaningful in practice. However, while we are always ultimately interested in absolute noise rates on single instances, hiding these constants and discussing scaling enables us to make important physical insights into the nature of noise sources and what the expected effect might be on quantum information.

Illustrating the utility of this scaling behavior perspective, Ref. [20] classifies many common qubit noise models such as dephasing, depolarizing, amplitude damping, leakage, and unitary errors according to this dichotomy. It is only the unitary errors and so-called “coherent leakage” that exhibit bad error scaling [20]. By combining some knowledge of the dominant noise process with measurements of  $r(\mathcal{E})$  and the aforementioned unitarity, Refs. [20,21] show that it is possible to obtain pertinent information about how close one’s qubits are to the fault-tolerance threshold.

This discussion of scaling focused on the case of *physical* errors, but even more desirable would be to understand the scaling of *logical* error rates in quantum error correcting codes [23]. Motivated by the above notions of good and bad error scaling in regards to physical errors, we seek to develop an analogous understanding of how coherent and incoherent errors scale at the logical level. It has been observed numerically that there can be orders of magnitude difference in the logical error rates after concatenating quantum codes [24], so *a priori* it is not clear that it is possible to obtain a simple scaling dichotomy as we seem to have in the case of physical errors.

### Summary of results

In this paper, we focus on physical qubit noise channels with both unitary control errors and dephasing noise. We aim to compare the performance of ideal error correction for coherent noise channels, for which the diamond norm distance greatly exceeds the average fidelity error, with dephasing, for which they are comparable. The main tool is to calculate effective noise channels for the encoded qubits after error correction, following Rahn *et al.* [25]. Broadly, the results of our investigations into the performance of ideal

error correction under coherent errors suggest that simply considering diamond norm error at the physical level overstates the effect that residual coherent errors will have on logical information in a quantum computer.

In order to develop our analytical methods, we begin in Sec. III by revisiting the case of the repetition codes studied by Greenbaum and Dutton [13]. Although these are not proper quantum error correcting codes, they can correct against  $Z$  rotations and dephasing, and are simple enough to be analytically tractable for arbitrarily high distance codes. As has been shown previously [24], the performance of error correction depends very strongly on how coherent the noise process is, and is not predicted by a single figure of merit for the noise, such as average fidelity or diamond norm distance.

One result of this analysis relates to the nature of the effective noise channels for codes where the distance is odd or even. If the noise process is a purely unitary rotation about  $Z$  and the repetition code has odd distance, we find that the effective noise channel conditioned on syndrome measurement is also unitary. Consequently, the statistics of syndrome measurement is independent of the logical state of the code. For the case of even distance repetition codes, however, we find that the effective channel conditioned on the syndrome measurement corresponds to a (very) weak measurement of logical  $Z$ . As a result, the syndrome measurement outcomes do depend weakly on the logical state of the qubit. However, the effective noise channel averaged over syndromes is pure dephasing, even when the underlying physical noise process is purely unitary. We have observed qualitatively similar behavior for even distance surface codes.

In Sec. IV we apply these methods of calculation of effective channels to general stabilizer codes. Our main analytical result is Theorem 1, which roughly states the following. For any  $[n, k, d]$  stabilizer code, the logical diamond norm error  $D'_\diamond$  after ideal error correction of a purely unitary error is bounded by  $D'_\diamond \leq c_{n,k} D_\diamond^d$  in terms of the physical diamond norm error  $D_\diamond$ , where  $c_{n,k}$  is a constant independent of the errors (but may depend on the code). This result does not just apply to uniform unitary noise, but is readily generalized to hold for all single-qubit unitary noise if we express the bound in terms of the largest value of  $D_\diamond$  across the physical qubits. This is very favorable behavior compared to the performance of error correction for dephasing noise. Recall that an error correcting code of odd distance  $d$  will correct  $t$  dephasing errors where  $d = 2t + 1$ . Consequently, for dephasing it is known that  $D'_\diamond$  scales like  $D_\diamond^{t+1}$ . Another way to assess this comparison between coherent and dephasing noise after error correction is to express this bound in terms of the average fidelity error  $r$  at the physical level  $D'_\diamond \leq cr^{t+1/2}$ , where  $c$  is a constant that depends on the code but not on the unitary error. (Recall that for a coherent error  $D_\diamond$  scales like  $r^{1/2}$ .) This compares quite favorably to the well-known scaling  $r^{t+1}$  for dephasing errors.

In Sec. V we compute the effective noise channels for a range of quantum codes using an automated procedure. We consider unitary qubit noise and study how the performance of the code depends on the axis of rotation in the Bloch sphere. There are frequently large effects. For example, for the surface code the error correction performs much better for rotations about the  $Y$  axis, than for rotations about  $X$  and  $Z$ . This

observation is analogous to the recently discussed behavior of the surface code under Pauli noise [26].

Finally, in Sec. VI we study concatenated codes for noise channels that combine  $Z$  rotations with dephasing using a hard decoder. Again, we use an automated procedure to generate the effective channel for a single level of encoding. By regarding these effective channels as a map on noise processes it is possible to find the threshold for a hard decoder as a fixed point of the map. We compare the two extreme limits of unitary  $Z$  rotations and pure dephasing and find that the threshold as a function of the diamond norm error is in every code we tested larger for the case of unitary errors.

## II. NOISE MODELS AND EFFECTIVE CHANNELS

Suppose we have a qubit Hamiltonian  $H$  with  $H^2 = I$ , then coherent noise channels result from unitary noise processes of the form

$$U_{H,\theta} = e^{-i\theta H} = \cos \theta I - i \sin \theta H. \quad (3)$$

Physically, such error processes arise from over or under rotation in qubit control pulses. Note that since we are considering qubit noise processes, the choice  $H^2 = I$  allows for a general Bloch sphere rotation axis.

It is frequently of interest to combine coherent and incoherent errors. To study this sort of noise process we will follow Kueng *et al.* [20] and specialize to rotations about the  $Z$  axis of angle  $\theta$  and dephasing with probability  $p$ . This simple phenomenological model allows us to make a detailed comparison between the effects of coherent and incoherent errors, but it also corresponds to the dominant noise processes in many experimental implementations of quantum computing. The resulting noise channel is

$$\mathcal{N}(\rho) = e^{-i\theta Z}[(1-p)\rho + pZ\rho Z]e^{i\theta Z} \quad (4)$$

$$= (1-x)\rho + xZ\rho Z - iy(Z\rho - \rho Z). \quad (5)$$

In the second expression we have used an alternative parametrization of the noise process that will greatly simplify certain calculations. The real parameters  $x, y$  are defined as follows:

$$x = p \cos^2 \theta + (1-p) \sin^2 \theta, \quad (6)$$

$$y = (1-2p) \cos \theta \sin \theta. \quad (7)$$

It has become common in the literature to compare a noise model such as this with some Pauli channel that approximates it [10,11,27,28]. For this example, the Pauli twirling approximation to this error is just to project down to the case where  $y = 0$ . Thus, for a given initial noise model given by  $p, \theta$  the Pauli twirling approximation is to consider instead the model with  $p' = p \cos^2 \theta + (1-p) \sin^2 \theta$  and  $\theta' = 0$ . Since we are mainly interested in the scaling when error rates are small, we will often expand expressions such as  $x$  and  $y$  to lowest nontrivial order in  $p$  and  $\theta$ . So, for example, we have  $x \approx p + \theta^2$  and  $y \approx \theta$ .

In the following, we will also occasionally be interested in the noise operations conditional on the outcome of a syndrome measurement in an error correction procedure. These noise

processes will likewise combine unitary and dephasing error but will not necessarily preserve trace, being of the form

$$\mathcal{N}_s(\rho) = \bar{x}\rho + xZ\rho Z - iy(Z\rho - \rho Z) \quad (8)$$

with the case  $\bar{x} = 1 - x$  preserving the trace, and  $x + \bar{x} \leq 1$  corresponding to some stochastic process occurring with probability  $x + \bar{x}$ . Here, the subscript  $s$  labels a particular syndrome.

We will frequently be concerned with measures of the strength of the noise process. Although there are many possible choices, we will confine ourselves to two frequently used measures. The first is the average gate infidelity  $r$  because it can be estimated accurately in randomized benchmarking experiments. The second is the diamond norm error  $D_\diamond$  because it can be used to bound the overall error when noise processes can occur sequentially in a computation and therefore appears in the statement of fault-tolerance threshold theorems. For our model channel, both of these error metrics were calculated by Kueng *et al.* [20] for the above unitary and dephasing channel:

$$r = \frac{2x}{3} = \frac{2}{3}[p \cos(2\theta) + \sin^2(\theta)] \approx \frac{2}{3}(p + \theta^2), \quad (9)$$

$$D_\diamond = \sqrt{x^2 + y^2} = \frac{1}{2}|1 - (1-2p)e^{2i\theta}| \approx \sqrt{p^2 + \theta^2}. \quad (10)$$

Notice that in the limit  $p \rightarrow 0$  corresponding to a unitary error,  $D_\diamond = |\sin \theta| \approx |\theta|$  while  $r \approx \theta^2$  so that the diamond norm distance can be much larger than the average fidelity error [17,20]. Also notice the simplifications in these formulas when expressed in terms of  $x$  and  $y$ , and in particular that  $r$  is independent of  $y$ , while the diamond norm distance is just the length of the vector  $(x, y)$ .

We wish to study the performance of  $[n, k, d]$  quantum error correcting codes that encode  $k$  logical qubits in  $n$  physical qubits with code distance  $d$ . As discussed in the Introduction, we will focus on ideal error correction where noise processes act only on the qubit memory and the encoding and error correction are performed ideally. Such an ideal error correction process defines an effective channel on the logical qubits. We wish to build intuition about the effectiveness of error correction for coherent errors by studying these error channels analytically.

The effective channel on the logical qubits arises from composing the operations of encoding, noise, error correction, and then decoding. As such, it is described by a completely positive trace-preserving map on the  $k$  logical qubits. It can be written in the form

$$\mathcal{N}_L = \mathcal{E}^\dagger \mathcal{R} \tilde{\mathcal{N}} \mathcal{E}, \quad (11)$$

where  $\mathcal{E}$  is the encoding map (an isometry),  $\tilde{\mathcal{N}}$  is the noise acting on the  $n$  qubits of the code,  $\mathcal{R}$  is recovery by syndrome measurement with correction, and  $\mathcal{E}^\dagger$  reverses the encoding. The application of quantum operations in our notation follows the same conventions as matrix multiplication, with operations occurring earlier being written to the right of later operations.

For independent qubit noise on an  $n$ -qubit code, the noise process acts on each physical qubit independently and

$$\tilde{\mathcal{N}} = \bigotimes_{m=1}^n \mathcal{N}, \quad (12)$$

where the single-qubit noise process  $\mathcal{N}$  is assumed to act independently and identically on each physical qubit. Most of our considerations can be generalized to the case of nonidentical noise processes acting independently on the different physical qubits, but for now we defer a discussion of this case.

The error correction process is made up of syndrome measurement followed by correction. We consider two natural cases. First, to obtain a trace-preserving map, we average over all syndromes to obtain the error correction map  $\mathcal{R}$ . Physically, this corresponds to the error process assigned by an observer external to the error correction process who is unaware of which syndrome arose. Alternatively, we can evaluate the conditional operation where a particular syndrome  $\mathbf{s}$  is specified and the recovery map  $\mathcal{R}_{\mathbf{s}}$  applies the projection corresponding to this syndrome and then performs the subsequent correction. The resulting conditional noise process on the logical qubits  $\mathcal{N}_{\mathbf{s}}$  will be completely positive but not trace preserving; it will be normalized by the probability of the specified syndrome.

Techniques for calculating the effective noise channel  $\mathcal{N}_L$  are discussed at length by Rahn *et al.* [25]. As described there we can regard the error correction procedure as a map on qubit noise channels  $\mathcal{N} \rightarrow \mathcal{N}_L$ . Note that while the encoding, error correction, and decoding operations are linear when regarded as a map on the noise process  $\tilde{\mathcal{N}}$ , the map  $\mathcal{N} \rightarrow \tilde{\mathcal{N}}$  is a polynomial map on the matrix elements that define  $\mathcal{N}$ . Specifically, we can expand a general qubit noise process as follows:

$$\mathcal{N}(\rho) = \sum_{P, P'} n_{PP'} P \rho P', \quad (13)$$

where the sum runs over all qubit Pauli matrices. Then, the coefficients in the corresponding expansion for either  $\mathcal{N}_L$  or  $\mathcal{N}_{\mathbf{s}}$  are  $n$ th-order polynomials of the coefficients  $n_{PP'}$ . In the case of the noise process [Eq. (5)] we conclude that the effective channels after error correction can be written as polynomials of the parameters  $x$  and  $y$ . We will make use of this observation to simplify analytical calculations in the following section.

In the majority of our examples below we have used *Mathematica* to automate the computation of the effective channel  $\mathcal{N}_L$  using the approach outlined by Rahn *et al.* [25]. The results of those calculations are discussed in detail in Secs. V and VI. However, when the error correcting code and recovery map have high symmetry it is possible to compute the effective channel explicitly for unitary error processes, and we will now describe that process using the repetition codes as a specific example (this case was also considered by Greenbaum and Dutton [13]). The techniques developed to analyze these codes will then be sufficient to allow us to place bounds on the diamond norm error after correction for general  $[n, k, d]$  stabilizer codes.

### III. EFFECTIVE ERROR CHANNELS FOR COHERENT ERRORS AND REPETITION CODES

In this section we will describe how to calculate effective error channels for coherent error processes, using the repetition codes as a primary example. The high symmetry of these codes allows us to find simple closed-form expressions for essentially all quantities of interest. Despite the fact that these are not proper quantum error correcting codes, the behavior of these simple examples is qualitatively similar to all the other (more interesting) examples that we have studied.

The effective channels for the repetition code with odd  $n$  were first calculated by Greenbaum and Dutton [13]. We present a slightly streamlined analysis and apply it to both the case of even  $n$  and the conditional noise processes  $\mathcal{N}_{\mathbf{s}}$ . We will also see that our techniques can be used to find the effective channel for more general codes and to obtain our main analytical result which is a bound on the diamond norm error after ideal correction for unitary errors and general stabilizer  $[n, k, d]$  codes.

Our approach involves considering the case of unitary errors initially. We are therefore motivated to consider correcting errors of the form  $U = \exp(-i\theta H)$  using an  $[n, k, d]$  error correcting code. So, we have  $n$  physical qubits each undergoing an error  $U$ . We will confine ourselves to identically distributed errors for now. Since  $U = \cos\theta I - i \sin\theta H$  and  $H^2 = I$  we can think of the fundamental error process as  $H$  and each qubit will either have experienced an error  $H$  or not. We can specify all possible tensor products of  $H$  by a binary vector  $\mathbf{w}$ , with ones indicating that  $H$  acts on the corresponding qubit. The Hamming weight  $w$  is just the number of ones  $|\mathbf{w}|$  in  $\mathbf{w}$ . Therefore,  $w$  is the number of qubits that have a nontrivial error. We can indicate error configurations by the operators  $H_{\mathbf{w}} = H^{\otimes w}$ . There are  $\binom{n}{w}$  configurations of weight  $w$ , and we can define the sum over these error configurations as follows:  $E_w = \sum_{\mathbf{w}:|\mathbf{w}|=w} H_{\mathbf{w}}$ . So, the overall  $n$ -qubit error operator is as follows:

$$\begin{aligned} \tilde{U} &= [\exp(i\theta H)]^{\otimes n} = [\cos\theta I - i \sin\theta H]^{\otimes n} \\ &= \sum_{w=0}^n (-i \sin\theta)^w (\cos\theta)^{n-w} E_w. \end{aligned} \quad (14)$$

This expression for the overall unitary error as a sum over Hamming weights holds for any independent unitary error regardless of the error correcting code of interest. In the following, it will be convenient to define the function

$$f_w(\theta) = (-i \sin\theta)^w \cos^{n-w}(\theta). \quad (15)$$

Notice that these functions are increasingly small for high weight. It is easy to see that  $|f_w(\theta)| \leq |\sin\theta|^w |\cos(\theta)|^{n-w} \leq D_{\diamond}^w$  where  $D_{\diamond}$  is the diamond norm error of the unitary noise.

At this point, we have characterized the noise process  $\tilde{\mathcal{N}}$  on the  $n$  physical qubits. The next step is to calculate the effect of error recovery  $\mathcal{R}$ . This will involve measuring the syndrome, and performing a corresponding correction. As a result,  $\mathcal{R}$  involves a sum over the  $2^{n-k}$  syndromes and it will not be possible to find a closed-form expression for a general code. In this work, we have used automated procedures to handle examples with small  $k$ . Moreover, the codes of most



interest possess high symmetry, and this enhanced symmetry can be used to find closed-form expressions in a number of interesting cases.

In order to determine the syndrome that will arise from stabilizer measurement, we now replace the sum over error weights in Eq. (14) by a sum over syndromes. At this point, we have to take advantage of the specific properties of the code. We will focus for the moment on a repetition code with an odd number  $n = 2t + 1$  of physical qubits. The stabilizer generators of the code are  $X_i X_{i+1}$  for  $i = 1, \dots, n - 1$ . There is one encoded qubit with logical operators  $\bar{X} = X^{\otimes n}$  and  $\bar{Z} = Z^{\otimes n}$ . Recall that logical operators of stabilizer codes are defined only up to multiplication by stabilizers. So, we could equally have chosen  $\bar{X}$  to be  $X \otimes I^{\otimes n-1}$ , but we make this choice to emphasize the permutation symmetry of the code. While this is only technically a quantum error correcting code, as it has distance  $d = 1$ , it does have distance  $n$  for  $Z$  errors. Consequently, it should be effective in correcting unitary errors when  $H = Z$ . We will now specialize to that case. Notice that this corresponds to our noise model (5) with  $p = 0$ , or equivalently  $x = \sin^2 \theta$  and  $y = \sin \theta \cos \theta$ .

Since the code can correct  $t$   $Z$  errors, each of the  $2^{n-1}$  error configurations  $Z_w$  with weight less than or equal to  $t$  results in a unique syndrome. For each error configuration  $w$  with weight  $w \leq t$  the weight  $n - w$  error configuration  $Z_w \bar{Z}$  results in the same syndrome. Together these account for all  $2^n$  error configurations. We can rewrite the error unitary (14) as follows:

$$\tilde{U} = \sum_{w=0}^t \sum_{\mathbf{w}:|\mathbf{w}|=w} Z_w [f_w(\theta) \bar{I} + f_{n-w}(\theta) \bar{Z}]. \quad (16)$$

Each term in the sum over  $w$  results in a unique syndrome and so the syndrome measurement step projects onto a single value of  $w$ . The vector  $w$ , which has weight less than or equal to  $t$ , specifies the correction  $Z_w$ . Consequently, when the syndrome requires a correction of weight  $w$  the logical qubit undergoes the effective channel

$$\mathcal{N}_{Lw}(\rho_L) = K_w \rho_L K_w^\dagger, \quad (17)$$

where

$$K_w = f_w(\theta) \bar{I} + f_{n-w}(\theta) \bar{Z}. \quad (18)$$

It is straightforward to show that  $K_w^\dagger K_w = (|f_w|^2 + |f_{n-w}|^2) \bar{I}$  and therefore  $K_w$  is proportional to a unitary. The normalization relates to the probability of the syndrome, and for small  $\theta$  this is approximately  $\theta^{2w}$ . The unitary is a rotation about  $Z$  and the rotation angle is approximately  $(-1)^{t-w} \theta^{2(t-w)+1}$  in this limit. Notice that this rotation angle is much larger for the less likely syndromes; we will discuss this further in a subsequent section. The fact that the effective channel only depends on the weight of the correction is a consequence of the permutation symmetry of the code. For general noise processes, the effective channel conditioned on a given syndrome is analyzed by Chamberland *et al.* [29] as a generalization of the approach of Rahn *et al.* [25].

We can now evaluate the overall noise map as follows:

$$\begin{aligned} \mathcal{N}_L(\rho) &= \sum_{w=0}^t \binom{n}{w} K_w \rho_L K_w^\dagger = (1 - x') \rho \\ &+ x' \bar{Z} \rho \bar{Z} - iy' (\bar{Z} \rho - \rho \bar{Z}), \end{aligned} \quad (19)$$

with

$$x' = \sum_{w=0}^t \binom{n}{w} x^{n-w} (1-x)^w = \sum_{w'=t+1}^n \binom{n}{w'} x^{w'} (1-x)^{n-w'}, \quad (20)$$

$$y' = y^n \sum_{w=0}^t \binom{n}{w} (-1)^{t-w} = \binom{2t}{t} y^{2t+1}. \quad (21)$$

The final expression for  $x'$  is a simple reorganization of the sum to make it clear that  $x' \simeq \binom{n}{t+1} x^{t+1}$  for small  $x$ . It is clear that for small  $x$  the sum for the effective error rate  $x'$  is dominated by the terms arising from syndromes with  $w = t + 1$ , the lowest weight uncorrected errors. [The repetition code corrects  $t$  errors with  $t \simeq n/2$  so the combinatorial factor  $\binom{n}{t+1}$  attains its maximum possible value. By contrast, for surface codes with large distance, for example, the lowest weight errors do not necessarily dominate the effective error rate for realistic values of the parameters.] The final expression for  $y'$  arises from applying Pascal's rule to consecutive terms in the sum [13]. It is clear from this expression that in contrast to  $x'$ , all Hamming weights contribute to the expression for  $y'$  at the same order  $y^{2t+1}$ . Thus, any characterization of the average error will get contributions from all syndromes, even those that correspond to very high weight errors and arise with very low probability. However, the combinatorial factors in the sum in Eq. (21) mean that the sum is largely determined by terms where  $w$  is not too different from  $t$ .

Inspecting the sum in Eq. (21), it is clear that the value of  $y'$  is reduced due to the fact that the sign of the rotation angle of the effective channel for each syndrome oscillates with the Hamming weight. By reducing the magnitude of  $y'$  this cancellation reduces the coherence of the effective channel. It is straightforward to assess the extent to which this cancellation is reducing the coherence of the average channel. If we remove the factor  $(-1)^{t-w}$  from the sum in Eq. (21) we would obtain the value  $2^n y^{2t+1}$ . So, this quantity scales with  $y$  as before but it is larger than  $y'$  by a factor roughly equal to  $\sqrt{\pi t}$  in the limit of large  $t$ .

We have established this identity only for the unitary case where  $x = \sin^2 \theta$  and  $y = \sin \theta \cos \theta$ . However, we will show below that this formula holds for all the combinations of unitary and dephasing errors in Eq. (5). Notice that since  $x'$  depends on  $x$  alone, the Pauli twirling approximation to the effective channel can be found starting from the Pauli twirling approximation at the physical level. This seems unlikely to be a general property of these noise maps.

We can also write the conditional channel (17) for a given syndrome  $w$ . Note that

$$\begin{aligned} \mathcal{N}_{Lw}(\rho_L) &= K_w \rho_L K_w^\dagger = \bar{x}'_w \rho_L + x'_w \bar{Z} \rho_L \bar{Z} \\ &- iy'_w (\bar{Z} \rho_L - \rho_L \bar{Z}), \end{aligned} \quad (22)$$

where

$$\bar{x}'_w = x^w (1-x)^{n-w}, \quad (23)$$

$$x'_w = x^{n-w} (1-x)^w, \quad (24)$$

$$y'_w = (-1)^{t-w} y^n. \quad (25)$$

Note that all of these are proportional to unitary processes satisfying  $y_w^2 = x_w \bar{x}'_w$  since the underlying process is unitary and has  $y^2 = x(1-x)$ .

The overall probability of this syndrome is

$$p_w = x^w(1-x)^{n-w} + x^{n-w}(1-x)^w, \quad (26)$$

and there are  $\binom{n}{w}$  syndromes that have a correction operation of weight  $w$ . Since the probability of each weight depends on  $x$  alone, it can be computed using the Pauli approximation to Eq. (5). Also, the probability of a given weight arising is independent of the logical state. We will see that these properties do not hold for all codes.

As for the unconditional effective channel, we have so far explained how to establish these results with unitary errors  $p = 0$ , but we have rewritten these expressions in terms of the parameters  $x$  and  $y$  such that they hold for all instances of the noise model (5). We will establish these formulas in the general case in the next subsection.

### A. Effective channels with dephasing as well as unitary errors

Given the calculations so far, it is straightforward to see that for the general case of the noise model of Eq. (5) we have

$$\tilde{\mathcal{N}}(\rho) = \sum_{w', \bar{w}, \bar{w}'} p^{w'}(1-p)^{n-w'} f_{\bar{w}}(\theta) f_{\bar{w}'}^*(\theta) Z_{w'+\bar{w}} \rho Z_{w'+\bar{w}'}. \quad (27)$$

Here and elsewhere, addition of the binary vectors  $w$  is modulo two.

If we consider a syndrome measurement with outcome  $w$ , then a projection onto the corresponding stabilizer subspace just picks out the four terms of this sum where both  $w' + \bar{w}$  and  $w' + \bar{w}'$  are equal to either  $w$  or  $n-w$ . ( $n$  is the  $n$ -component binary vector with Hamming weight  $n$ .) The corresponding correction operator is  $Z_w$  as before, and this results in a conditional effective channel of the form of Eq. (22) as expected.

We now just need to evaluate the parameters  $\bar{x}'_w, x'_w, y'_w$ . Considering first the coefficient of  $\rho_L$  we find

$$\begin{aligned} \bar{x}'_w &= \sum_{w', \bar{w}=w-w'} p^{w'}(1-p)^{n-w'} |f_{\bar{w}}(\theta)|^2 \\ &= \sum_{w'=0}^n \sum_{\bar{w}=0}^{w'} \binom{w}{\bar{w}} \binom{n-w}{w'-\bar{w}} p^{w'}(1-p)^{n-w'} |f_{w+w'-2\bar{w}}(\theta)|^2 \\ &= x^w(1-x)^{n-w}. \end{aligned} \quad (28)$$

In the first equality we sum over all binary vectors  $w'$  and  $\bar{w} = w - w'$ . Consider a fixed  $w'$ . Let  $\tilde{w}$  be the number of locations where both  $w$  and  $w'$  have ones. There are  $w' - \tilde{w}$  locations where  $w'$  has a one and  $w$  has a zero. Consequently,  $\bar{w}$  has weight  $|\bar{w}| = w - \tilde{w} + (w' - \tilde{w}) = w + w' - 2\tilde{w}$ . For each  $0 \leq \tilde{w} \leq w'$  there are  $\binom{w}{\tilde{w}} \binom{n-w}{w'-\tilde{w}}$  distinct choices of  $w'$  that have this value of  $\tilde{w}$ . This establishes the second equality above. The final equality is most easily seen by observing that  $x = (1-p)\sin^2\theta + p\cos^2\theta$  and  $1-x = (1-p)\cos^2\theta + p\sin^2\theta$  and then applying the binomial expansion to  $x^w$  and  $(1-x)^{n-w}$ .

The same procedure can be used to compute the other parameters. The case of  $x'_w$  is precisely the same as for  $\bar{x}'_w$ .

For  $y'_w$  we find

$$\begin{aligned} -iy'_w &= \sum_{w', \bar{w}=n-w-w', \bar{w}'=w-w'} p^{w'}(1-p)^{n-w'} f_{\bar{w}}(\theta) f_{\bar{w}'}^*(\theta) \\ &= \sum_{w'=0}^n \sum_{\bar{w}=0}^{w'} \binom{w}{\bar{w}} \binom{n-w}{w'-\bar{w}} p^{w'}(1-p)^{n-w'} \\ &\quad \times f_{n+2\bar{w}-w-w'}(\theta) f_{w+w'-2\bar{w}}^*(\theta) \\ &= -i(-1)^{t-w} [\sin\theta \cos\theta]^n \sum_{w'=0}^n \sum_{\bar{w}=0}^{w'} \binom{w}{\bar{w}} \\ &\quad \times \binom{n-w}{w'-\bar{w}} (-p)^{w'}(1-p)^{n-w'} \\ &= -i(-1)^{t-w} y^n. \end{aligned} \quad (29)$$

One way to see the final equality is to note that  $y = \cos\theta \sin\theta[(1-p) - p]$  and apply the binomial expansion to  $y^w$  and  $y^{n-w}$ .

This establishes the claimed result for the conditional effective channels. Averaging over syndromes recovers the result for the unconditional effective channel.

### B. Repetition codes with even distance

So far, we have considered the case where  $n = 2t + 1$  is odd. It is also interesting to look at the case of an even distance code where  $n = 2t + 2$  physical qubits. We will note two interesting features of this case. First, the unitary part of the logical error channel ( $y'$ ) vanishes after a single round of error correction, so that the effective logical channel is pure dephasing. Second, the probabilities of various syndrome measurement outcomes will depend on the initial logical state (albeit weakly).

The stabilizer generators of the code are  $X_i X_{i+1}$  for  $i = 1, \dots, n-1$ . There is one encoded qubit with logical operators  $\bar{X} = X \otimes I^{\otimes n-1}$  and  $\bar{Z} = Z^{\otimes n}$ . For the odd repetition codes, each syndrome indicates either a weight  $w$  or a weight  $n-w$  error and the symmetric decoder corrects the lowest weight error. When  $n$  is even there is no good way of correcting the Hamming weight  $t+1$  errors, and we just need to decide on some procedure that does not lower the symmetry of the code. In this case for each of the  $w = t+1$  syndromes we will choose one of the two corresponding weight  $t+1$  errors as the correction operation. So, for example, one of the pair will act nontrivially on the first qubit, and we will choose this as the correction operator. This procedure results in a state in the code space and preserves the overall structure and symmetry of the decoder. Looking at the sum over the syndromes in Eq. (16), we should therefore treat the  $w = t+1$  term separately as follows:

$$\begin{aligned} \tilde{U} &= \sum_{w=0}^t \sum_{\mathbf{w}: |\mathbf{w}|=w} Z_{\mathbf{w}} [f_w(\theta) \bar{I} + f_{n-w}(\theta) \bar{Z}] \\ &\quad + \sum_{\mathbf{w}: |\mathbf{w}|=t+1, w_1=1} f_{t+1}(\theta) Z_{\mathbf{w}} [\bar{I} + \bar{Z}]. \end{aligned} \quad (30)$$

It is already clear from the expression for the noise operator in terms of the syndromes that the effective chan-

nel conditioned on a stabilizer measurement is quite different in the even distance case as compared to the odd distance case. We can read off from Eq. (30) that a weight  $t+1$  syndrome results in a projection on a  $\bar{Z}$  eigenstate of the logical qubit, which would of course totally destroy the quantum information. In contrast for odd distance codes, the conditional channel was unitary. For the other syndromes we have as before a conditional channel given by Eq. (17) with the Kraus operator  $K_w$  given by Eq. (18). However, in the even distance case it is easy to show that the Kraus operator is not unitary and in fact  $K_w^\dagger = \pm K_w$ . For small  $\theta$  we find  $K_w \approx (-i)^w \theta^w [\bar{I} - (-1)^{t-w} \theta^{2(t-w)+2} \bar{Z}]$ . Each of these Kraus operators corresponds to performing a weak measurement of  $\bar{Z}$  with a measurement outcome that depends on the weight of the stabilizer.

An important takeaway is this: whereas previously the syndrome probabilities were completely independent of the logical state, this is no longer true for the even distance codes. For these codes, the unitary error is generating some entanglement between the logical and stabilizer qubits and the stabilizer measurement outcomes can depend on the initial logical state.

We can now evaluate the overall noise map for even distance repetition codes, which has the same form as Eq. (19) but with the parameters as follows:

$$x' = \sum_{w=0}^t \binom{n}{w} x^{n-w} (1-x)^w + \frac{1}{2} \binom{n}{t+1} x^{t+1} (1-x)^{t+1} \quad (31)$$

$$= \frac{1}{2} \binom{n}{t+1} x^{t+1} (1-x)^{t+1} + \sum_{w'=t+2}^n \binom{n}{w'} x^{w'} (1-x)^{n-w'}, \quad (32)$$

$$iy' = y^n \left[ \sum_{w=0}^t \binom{n}{w} (-1)^{t-w} - \frac{1}{2} \binom{n}{t+1} \right] = 0. \quad (33)$$

As we argued in the odd case, this expression holds for all initial  $x$  and  $y$ . A cancellation arises such that the unitary component of the error is totally removed after a single round of error correction. Once again, this cancellation depends on the fact that all Kraus operators contribute at the same order, so in practice for a large  $t$  code this involves averaging over very rare events.

As for the case of odd distance repetition codes, it is possible to generalize the conditional effective channels to the general noise model of Eq. (5). We will not write the formulas for  $\bar{x}'_w, x'_w, y'_w$  explicitly. However, note that the overall probability for a syndrome  $w$  is

$$p_w = x^w (1-x)^{n-w} + x^{n-w} (1-x)^w - 2(-1)^{t-w} y^n \langle \bar{Z} \rangle. \quad (34)$$

In contrast to the case of odd distance, this probability depends on  $y$  as well as  $x$  and so the behavior is distinct from that for the Pauli channel approximation. It also depends on the logical state. Measuring the syndrome implements a weak measurement of  $\bar{Z}$  and this is reflected in the syndrome outcome probabilities. Note that for low  $w$ , this dependence is rather weak.

### C. Nonuniform single-qubit errors

These considerations work equally well when the unitary error on each physical qubit can be distinct. The generalization from the case of uniform unitary errors can largely be handled by modifying the notation. We will consider the case of rotations about the  $z$  axis by angles  $\theta_j$  that differ between the qubits.

Stepping through the calculations above to find the effective channel for a given syndrome, we first write the error process on the  $n$  qubits as a sum over syndromes as follows:

$$\tilde{U} = \bigotimes_{i=1}^n U(\theta_i) = \sum_{w=0}^t \sum_{\mathbf{w}:|\mathbf{w}|=w} Z_{\mathbf{w}} [f_{\mathbf{w}}(\boldsymbol{\theta}) \bar{I} + f_{n-w}(\boldsymbol{\theta}) \bar{Z}]. \quad (35)$$

We have defined the following function:

$$f_{\mathbf{w}}(\boldsymbol{\theta}) = \prod_{j=1}^n (-i \sin \theta_j)^{w_j} \cos^{1-w_j}(\theta_j). \quad (36)$$

Consequently, the Kraus operator corresponding to the syndrome  $w$  is as follows:

$$K_w = f_{\mathbf{w}}(\boldsymbol{\theta}) \bar{I} + f_{n-w}(\boldsymbol{\theta}) \bar{Z}. \quad (37)$$

Now that the unitary errors are not uniform, the effective channels for each syndrome are distinct. The number of distinct effective channels has risen from  $t$  to  $2^{n-1}$ .

We can write these effective channels in terms of the parameters  $x_j$  and  $y_j$  as follows:

$$\bar{x}'_{\mathbf{w}} = \prod_{j=1}^n x_j^{w_j} (1-x_j)^{1-w_j}, \quad (38)$$

$$x'_{\mathbf{w}} = \prod_{j=1}^n x_j^{1-w_j} (1-x_j)^{w_j}, \quad (39)$$

$$y'_{\mathbf{w}} = (-1)^{t-w} \prod_{j=1}^n y_j. \quad (40)$$

As in our earlier examples, these expressions have been established only for unitary rotations about the  $z$  axis but hold also for the model with nonuniform dephasing included. In this case, the generalization can be established more straightforwardly. The map from the physical noise process, specified by the parameters  $\{x_j\}$ ,  $\{y_j\}$ , and  $\{\bar{x}_j = 1-x_j\}$  to the effective logical noise process for the syndrome  $w$ , given by  $\bar{x}'_{\mathbf{w}}, x'_{\mathbf{w}}, y'_{\mathbf{w}}$  is a polynomial where each term is  $n$ th order overall and first order in  $\bar{x}_j, x_j, y_j$  for each  $j$ . These polynomials are fixed uniquely by considering just the unitary case. We were not able to use this argument previously because in the unitary case  $y^2 = \bar{x}x$  which generally leads to an ambiguity in generalizing from the unitary to the general case using this method. This ambiguity is avoided here because no quadratics in the parameters  $\bar{x}_j, x_j, y_j$  arise. Having established these results for nonuniform noise without explicitly considering dephasing, we could reobtain the results of Sec. III A by specializing to the uniform case and then averaging over syndromes.

#### IV. EFFECTIVE CHANNELS WITH COHERENT ERRORS FOR GENERAL STABILIZER CODES

The same approach can be used to find effective channels for more general stabilizer codes, including the channel conditioned on a given syndrome. The main ideas have already arisen in the context of the repetition codes, but for a general stabilizer code there is a greater overhead of notation and some technical details. The main result of this section is that the diamond distance error for unitary noise processes after ideal correction scales like  $\theta^d$  where  $d$  is the distance of the code when a minimum distance decoder is used. Readers who are more interested in the specific effective channels may wish to skip ahead to the next section.

As discussed in the Appendix, a general product of Pauli matrices on  $n$  qubits can be specified by a binary vector  $\mathbf{b}$  of length  $2n$ . Ones in the first  $n$  entries specify the locations of  $X$  operators, while ones in the second group of  $n$  entries specify the locations of  $Z$  operators. A qubit acted on by both an  $X$  and a  $Z$  is understood to be acted on by  $Y$ . We will use the notation  $P_{\mathbf{b}}$  to indicate the resulting product of Paulis. If  $m_x$ ,  $m_y$ , and  $m_z$  are the numbers of  $X$ ,  $Y$ , and  $Z$  operators in the product, then the Hamming weight is  $w = m_x + m_y + m_z$ . We refer the reader to [30] for further information on stabilizer quantum error correcting codes. The discussion in the Appendix provides a brief summary and fixes our notation.

We now wish to consider general unitary rotations, so we have  $H = \alpha_x X + \alpha_y Y + \alpha_z Z$ . Since  $H^2 = I$  we have  $\alpha_x^2 + \alpha_y^2 + \alpha_z^2 = 1$ . We can therefore rewrite our overall unitary on  $n$  qubits (14) as follows:

$$\tilde{U} = \sum_{\mathbf{b}} f_w(\theta) \alpha^m P_{\mathbf{b}}, \quad (41)$$

where we have used the shorthand notation  $\alpha^m = \alpha_x^{m_x} \alpha_y^{m_y} \alpha_z^{m_z}$ .

As before, we can generalize to the case of nonuniform unitary noise straightforwardly. We obtain

$$\tilde{U} = \sum_{\mathbf{b}} g_{\mathbf{b}}(\boldsymbol{\theta}, \boldsymbol{\alpha}) P_{\mathbf{b}}, \quad (42)$$

with the definition

$$g_{\mathbf{b}}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \prod_{j=1}^n (-i \sin \theta_j)^{w_j(\mathbf{b})} \cos^{1-w_j(\mathbf{b})}(\theta_j) \times \alpha_{x_j}^{m_{x_j}(\mathbf{b})} \alpha_{y_j}^{m_{y_j}(\mathbf{b})} \alpha_{z_j}^{m_{z_j}(\mathbf{b})}. \quad (43)$$

Both the rotation angles  $\theta_j$  and the rotation axes, given by  $\alpha_{x_j}$ ,  $\alpha_{y_j}$ ,  $\alpha_{z_j}$ , now change from qubit to qubit. We have used a notation such that  $m_{x_j}(\mathbf{b})$  is one if  $P_{\mathbf{b}}$  has an  $X$  for qubit  $j$  and is zero otherwise.

Consider now a stabilizer code. The stabilizer generators are signed products of Pauli matrices  $\pm P_{\mathbf{g}_i}$  where the binary vectors  $\mathbf{g}_i$  satisfy certain constraints, for example, that the stabilizer generators all commute. A general member of the stabilizer group  $S_t$  is described by the length  $n - k$  binary vector  $\mathbf{t}$ , where those in the vector signify that the corresponding stabilizer generator is part of the product that results in  $S_t$ . The code has  $k$  logical  $X$  operators  $\bar{X}_i = P_{\mathbf{x}_i}$  and  $k$  logical  $Z$  operators  $\bar{Z}_i = P_{\mathbf{z}_i}$ . These logical operators are defined only up to multiplication by elements of the stabilizer group, but

here we will pick particular coset representatives. We can describe products of logical Pauli operators  $L_l$  by a length  $2k$  binary vector  $\mathbf{l} = (\mathbf{l}_x, \mathbf{l}_z)$ , in analogy to the Pauli matrices.

Now consider implementing the error correction with a specific decoder. First, the stabilizers are measured, and there are  $2^{n-k}$  syndromes. For a given Pauli error  $P_{\mathbf{b}}$ , we associate the syndrome  $\mathbf{y} = \mathbf{s}(\mathbf{b})$ . We will choose a fiducial Pauli error  $E_{\mathbf{y}}$  having the lowest possible weight among all errors that lead to that syndrome. There may be more than one choice for some syndromes, particularly for degenerate codes, and in that case we choose the lexicographically first choice of Pauli. We therefore define the *symmetric decoder* (or *minimum weight decoder*) to be the decoder that chooses the following correction:

$$E_{\mathbf{y}}^{\dagger} = P_{r(\mathbf{y})}, \quad r(\mathbf{y}) = \text{lex arg min}_{\mathbf{b}} \{|\mathbf{b}| : \mathbf{s}(\mathbf{b}) = \mathbf{y}\}, \quad (44)$$

where recall that  $|\mathbf{b}|$  is the Hamming weight of  $\mathbf{b}$ . We note that this decoder is generally not optimal, so some decoders could potentially do even better than suggested by our analysis.

Later in the paper, we will discuss two slight variants of this decoder. In Sec. IV we consider simplifications to the effective noise channels that arise from symmetries of the code. In that case, choosing the correction to be the lexicographically first choice of Pauli with the required syndrome and lowest possible weight can result in a correction procedure that has less symmetry than the code itself. An example is the Steane code, where certain two-qubit errors can be corrected. If these are chosen to be the two-qubit Paulis made up of one  $X$  and one  $Z$  error, then the effective noise channels simplify. In such examples it is possible to specify the corrections  $E_{\mathbf{y}}$  out of the set of possible lowest weight corrections for that syndrome so as to preserve the symmetry of the code. A second modification of the decoder is used in Sec. VI where we consider noise processes that are combinations of dephasing and  $Z$  rotations. Since for these noise processes only  $Z$  errors can occur, we can improve the performance of the code by choosing  $E_{\mathbf{y}}$  to be the lowest weight product of Pauli  $Z$  errors that could produce the syndrome  $\mathbf{y}$ . If no product of  $Z$ 's results in the syndrome we can, as before, choose the lowest weight correction, however, this syndrome will not arise.

Having specified a decoder we know that for any Pauli matrix  $P_{\mathbf{b}}$  there is a unique syndrome  $\mathbf{s}(\mathbf{b})$  and correction  $E_{\mathbf{s}(\mathbf{b})}^{\dagger}$  such that  $E_{\mathbf{s}(\mathbf{b})}^{\dagger} P_{\mathbf{b}}$  commutes with the stabilizer group. Thus,  $E_{\mathbf{s}(\mathbf{b})}^{\dagger} P_{\mathbf{b}}$  is in the normalizer group that is generated by the logical operators and the stabilizer group (see for example [30]). Thus, there exists a logical operator  $L_{\mathbf{l}(\mathbf{b})}$ , stabilizer  $S_{\mathbf{t}(\mathbf{b})}$ , and phase factor  $\eta_{\mathbf{b}}$  equal to  $\pm 1$  or  $\pm i$  such that  $E_{\mathbf{s}(\mathbf{b})}^{\dagger} P_{\mathbf{b}} = \eta_{\mathbf{b}} L_{\mathbf{l}(\mathbf{b})} S_{\mathbf{t}(\mathbf{b})}$ . Consequently, every Pauli matrix can be written uniquely in the form  $P_{\mathbf{b}} = \eta_{\mathbf{b}} E_{\mathbf{s}(\mathbf{b})} L_{\mathbf{l}(\mathbf{b})} S_{\mathbf{t}(\mathbf{b})}$ . This mapping is a property of the chosen error correcting code, encoding unitary and decoding schemes. Equally for any choice of syndrome  $\mathbf{s}$ , logical operator  $\mathbf{l}$ , and stabilizer element  $\mathbf{t}$  their product is a Pauli matrix up to a phase so that there is a unique  $\mathbf{b}(\mathbf{s}, \mathbf{l}, \mathbf{t})$  and  $\eta_{\mathbf{s}, \mathbf{l}, \mathbf{t}}$  such that  $P_{\mathbf{b}} = \eta_{\mathbf{s}, \mathbf{l}, \mathbf{t}} E_{\mathbf{s}} L_{\mathbf{l}} S_{\mathbf{t}}$ . So, the mapping between  $\mathbf{b}$  and  $(\mathbf{s}, \mathbf{l}, \mathbf{t})$  is one-to-one and onto. Note that the  $4^n$  Pauli products  $P_{\mathbf{b}}$  are accounted for since there are  $2^{n-k}$  syndromes,  $4^k$  logical operators, and  $2^{n-k}$  stabilizer



elements, which can together result in  $2^{n-k} \times 4^k \times 2^{n-k} = 4^n$  distinct products.

With these preliminaries completed, we can now rewrite the expression for  $\tilde{U}$  from Eq. (41) as a sum over syndromes, rather than a sum over products of Paulis:

$$\tilde{U} = \sum_s \sum_l \sum_t f_{w(s,l,t)}(\theta) \alpha^{m(s,l,t)} \eta_{s,l,t} E_s L_l S_t. \quad (45)$$

This expansion for  $\tilde{U}$  makes it very straightforward to read off the Kraus operators of the effective channel. Since we start in the code space, the stabilizers in this product act trivially. The syndrome measurement projects the system onto a single value of  $s$  and the correction  $E_s^\dagger$  removes the fiducial error for that syndrome. Consequently, we obtain  $2^{n-k}$  Kraus operators indexed by  $s$  acting on the logical qubit as follows:

$$K_s = \sum_l \sum_t f_{w(s,l,t)}(\theta) \alpha^{m(s,l,t)} \eta_{s,l,t} L_l. \quad (46)$$

The nonuniform case is nearly identical:

$$K_s = \sum_l \sum_t g_{b(s,l,t)}(\theta, \alpha) \eta_{s,l,t} L_l. \quad (47)$$

This expression can be used to study general properties of the effective channel. We will restrict our discussion to the following theorem.

*Theorem 1.* For any  $[n, k, d]$  stabilizer code, the diamond norm error  $D'_\diamond$  of the logical qubits after ideal error correction using a minimum weight decoder satisfies

$$D'_\diamond \leq c_{n,k} D_\diamond^d \quad (48)$$

for independent unitary errors with rotation angles satisfying  $|\theta_i| \leq \pi/4$ , where  $D_\diamond = \max_i |\sin \theta_i|$ . The constant  $c_{n,k}$  depends on  $n$  and  $k$  but is independent of the errors.

*Proof.* Our result is analogous to Theorem 5.5 of [31] when applied to unitary errors, and our proof follows the contours of that theorem. We will explicitly consider the case of uniform errors. The generalization to nonuniform errors is a straightforward modification using the approach of Sec. III C and Eq. (47) for the Kraus operator arising from the syndrome  $s$  with nonuniform unitary noise.

We must first explicitly separate the nontrivial logical operators from the identity in the expansion over syndromes. So, consider

$$\begin{aligned} \tilde{U} &= \sum_s \sum_t f_{w(s,0,t)}(\theta) \alpha^{m(s,0,t)} \eta_{s,0,t} E_s \bar{I} S_t \\ &+ \sum_s \sum_{l \neq 0} \sum_t f_{w(s,l,t)}(\theta) \alpha^{m(s,l,t)} \eta_{s,l,t} E_s L_l S_t. \end{aligned} \quad (49)$$

We have written  $L_l$  in the second term above just to emphasize that the logical identity operator does not arise in the sum. We seek to understand the relative size of the terms in this expansion. We have restricted attention to  $|\theta| \leq \pi/4$  so that higher weight errors are less likely than lower weight errors. Notice that in this regime we have  $|\tan \theta| \leq 1$  and therefore  $|f_w(\theta)| \leq |f_{w'}(\theta)|$  when  $w \geq w'$ . On the other hand, all the factors  $\alpha^m$  satisfy  $|\eta_b \alpha^m| \leq 1$  and the operator norm of the product of Pauli matrices  $E_s L_l S_t$  is  $\leq 1$ .

We now analyze the Hamming weights of the various terms in Eq. (49). In each term of this sum we will specify the

weight of the correction operator  $E_s^\dagger$  by  $w_s$ . Considering the first of the two terms in Eq. (49), the identity is a member of the stabilizer group, so that one of the terms is just  $E_s$  itself and has Hamming weight  $w_s$ . The other terms in the sum are of the form  $E_s S_t$  and each of these terms is a possible error process that leads to the same syndrome  $s$ . But, we have chosen our decoder such that  $w_s$  is the lowest possible weight for an error with this syndrome. So, all the contributions to the first term in Eq. (49) have  $w \geq w_s$ . Since we have a quantum error correcting code, all errors up to and including some Hamming weight  $t$  will be corrected by the code. For odd  $d$  we have  $d = 2t + 1$ , while for even  $d$ ,  $d = 2t + 2$ . Every product of Paulis of Hamming weight  $w \leq t$  is either a member of the stabilizer group, or one of the fiducial errors  $E_s$ , or is equivalent to some  $E_s$  up to multiplication by a stabilizer operator. Consequently, every product of Paulis with Hamming weight  $w \leq t$  occurs somewhere in the first term of Eq. (49). Considering now the second term in Eq. (49), clearly we have  $w \geq t + 1$  for all the contributions, regardless of  $s$ . For specific values of  $s$  we can be more precise. For each syndrome, the factor  $L_l S_t$  is a nontrivial logical operator and therefore has weight at least  $d$ . Therefore, in each contribution to the syndrome  $s$  the product of Pauli matrices  $E_s L_l S_t$  has weight  $w \geq d - w_s$ . (Note that in examples like the Steane code it is possible to correct certain errors  $E_s$  with  $w_s > t$ . For these syndromes,  $w \geq t + 1$  will be a better bound on the Hamming weight of  $E_s L_l S_t$ .)

Consequently, we can consider the corresponding expansion of  $K_s$ :

$$\begin{aligned} K_s &= \sum_t f_{w(s,0,t)}(\theta) \alpha^{m(s,0,t)} \eta_{s,0,t} \bar{I} \\ &+ \sum_{l \neq 0} \sum_t f_{w(s,l,t)}(\theta) \alpha^{m(s,l,t)} \eta_{s,l,t} L_l'. \end{aligned} \quad (50)$$

In the first term we have  $w \geq w_s$  and in the second we have  $w \geq \max(d - w_s, t + 1)$ . The overall effective channel is

$$\mathcal{N}_L(\rho) = \sum_s K_s \rho K_s^\dagger = \sum_{l,l'} r_{l,l'} L_l \rho L_{l'}^\dagger. \quad (51)$$

Specifically,

$$r_{l,l'} = \sum_{s,t,t'} f_{w(s,l,t)}(\theta) f_{w(s,l',t')}(\theta) \alpha^{m(s,l,t)} \alpha^{m(s,l',t')} \eta_{s,l,t} \eta_{s,l',t'}. \quad (52)$$

This is a trace-preserving, unital, completely positive map since the encoding, recovery, and decoding maps are all unital and trace preserving.

If we specify that both  $l$  and  $l'$  are nontrivial so  $L_l, L_{l'} \neq \bar{I}$ , then

$$\begin{aligned} |r_{l,l'}| &\leq \sum_{s,t,t'} |f_{w(s,l,t)}(\theta) f_{w(s,l',t')}(\theta)| \leq 8^{n-k} |f_{t+1}(\theta)|^2 \\ &\leq 8^{n-k} |f_{2t+2}(\theta)| \leq 8^{n-k} |f_d(\theta)|. \end{aligned} \quad (53)$$

The first inequality is just the triangle inequality combined with the fact that  $|\alpha^m| \leq 1$  and  $\eta_{s,l,t} \leq 1$ . The second inequality follows from the fact that each  $w$  appearing in the sum satisfies  $w \geq t + 1$  and so  $|f_w| \leq |f_{t+1}|$  when

$|\theta| \leq \pi/4$ . The prefactor just counts the number of terms in the sum over the syndromes and two copies of the stabilizer group. To see the third inequality notice that if  $w + w' \leq n$ , then  $|f_w f_{w'}| = |\sin(\theta)|^{w+w'} |\cos(\theta)|^{2n-w-w'} \leq |\sin(\theta)|^{w+w'} |\cos(\theta)|^{n-w-w'} = |f_{w+w'}|$ .

Applying the same logic to the case where  $\mathbf{l}' = \mathbf{0}$ , so that  $L_{\mathbf{l}'} = \bar{I}$ , and  $\mathbf{l} \neq \mathbf{0}$  we see that

$$\begin{aligned} |r_{\mathbf{l}, \mathbf{0}}| &\leq \sum_{s, \mathbf{l}, \mathbf{l}'} |f_{w(s, \mathbf{l}, \mathbf{l}')}(\theta) f_{w(s, \mathbf{0}, \mathbf{l}')}(\theta)| \\ &\leq 4^{n-k} \sum_s |f_{w_s}(\theta) f_{d-w_s}(\theta)| \leq 8^{n-k} |f_d(\theta)|. \end{aligned} \quad (54)$$

The second inequality here results from the fact discussed above that for each term in the sum  $w(s, \mathbf{l}, \mathbf{l}') \geq d - w_s$  and  $w(s, \mathbf{0}, \mathbf{l}') \geq w_s$ . Clearly, the case of  $r_{\mathbf{0}, \mathbf{l}'}$  with  $\mathbf{l}' \neq \mathbf{0}$  is the same.

Finally, we can bound  $r_{\mathbf{0}, \mathbf{0}}$  since  $\mathcal{N}_L(\rho)$  is trace preserving and unital and so  $\sum_{\mathbf{l}} r_{\mathbf{l}, \mathbf{l}} = 1$ , and therefore

$$|r_{\mathbf{0}, \mathbf{0}} - 1| \leq (4^k - 1) 8^{n-k} |f_d(\theta)|. \quad (55)$$

The result follows by applying the triangle inequality and the fact that if  $\mathcal{N}_{\mathbf{l}, \mathbf{l}'}(\rho) = L_{\mathbf{l}} \rho L_{\mathbf{l}'}^\dagger$  then  $\|\mathcal{N}_{\mathbf{l}, \mathbf{l}'}\|_\diamond = 1$ . This last statement follows from the characterization of the diamond norm in terms of the maximum output fidelity in [32, Theorem 5]. ■

A couple of remarks are in order. First, collecting the constant factors and simplifying, we see that the statement holds with  $c_{n,k} = 2^{3n+k} + 2^{3n-k} \leq 2^{3n+k+1}$ , but no attempt was made to optimize the constant and significantly better bounds might exist. Second, we note that most of the error correcting codes that we have studied saturate this inequality in the sense that the diamond norm distance after error correction scales as  $|\theta|^d$  for a small initial unitary angle, regardless of code or rotation axis. Only for certain codes and specific rotation axes have we observed a more favorable scaling. Third, our theorem bounds the diamond norm error in terms of  $\theta^d$  for all stabilizer codes. However, for large distance codes operating close to threshold, this bound will be very weak. For large distance codes, the expected  $\theta^d$  dependence of the effective channel will occur but only for very small values of  $\theta$ , very far below threshold. We also note that the intermediate stages of our proof reproduce the observations about unitary channels made at the level of the process matrix by Gutiérrez *et al.* [11]. Finally, we note that a related version of this result in terms of the average gate infidelity was recently proven independently by Beale *et al.* [33].

### Effective channels for symmetric codes

It has been observed that the effective channels for widely studied error correcting codes, such as the five-qubit and Steane codes, typically have many fewer distinct conditional channels than there are syndromes of the code [29]. We saw an example of this for the repetition code where, due to the permutation symmetry of the code, the conditional channel just depends on the weight of the correction operator and not on the detailed syndrome for uniform noise. Clearly, symmetries of the error correcting code can simplify the effective channels that arise for uniform noise. Here, we briefly discuss

these simplifications. Symmetries of stabilizer codes have been studied previously (see for example [34–37]).

We will largely focus on the automorphism group of the code under permuting qubits, but we can formulate the notion of symmetry more generally. For our purposes, a stabilizer code together with its error correction procedure will have interesting symmetries when there is some subgroup  $\mathbf{G}$  of the Clifford group acting on the physical qubits that preserves the error correction procedure. We will say that an error correcting code is symmetric under  $\mathbf{G}$  if the logical operators can be chosen to be invariant under  $\mathbf{G}$  and the stabilizer group is preserved under the action of  $\mathbf{G}$ . Specifically, for all  $G \in \mathbf{G}$  and stabilizer elements  $\mathbf{t}$  there exists some  $\mathbf{t}'$  such that  $GS_{\mathbf{t}}G^\dagger = S_{\mathbf{t}'}$ , while for all  $G \in \mathbf{G}$  and  $\mathbf{l}$  we have  $GL_{\mathbf{l}}G^\dagger = L_{\mathbf{l}}$ . We will say that the decoding procedure is symmetric if the code is symmetric and for all  $G$  and  $\mathbf{t}$  there is some  $\mathbf{t}'$  such that  $GE_{\mathbf{t}}G^\dagger = E_{\mathbf{t}'}$ . Up until now we have used the term symmetric decoder if the correction  $E_{\mathbf{t}}^\dagger$  has the lowest weight of any error that could have caused the syndrome  $\mathbf{t}$ . We now place this further requirement on the choice of correction.

When the error correction procedure is symmetric, the syndromes will break up into orbits under the action of  $\mathbf{G}$ . Rather than considering general Clifford symmetries of stabilizer codes, we will now specialize to the case where the symmetry of the code is some group of permutations of physical qubits. Qubit permutations are very special because they preserve the Hamming weight and  $\alpha^n$ . Thus, if we perform a permutation  $GP_bG^\dagger = P_{b'}$ , then  $w(\mathbf{b}') = w(\mathbf{b})$  and so on. We will see that conditional channels of all syndromes in a given orbit are equal for such permutation symmetries of stabilizer codes. This explains the simplifications observed in Sec. III for the repetition code and in [29] for a range of other codes.

To demonstrate that the effective channels for syndromes in a single orbit are equal, consider a fixed syndrome  $s$  and a fixed permutation  $G \in \mathbf{G}$ . Then, we define  $s'$  such that  $GE_sG^\dagger = E_{s'}$ . Now, let us consider a single term in Eq. (45) that has the chosen  $s$ , we have

$$P_b = \eta_{s, \mathbf{l}, \mathbf{l}'} E_s L_{\mathbf{l}} S_{\mathbf{l}'} \quad (56)$$

and if we define  $\mathbf{b}'$  such that  $GP_bG^\dagger = P_{b'}$  we have

$$\begin{aligned} P_{b'} &= \eta_{s, \mathbf{l}, \mathbf{l}'} (GE_sG^\dagger)(GL_{\mathbf{l}}G^\dagger)(GS_{\mathbf{l}'}G^\dagger) \\ &= \eta_{s, \mathbf{l}, \mathbf{l}'} E_{s'} L_{\mathbf{l}} S_{\mathbf{l}'} = \eta_{s', \mathbf{l}, \mathbf{l}'} E_{s'} L_{\mathbf{l}} S_{\mathbf{l}'}. \end{aligned} \quad (57)$$

The second equality just arises from the definitions and the fact that the error correction procedure is symmetric under  $\mathbf{G}$ . The third equality holds because  $\eta$  can be written in terms of the commutators of the various factors, and these are preserved under the action of the unitary. (Alternatively, it can be written in terms of the symplectic inner product which is manifestly invariant under swapping qubits; see the Appendix for more details.) Moreover, we have

$$w(s, \mathbf{l}, \mathbf{l}') = w(\mathbf{b}) = w(\mathbf{b}') = w(s', \mathbf{l}, \mathbf{l}') \quad (58)$$

since permutations leave the weight unchanged, and the same identity holds for  $\alpha^n$ . The mapping takes  $S_{\mathbf{l}}$  to  $S_{\mathbf{l}'}$  under the group  $\mathbf{G}$  so in the formula for the Kraus operator of the conditional channel (46) each term in the sum for  $K_s$  maps invertibly to a single term in the sum for  $K_{s'}$ . Consequently,

we have

$$K_s = \sum_l \sum_{l'} f_{w(s,l,l')}(\theta) \alpha^{n(s,l,l')} \eta_{s,l,l'} L_l \quad (59)$$

$$= \sum_l \sum_{l'} f_{w(s',l,l')}(\theta) \alpha^{n(s',l,l')} \eta_{s',l,l'} L_l = K_{s'}. \quad (60)$$

Thus, all the Kraus operators corresponding to syndromes in a given orbit of the symmetry are equal.

We have shown these simplifications only for unitary noise at the physical level. It would be interesting to extend this analysis to noise processes with more than a single Kraus operator. In the case of repetition codes with an odd number of qubits, we can choose the code and decoding procedure to have the full permutation symmetry of the  $d$  qubits. This implies that the Kraus operators can only depend on the Hamming weight of the correction procedure, as we observed in Sec. III.

In the case of the five-qubit code [38], the stabilizer group is generated by the operators  $XZZXI$  and its cyclic permutations and the logical operators can be chosen to be  $\bar{X} = XXXXX$  and  $\bar{Z} = ZZZZZ$ . It is clear that this code is symmetric under cyclic permutations of the qubits. It is also symmetric under the permutation  $(1, 5)(2, 4)$ , which is a reflection if the five qubits are arranged on the vertices of a pentagon. This generates the group  $D_{10}$  of symmetries of the pentagon. (We believe that this is the full symmetry group of the code.) The five-qubit code is a  $[5,1,3]$  code and there are  $2^{n-k} = 16$  distinct syndromes. In the symmetric decoder we choose correction operations that are the lowest weight Paulis that can result in each syndrome. These are the identity  $IIIII$ , and the 15 single-qubit Pauli's. This set of correction operators is symmetric under the full permutation group, so we can conclude that the overall error correction procedure is symmetric under  $D_{10}$ . The correction operations break into four orbits under this group. Representatives of these orbits are  $IIIII$ ,  $XIIII$ ,  $YIIII$ , and  $ZIIII$ . Consequently, there are four distinct conditional effective channels for the five-qubit code under the symmetric decoder. This agrees with the number of conditional effective channels in [29].

For the Steane code [39] we can use similar arguments. This is a CSS code with its  $X$ -type stabilizers generated by  $XXXIXII$ ,  $XIXXIXI$ , and  $XXIXIIX$ . The  $Z$ -type stabilizer generators are obtained by replacing  $X$  with  $Z$ . The logical operators can be chosen to be  $\bar{X} = XXXXXX$  and  $\bar{Z} = ZZZZZZ$ . The code is symmetric under the permutations  $(2, 3)(6, 7)$ ,  $(2, 3, 4)(5, 6, 7)$ ,  $(1, 3)(2, 5)$ , and  $(1, 3)(4, 6)$ . These permutations generate the 168-element group  $GL(3, 2)$ , which is the group of symmetries of the Fano plane. One can check by exhaustive search that this is the full symmetry group of the code. In order to identify the symmetry group we found it useful to picture the Steane code as the smallest possible color code. From this point of view, the qubits in the code correspond one-to-one with the points of the Fano plane. Moreover, the lines of the Fano plane correspond to the logical operators with weight 3. (It would be interesting to know whether some relationship of this kind holds between other color codes and other finite geometries.) Since the Steane code is a  $[7,1,3]$  code, there are  $2^{n-k} = 64$  distinct syndromes. In the symmetric decoder, the

correction operations are the identity  $IIIIIII$ , the 21 single-qubit Paulis, and the 42 Paulis that are the tensor product of a single  $X$  with a single  $Z$ . This set of correction operators is symmetric under the full permutation group, so we can conclude that the overall error correction procedure is symmetric under  $GL(3, 2)$ . The correction operations break into five orbits under this group. Representatives of these orbits are  $IIIIIII$ ,  $XIIIIII$ ,  $YIIIIII$ ,  $ZIIIIII$ , and  $XZIIIII$ . This number of orbits does not agree with the seven inequivalent conditional effective channels stated in [29] for general noise processes. However, our result only applies to unitary channels (which may have higher symmetries than the more general channels considered in [29]) and it is not clear whether the decoder in [29] has the highest possible symmetry. So, we expect that these statements are not in disagreement.

Note that these simplifications are independent of the exact unitary error that has occurred and for certain highly symmetric noise processes it is possible that there are further reductions in the number of distinct conditional effective channels. We will see that an example of this is the five-qubit code with unitary errors of the form  $\exp[-i\theta(X + Y + Z)/\sqrt{3}]$ .

For specific examples that we considered there are still patterns in the expressions for the conditional effective channels that arise from the structure of the code. One instance where such examples arise is when the code has a nontrivial single-qubit transversal gate. Such a transversal gate maps the stabilizer group to itself but acts nontrivially on the logical operations, the corrections  $E_s$ , and the factors  $\alpha^n$ . In the example of the five-qubit code, the Clifford gate that maps  $X \rightarrow Y \rightarrow Z \rightarrow X$  is transversal. This makes it possible to infer the conditional effective channel for an error in the orbit given by  $YIIII$  in terms of the one for errors in the orbit containing  $XIIII$ .

## V. PERFORMANCE OF STABILIZER CODES UNDER COHERENT ERRORS

In this section we will review the performance of a variety of quantum error correcting codes against coherent errors using the techniques developed so far. Recall that for a unitary error we have  $D_\diamond \approx 3\sqrt{r}/2 \approx |\theta|$  so that there is a large difference between the size of errors as estimated in randomized benchmarking and errors as measured by diamond norm distance. Characteristically we find that after error correction the diamond norm error of the effective channel  $D'_\diamond$  is much smaller than would be expected based on the diamond norm error at the physical level. For dephasing errors,  $D'_\diamond \propto D_\diamond^{t+1}$  for a small error. On the other hand, we have seen that for a unitary error  $D'_\diamond \leq cD_\diamond^d$ .

The difference in this behavior is almost sufficient to overcome the distinction between diamond norm error and average gate infidelity at the physical level. If  $r$  is the average fidelity error at the physical level, then for an odd distance code we find  $D'_\diamond \leq c(3/2)^{2t+1} r^{t+1/2}$  for unitary errors, while for dephasing errors  $D'_\diamond \propto r^{t+1}$ . For even distance codes this comparison is even more striking, we find  $D'_\diamond \leq c(3/2)^{2t+2} r^{t+1}$  for unitary errors, compared to  $D'_\diamond \propto r^{t+1}$  for dephasing errors. So, the scaling with  $r$  is the same for both unitary and dephasing errors. We emphasize that these results assume ideal error correction, and do not analyze

fault-tolerant gadgets and noisy gates and measurements. They do suggest that there is a need for a sharper analysis of the performance of fault-tolerant gadgets against unitary errors.

The remainder of this section explores the logical effective channels and the logical diamond norm error scaling relative to the physical diamond norm error and the physical average gate infidelity for a variety of small quantum codes using pure unitary errors around arbitrary axes. In the next section, we will also consider the more general noise model that includes dephasing and demonstrate with the Steane code how the computation of the effective logical channel can be used to find thresholds for concatenated codes with the symmetric decoder.

### A. Repetition codes

In the limit of low error  $x, y \ll 1$ , we can compute simple approximations for the diamond norm error of the effective channel of the repetition codes discussed in Sec. III. For pure dephasing noise we have  $y = 0$  and  $D'_\diamond \simeq \binom{n}{t+1} x^{t+1} = \binom{n}{t+1} D_\diamond^{t+1} = \binom{n}{t+1} (3/2)^{t+1} r^{t+1}$  which is the expected scaling with error probability, physical diamond norm error, and average gate infidelity, respectively.

For unitary errors we have  $y^2 = x(1-x)$  and it is possible to see that  $D'_\diamond$  is dominated by the contribution arising from  $y'$ . Consequently, we have  $D'_\diamond \simeq \binom{2t}{t} |y|^{2t+1} \simeq \binom{2t}{t} D_\diamond^{2t+1} \simeq \binom{2t}{t} (3/2)^{t+1/2} r^{t+1/2}$ , which is the scaling suggested by Theorem 1. Notice that these results imply that a simple Pauli twirling approximation to a unitary error, which just sets  $y = 0$ , would underestimate the logical error, finding  $D'_{PTA_\diamond} \simeq \binom{2t+1}{t+1} |\theta|^{2t+2}$  in the limit of small  $|\theta|$ , as opposed to  $\binom{2t}{t} |\theta|^{2t+1}$ . In the terminology of [40], this implies that the Pauli twirling approximation to the effective noise channel is *dishonest* but it is less dishonest than when the PTA is applied at the logical level.

### B. Effective channels for unitary errors: The five-qubit code and others

As discussed in Sec. IV, the five-qubit code is symmetric under cyclic permutation of the qubits of the code, and also possesses a symmetry resulting from the transversal Clifford gate operation that maps  $X \rightarrow Y \rightarrow Z \rightarrow X$ . Consequently, we begin by considering unitary errors of the form  $\exp[-i\theta(X+Y+Z)/\sqrt{3}]$  which enhances the symmetry of the various effective error channels. The five-qubit code has 16 syndromes, and in this highly symmetric case there are only two conditional effective channels, one corresponding to the trivial syndrome. All the nontrivial syndromes result in the same effective channel. These two channels are each described by a single Kraus operator  $K_0$  in the trivial case and  $K_1$  in the nontrivial case. They can be written as follows:

$$K_0 = (g_0 + 15g_4)\bar{I} - (10g_3 - 6g_5)(\bar{X} + \bar{Y} + \bar{Z}), \quad (61)$$

$$K_1 = (g_1 + 4g_3 + 3g_5)\bar{I} - (2g_2 + 2g_4)(\bar{X} + \bar{Y} + \bar{Z}), \quad (62)$$

where  $g_w = f_w/(\sqrt{3})^w$ .

Inspecting the expressions for  $K_0$  and  $K_1$  we can see that this behavior is qualitatively very similar to what we observed

for the repetition codes. It is straightforward to check that both  $K_0$  and  $K_1$  are proportional to a unitary and correspond to rotation about the same axis  $(X+Y+Z)/\sqrt{3}$  as occurs at the physical level. In the case of a small rotation error  $\theta \ll 2\pi$ , we find that the trivial syndrome occurs with high probability and results in a rotation  $\theta'_0 \simeq 10\theta^3/3$ . A nontrivial syndrome occurs with probability  $\simeq 15\theta^2$  and results in a rotation  $\theta'_1 \simeq -2\theta$ . Notice that conditioned on a nontrivial syndrome, the rotation angle has actually increased. Nevertheless, this pattern is exactly the one seen in the repetition code, and saturates the scaling limit of Theorem 1 since we have  $D'_\diamond \simeq cD_\diamond^3$  for small  $\theta$ . The qualitative behaviors that we observed for the repetition codes and special rotation axes are also seen in proper stabilizer codes for typical rotation axes.

Consider now a rotation about the  $Z$  axis. In this case, we find four inequivalent Kraus operators. The first corresponds to the trivial syndrome. There are then five syndromes that detect single  $Z$  errors and result in a effective error that is rotation about the  $Z$  axis. The other two classes of syndromes would generally detect single  $Y$  and  $X$  errors. But, since these do not arise for this noise model, they detect two-qubit  $Z$  errors of the form  $ZZIII$  and  $ZIZII$ , respectively. The resulting Kraus operators are

$$K_0 = f_0\bar{I} + f_5\bar{Z}, \quad (63)$$

$$K_1 = f_1\bar{I} + f_4\bar{Z}, \quad (64)$$

$$K_2 = -f_2\bar{Y} - if_3\bar{X}, \quad (65)$$

$$K_3 = -f_2\bar{X} + if_3\bar{Y}. \quad (66)$$

Here,  $K_1$  is the conditional channel that results when the correction operation is a  $Z$  operation,  $K_2$  corresponds to a  $Y$  operation, and  $K_3$  is an  $X$  operation. The Kraus operators are proportional to unitaries, but this time the rotation axis is not necessarily the same as the original unitary. Notice that from these conditional effective channels we find  $D'_\diamond \simeq cD_\diamond^4$  so that unitary errors about this axis are much better corrected than for a typical rotation axis.

We see that for this Pauli axis rotation, the five-qubit code behaves as if it corrects two errors rather than one, and if we modified the correction operation to correct for two-qubit  $Z$  errors it would be possible to correct the unitary rotation such that  $D'_\diamond \simeq cD_\diamond^5$ . This is consistent with the fact that the five-qubit code has distance 5 if there are only  $Z$  errors, as it becomes a repetition code in this limit. Note that this improvement requires that the rotation axis of the unitary is known.

We will not write the effective channels for a general rotation axis, but it is illuminating to see how the performance of the code varies as the rotation axis is changed. This is shown in Fig. 1, where  $D'_\diamond/D_\diamond^d$  is plotted for various rotation axes. The improved performance for rotations about the Pauli axes is clearly visible. It can be seen that rotations about axes like  $(X+Y+Z)/\sqrt{3}$  are local maxima for the diamond norm error after error correction.

We can study the performance of many other stabilizer codes in the same way. The general behavior is similar to the five-qubit code, although for typical codes the scaling of



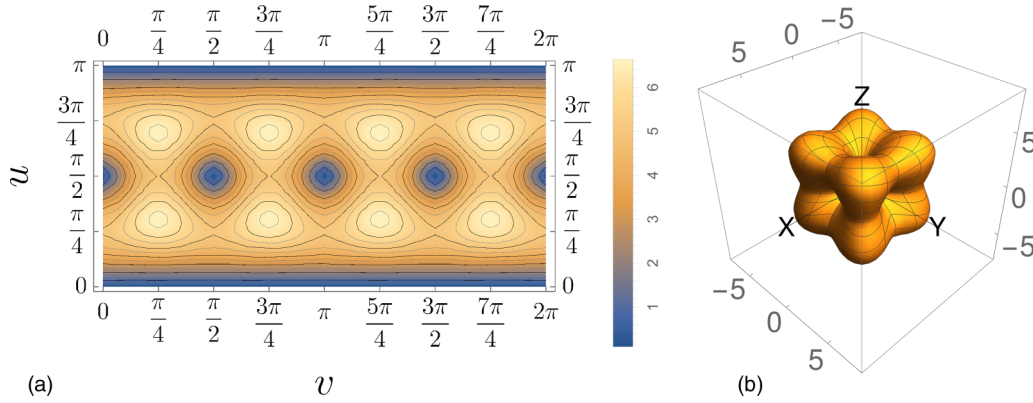


FIG. 1. (a) Contour plot of the ratio  $D'_o/D_o^d$  against the rotation axis of a single-qubit unitary error for the [5,1,3] five-qubit code using the symmetric decoder. The polar angle is  $u$  and the azimuthal angle is  $v$ . The rotation angle for the unitary error was chosen to be  $\theta = 0.01$ . (b) 3D spherical plot of the ratio  $D'_o/D_o^d$  for the [5,1,3] five-qubit code calculated as for (a). Rotations about Pauli axes are suppressed at a higher order than all other axes.

$D'_o$  with  $D_o^d$  holds for all rotation axes. This is true of the Steane code, for example, as shown in Fig. 2(a), where we see improved performance for rotations about  $(X + Y + Z)/\sqrt{3}$  and reduced performance for rotations about the Pauli axes.

We provide a few further examples of the dependence of the effective channel on the rotation axis of the unitary which arises when using the symmetric decoder. In Fig. 2(b) we show the analogous plot for the Bare [7,1,3] code of Li *et al.* [41] and in Fig. 2(c) the [9,1,3] Shor code [42]. In the case of the Shor code, errors for rotations about  $Z$  are corrected less well than rotations about  $X$  due to the structure of this code as a concatenation of codes correcting bit flips and phase flips. This concatenation structure means that the subgroup of the stabilizer group made up of products of  $X$  operators is larger than the subgroup made of products of  $Z$ 's. Lastly, we show the [9,1,3] surface code [43] in Fig. 2(d). Here again we see that there is enhanced correction along  $(X + Z)/\sqrt{2}$  and especially along  $Y$ . This is consistent with the large increase in the error correction threshold observed for incoherent  $Y$  noise in the surface code when using an optimal decoder [26]. Unfortunately, we do not have a general understanding of the symmetry axes that show enhanced performance for all of the codes that we plot in Fig. 2, and it remains an intriguing open problem to better explain or predict such behavior in the most general case.

## VI. CONCATENATED CODE THRESHOLDS FOR JOINT UNITARY AND DEPHASING ERRORS

We now return to the noise model of Eq. (5), which combines  $Z$ -axis rotation with dephasing. Whereas previously we

have always used the symmetric, minimum distance decoder, we now specialize to a  $Z$ -only decoder that finds the minimum distance correction among all Paulis consisting only of  $Z$  and  $I$ . This is sensible since for these noise channels only  $Z$ -type errors are supported.

For any given code we can regard the ideal error correction as a map on the noise channel. For most of the codes we have investigated we find that the form of the noise is unchanged and there exist some  $(x', y')$  such that the error correction implements the map  $(x, y) \rightarrow (x', y')$ . This type of calculation was also performed by Gutiérrez *et al.* [11], but in contrast we do not need the full process matrix since our noise model is restricted to Eq. (5) and so is a function solely of  $x$  and  $y$ . For example, for the Steane code we find

$$x' = 21x^2 - 98x^3 + 210x^4 - 252x^5 + 168x^6 - 48x^7 + 42y^4 - 252xy^4 + 504x^2y^4 - 336x^3y^4, \quad (67)$$

$$y' = 14y^3 - 168xy^3 + 504x^2y^3 - 672x^3y^3 + 336x^4y^3 + 48y^7, \quad (68)$$

so to lowest order, under unitary noise only,  $p' \approx 63\theta^4$  and  $\theta' \approx 14\theta^3$ , which is consistent with Theorem 1.

This procedure can handle rather larger codes. As a further example, consider the [16,1,4] surface code described in [44]. We consider the noise model of Eq. (5) and perform correction by measuring  $X$  stabilizers only. With the  $Z$ -only decoder, we find  $y' = 0$  and

$$\begin{aligned} x' = & 32x^2 - 188x^3 + 484x^4 - 500x^5 - 612x^6 + 3136x^7 - 5680x^8 + 6080x^9 - 4032x^{10} + 1536x^{11} - 256x^{12} + 28xy^2 \\ & - 222x^2y^2 + 756x^3y^2 - 1378x^4y^2 + 1104x^5y^2 + 752x^6y^2 - 2880x^7y^2 + 3120x^8y^2 - 1600x^9y^2 + 320x^{10}y^2 \\ & + 12y^4 - 96xy^4 + 208x^2y^4 + 416x^3y^4 - 3088x^4y^4 + 7040x^5y^4 - 8320x^6y^4 + 5120x^7y^4 - 1280x^8y^4 - 10y^6 \\ & - 32xy^6 + 704x^2y^6 - 2880x^3y^6 + 5280x^4y^6 - 4608x^5y^6 + 1536x^6y^6 - 48y^8 + 384xy^8 - 1152x^2y^8 + 1536x^3y^8 \\ & - 768x^4y^8 - 16y^{10} + 64xy^{10} - 64x^2y^{10}. \end{aligned} \quad (69)$$

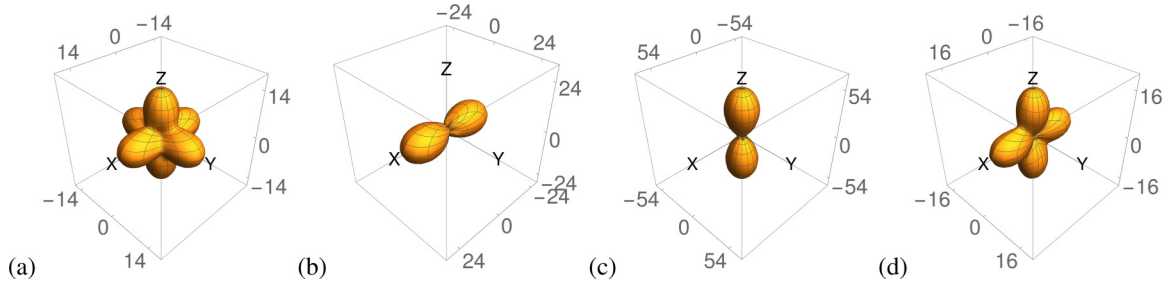


FIG. 2. 3D spherical plots of the ratio  $D'_\diamond/D_\diamond^d$  against the rotation axis of a single-qubit unitary error for various error correcting codes using the symmetric decoder. The rotation angle for the unitary error was chosen to be  $\theta = 0.01$ . The codes are (a) the Steane code, (b) the Bare [7,1,3] code [41], (c) the [9,1,3] Shor code, and (d) the [9,1,3] surface code. Different codes have different axes along which the post-error correction diamond norm error is more favorable than others.

It is striking that for this even distance error correcting code we reproduce the behavior for even distance repetition codes where we also found  $y' = 0$ . This confirms that the behavior seen there also arises for other more interesting quantum error correcting codes, although we still do not know if this is true for every even distance code in general.

As discussed by Rahn *et al.* [25], the effective channel can be used to study the performance of concatenated codes. The effective channel can be regarded as a map on noise processes, and iterating this map corresponds to concatenated error correction. As always, the noise map corresponds to a particular choice of decoder. In this case, the corrections are made at each level of the code using only the syndrome at that level. The full syndrome of the concatenated error correcting code is *not* used to determine the correction. Such decoders are known as *hard decoders* and it has been shown that soft decoders, that do consider the full syndrome, can lead to improved performance [45]. Moreover, simply iterating the mapping means that at each level of the code the symmetric decoder is used. It has also been shown that there are hard decoders that can improve on this [29]. (The improvements in [29] typically depend on knowledge of the noise model, whereas the concatenated symmetric or Z-only decoders can be specified independently of the precise noise model.) Nevertheless,

it is of interest to study the threshold of the concatenated symmetric decoder under unitary noise. For example, it was shown in [13] that after two levels of concatenation through a repetition code a unitary error of the form  $\exp(-i\theta Z)$  results in an effective noise process dominated by dephasing.

Looking at the Steane code again, the mapping in Eq. (67) can be concatenated to further reduce the noise in terms of the diamond distance for unitary noise as in Fig. 3(a). The level-1 pseudothreshold is

$$\theta_{\text{th1}} \approx 0.3276, \quad D_{\diamond, \text{th1}} \approx 0.3218. \quad (70)$$

For comparison, under incoherent dephasing noise alone, the level-1 pseudothreshold for concatenation of the Steane code is

$$p_{\text{th1}} \approx 0.0646, \quad D_{\diamond, \text{th1}} \approx 0.0646. \quad (71)$$

As can be seen from Fig. 3(b), this is also equal to the true threshold. This is because the logical noise model is exactly of the same form at each level of concatenation. By contrast, the true threshold for unitary Z rotation under concatenation is

$$\theta_{\text{th}} \approx 0.1918, \quad D_{\diamond, \text{th}} \approx 0.1906, \quad (72)$$

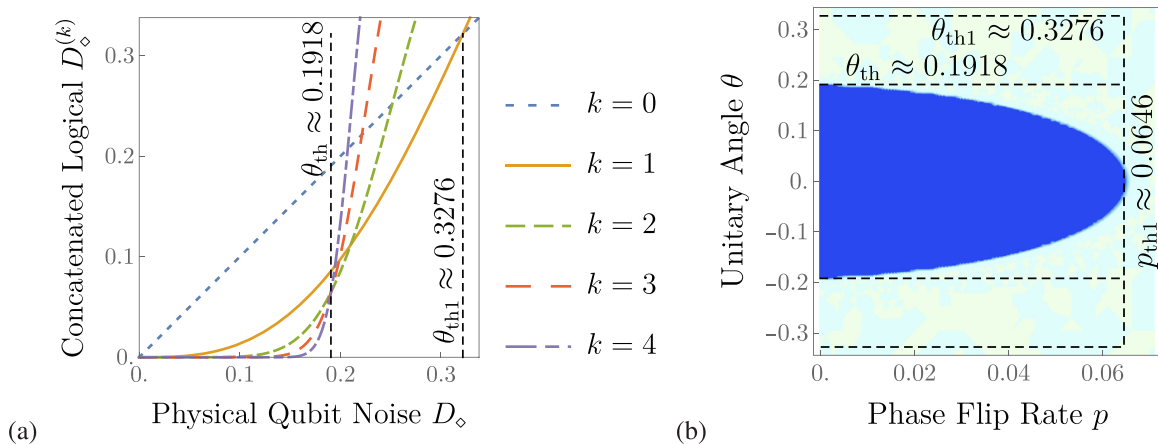


FIG. 3. (a) Plots of the diamond distance  $D_\diamond^{(k)}$  of the effective channel for the  $k$ th concatenation of the Steane code over the diamond distance  $D_\diamond$  of the original unitary  $X$  noise channel for  $k = 0, 1, 2, 3, 4, 5$ . The true threshold and level-1 pseudothresholds are indicated by dashed lines. (b) Thresholds for concatenation of the Steane code under unitary bit-flip noise as a basin of attraction in  $p$  and  $\theta$ . Thresholds and level-1 pseudothresholds for coherent and incoherent noise are labeled with dashed lines.

TABLE I. Concatenated code thresholds in terms of physical diamond norm error and average fidelity error for unitary  $Z$  noise and incoherent dephasing using minimum-weight  $Z$  decoders. In each case, the threshold diamond norm error exhibits two important effects. First, the level-1 pseudothreshold differs by a significant amount from the true threshold only for the case of unitary noise. Second, the threshold is substantially higher in the case of unitary errors compared to incoherent errors. Notice, however, that when the thresholds are expressed in terms of the average fidelity error, then the threshold for unitary error is less than or equal to the threshold for a dephasing error.

Code		Incoherent		Coherent			
$[n, k, d]$	Name	$D_\diamond(p_{\text{th}})$	$r(p_{\text{th}})$	$D_\diamond(\theta_{\text{th}})$	$r(\theta_{\text{th}})$	$D_\diamond(\theta_{\text{th}1})$	$r(\theta_{\text{th}1})$
[5,1,3]	Five-qubit	0.5000	0.3333	0.7071	0.3333	0.8284	0.4575
[7,1,3]	Steane	0.0646	0.0431	0.1906	0.0242	0.3218	0.0690
[7,1,3]	Bare	0.5000	0.3333	0.7077	0.3333	0.8761	0.5117
[9,1,3]	Shor	0.0499	0.0333	0.1177	0.0092	0.1477	0.0145
[9,1,3]	Surface	0.0753	0.0502	0.2053	0.0281	0.3085	0.0634
[16,1,4]	Surface	0.0395	0.0263	0.1613	0.0173	0.3180	0.0674

which is significantly lower than the level-1 pseudothreshold. We can look at the behavior over the full range of parameters to find the region for which the noise is below threshold and can be reduced to an arbitrarily small value by concatenating the code. This is done by considering the iterative map as a dynamical system and finding the basin of attraction, as indicated in Fig. 3(b).

We can repeat these calculations for a range of codes, as before. The results are summarized in Table I. It is worth noting that the true threshold for the two extremal noise models of pure unitary  $Z$ -rotation error or pure dephasing error exhibit a large gap in each of the cases we examine, and perhaps surprisingly the threshold for coherent noise is higher in every case.

## VII. CONCLUSION

We have studied the performance of quantum error correction against coherent noise, and coherent noise together with dephasing. We have proven that the logical effective channel corrects coherent errors much more effectively than incoherent errors, in the precise sense of Theorem 1. We have also shown that a natural family of decoders exhibits this effect for a wide range of codes both in the single-level coding regime, and for concatenated codes. The thresholds that we calculate suggest that quantum error correction performs better than expected when the errors are coherent, at least in the sense that this seems to lead to higher thresholds when these thresholds are expressed in terms of diamond norm errors.

There are many open questions suggested by this work, but the most interesting is to extend this analysis to the context of fault-tolerant quantum computation. It would be extremely interesting if the increased thresholds that we observed for ideal error correction continued to hold for fault-tolerant circuits. This would in turn strongly suggest that a sharper analysis of the threshold is needed to get accurate estimates of the threshold when coherent noise is taken into account. It would also be especially interesting to investigate what further

improvements could be obtained for these channels using optimal Pauli recovery channels instead of minimum distance decoding, or other structured near-optimal recoveries [46].

## ACKNOWLEDGMENTS

This work was supported by the Australian Research Council through the Centre of Excellence in Engineered Quantum Systems (Grants No. CE110001013 and No. CE170100009), and by the U. S. Army Research Office Grants No. W911NF-14-1-0098 and No. W911NF-14-1-0103.

## APPENDIX

In this Appendix we will collect some formulas from the stabilizer formalism that are standard (see for example [30]), but the discussion here will fix notation that will be required in the main text. We will write every Pauli matrix on  $n$  qubits in terms of two length  $n$  binary vectors  $\mathbf{b}_x$  and  $\mathbf{b}_z$ . Ones in this vector will correspond to a Pauli matrix acting on the corresponding qubit. The *support* of the vector  $\mathbf{b}_x$  will be the set of qubits for which the corresponding entries of  $\mathbf{b}_x$  are equal to one. The size of the support will be  $|\mathbf{b}_x|$ . We will also define the binary vector  $\mathbf{b}_x \cdot \mathbf{b}_z$  which is the bitwise product of the two vectors, and has ones on the intersection of the supports of  $\mathbf{b}_x$  and  $\mathbf{b}_z$ . These are the qubits on which the corresponding Pauli matrix acts with  $Y$ . Recalling that  $XZ = -iY$ , we have that a general Pauli matrix can be written

$$P_{\mathbf{b}} = i^{|\mathbf{b}_x \cdot \mathbf{b}_z|} X^{\otimes \mathbf{b}_x} Z^{\otimes \mathbf{b}_z} \quad (\text{A1})$$

and it is clear that the Pauli matrices are in one-to-one correspondence to the  $4^n$  binary vectors  $\mathbf{b}_x, \mathbf{b}_z$ . Note that the Pauli matrices  $P_{\mathbf{b}}$  so defined are the products of Paulis with a coefficient +1 and are therefore Hermitian. This Pauli has weight  $w = |\mathbf{b}_x| + |\mathbf{b}_z| - |\mathbf{b}_x \cdot \mathbf{b}_z|$ , there are  $m_x = |\mathbf{b}_x| - |\mathbf{b}_x \cdot \mathbf{b}_z|$   $X$  operators,  $m_z = |\mathbf{b}_z| - |\mathbf{b}_x \cdot \mathbf{b}_z|$   $Z$  operators, and  $m_y = |\mathbf{b}_x \cdot \mathbf{b}_z|$   $Y$  operators in this product, so that  $m_x + m_y + m_z = w$ .

We will imagine that  $\mathbf{b}_x$  and  $\mathbf{b}_z$  are row vectors, and so we can use the length  $2n$  row vector  $\mathbf{b} = (\mathbf{b}_x, \mathbf{b}_z)$  to denote a general Pauli matrix. We can define a  $2n \times 2n$  symplectic matrix

$$\Lambda = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}. \quad (\text{A2})$$

We can define a symplectic inner product as follows:  $\mathbf{b} \Lambda \mathbf{b}'^T = |\mathbf{b}_x \cdot \mathbf{b}'_z| - |\mathbf{b}_z \cdot \mathbf{b}'_x|$ . The two elements  $\mathbf{b}$  and  $\mathbf{b}'$  commute if and only if  $(-1)^{\mathbf{b} \Lambda \mathbf{b}'^T} = 1$ . While we have defined the symplectic inner product as a number, we will only ever use its value modulo 4. (Notice that there is a sign that is different compared to Nielsen and Chuang [30] and we do not evaluate inner products modulo 2, but rather use the conventional inner product.)

We can therefore compute products straightforwardly:

$$P_{\mathbf{b}} P_{\mathbf{b}'} = i^{|\mathbf{b}_x \cdot \mathbf{b}_z|} i^{|\mathbf{b}'_x \cdot \mathbf{b}'_z|} X^{\otimes \mathbf{b}_x} Z^{\otimes \mathbf{b}_z} X^{\otimes \mathbf{b}'_x} Z^{\otimes \mathbf{b}'_z} \quad (\text{A3})$$

$$= (-1)^{|\mathbf{b}'_x \cdot \mathbf{b}_z|} i^{|\mathbf{b}_x \cdot \mathbf{b}_z|} i^{|\mathbf{b}'_x \cdot \mathbf{b}'_z|} X^{\otimes \mathbf{b}_x} X^{\otimes \mathbf{b}'_x} Z^{\otimes \mathbf{b}_z} Z^{\otimes \mathbf{b}'_z} \quad (\text{A4})$$

$$= (-1)^{|\mathbf{b}'_x \cdot \mathbf{b}_z|} (-i)^{|\mathbf{b}'_x \cdot \mathbf{b}_z|} (-i)^{|\mathbf{b}_x \cdot \mathbf{b}'_z|} P_{\mathbf{b}+\mathbf{b}'}$$

$$= (-i)^{|\mathbf{b}_x \cdot \mathbf{b}'_z| - |\mathbf{b}'_x \cdot \mathbf{b}_z|} P_{\mathbf{b}+\mathbf{b}'} = (-i)^{\mathbf{b} \Lambda \mathbf{b}'^T} P_{\mathbf{b}+\mathbf{b}'}. \quad (\text{A5})$$

The second line makes use of the commutation formula we have just derived. In the final line we use  $(\mathbf{b}_x + \mathbf{b}'_x) \cdot (\mathbf{b}_z + \mathbf{b}'_z) = \mathbf{b}_x \cdot \mathbf{b}_z + \mathbf{b}'_x \cdot \mathbf{b}'_z + \mathbf{b}'_x \cdot \mathbf{b}_z + \mathbf{b}_x \cdot \mathbf{b}'_z$ . As claimed, the symplectic inner product only needs to be evaluated modulo 4. Since two Paulis commute if and only if the symplectic inner product is  $0 \pmod 2$ , when we multiply two commuting Paulis the phase factor is  $\pm 1$ , conversely when multiplying two anticommuting Paulis the phase factor is  $\pm i$ .

Consider now a stabilizer code. The stabilizer generators are signed products of Pauli matrices  $\pm P_{\mathbf{g}_i}$  where the binary vectors  $\mathbf{g}_i$  satisfy certain constraints, for example, that the stabilizer generators all commute. Since the generators are associated with  $n - k$  row vectors  $\mathbf{g}_i$  we can collect these vectors into a  $(n - k) \times 2n$  binary matrix  $G$  whose rows are the  $\mathbf{g}_i$ . For any given error  $E = P_{\mathbf{b}}$  the syndrome is a length  $n - k$  binary column vector  $\mathbf{s}$  given by

$$\mathbf{s}(\mathbf{b}) = G\Lambda\mathbf{b}^T \pmod 2. \quad (\text{A6})$$

A general member of the stabilizer group  $S_t$  is described by the length  $n - k$  binary vector  $\mathbf{t}$  such that

$$S_t = \pm \prod_{i=1}^{n-k} P_{\mathbf{g}_i}^{\mathbf{t}_i} = \pm P_{tG}. \quad (\text{A7})$$

The first factor of  $\pm 1$  is there to account for the signs of the stabilizer generators in product. In examples of interest to us, this will always be one. The phase factor in the final expression arises from iterating equation (A5) and is real because all the stabilizer generators commute. It has a definite

value that is suppressed in the interests of lightening the notation.

The code has  $k$  logical  $X$  operators  $\bar{X}_i = P_{x_i}$  and  $k$  logical  $Z$  operators  $\bar{Z}_i = P_{z_i}$ . Since the  $X$  operators are specified by  $k$  binary row vectors of length  $2n$ , we can specify the logical  $X$  operators by the  $k \times 2n$  binary matrix  $X$  whose rows are the  $x_i$ , and likewise a corresponding matrix  $Z$  for the logical  $Z$  matrices.

We can describe products of logical Pauli operators  $L_l$  by a length  $2k$  binary vector  $\mathbf{l} = (\mathbf{l}_x, \mathbf{l}_z)$ , in analogy to the Pauli matrices as follows:

$$L_l = (-i)^{|\mathbf{l}_x \cdot \mathbf{l}_z|} \prod \bar{X}_i^{\mathbf{l}_{x_i}} \prod \bar{Z}_j^{\mathbf{l}_{z_j}} = \pm (-i)^{|\mathbf{l}_x \cdot \mathbf{l}_z|} P_{l_x X} P_{l_z Z}. \quad (\text{A8})$$

The phase factor  $\pm 1$  arises because the product of logical  $X$ 's is a product of commuting operators and likewise the product of logical  $Z$ 's. Again, it has a definite value, given by iterating equation (A5), but suppressing this factor lightens the notation. Notice that the matrix product in the subscript results in a length  $2n$  binary row vector as expected.

There are  $2^{n-k}$  syndromes, and with each syndrome  $\mathbf{s}$  we will associate the lowest weight error  $E_{\tilde{\mathbf{b}}}$  that leads to that syndrome. Our decoder will reverse this error. Explicitly then  $E_{\tilde{\mathbf{b}}} = P_{\tilde{\mathbf{b}}}$  where  $\tilde{\mathbf{b}}$  is the lowest weight solution to  $\mathbf{s} = G\Lambda\tilde{\mathbf{b}}^T$ . Then, it can be shown that every Pauli matrix can be written in the form  $P_{\mathbf{b}} = \eta_{\mathbf{b}} E_{s(\mathbf{b})} L_{l(\mathbf{b})} S_{t(\mathbf{b})}$  where  $\eta_{\mathbf{b}}$  is a phase factor equal to  $\pm 1$  or  $\pm i$ .

- 
- [1] E Knill, R Laflamme, and W Zurek, Threshold accuracy for quantum computation, [arXiv:quant-ph/9610011](https://arxiv.org/abs/quant-ph/9610011).
- [2] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, in *29th ACM Symposium on Theory of Computing (STOC)* (ACM, New York, 1997), pp. 176–188.
- [3] A. Yu. Kitaev, Quantum error correction with imperfect gates, *Quantum Communication, Computing, and Measurement* (Springer, New York, 1997), pp. 181–188.
- [4] P. W. Shor, Fault-tolerant quantum computation, in *Proceedings of the 37th IEEE Symposium on Foundations in Computer Science (FOCS)* (IEEE, Piscataway, NJ, 1996), p. 56.
- [5] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Providence, RI, 2002).
- [6] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, *Quantum Inf. Comput.* **6**, 97 (2006).
- [7] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, Surface code quantum computing with error rates over 1%, *Phys. Rev. A* **83**, 020302(R) (2011).
- [8] P. Aliferis and A. W. Cross, Subsystem Fault Tolerance with the Bacon-Shor Code, *Phys. Rev. Lett.* **98**, 220502 (2007).
- [9] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, *Phys. Rev. A* **85**, 042311 (2012).
- [10] D. Puzzioli, C. Granade, H. Haas, B. Criger, E. Magesan, and D. G. Cory, Tractable simulation of error correction with honest approximations to realistic fault models, *Phys. Rev. A* **89**, 022306 (2014).
- [11] M. Gutiérrez, C. Smith, L. Lulushi, S. Janardan, and K. R. Brown, Errors and pseudo-thresholds for incoherent and coherent noise, *Phys. Rev. A* **94**, 042338 (2016).
- [12] A. S. Darmawan and D. Poulin, Tensor-Network Simulations of the Surface Code Under Realistic Noise, *Phys. Rev. Lett.* **119**, 040502 (2017).
- [13] D. Greenbaum and Z. Dutton, Modeling coherent errors in quantum error correction, *Quantum Sci. Technol.* **3**, 015007 (2018).
- [14] Y. Suzuki, K. Fujii, and M. Koashi, Efficient Simulation of Quantum Error Correction Under Coherent Error Based on Non-Unitary Free-Fermionic Formalism, *Phys. Rev. Lett.* **119**, 190503 (2017).
- [15] S. Bravyi, M. Englbrecht, R. Koenig, and N. Peard, Correcting coherent errors with surface codes, *npj Quantum Inf.* **4**, 55 (2018).
- [16] 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).
- [17] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding Quantum Gate Error Rate Based on Reported Gate Fidelity, *New J. Phys.* **18**, 012002 (2015).
- [18] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nat. Commun.* **8**, 14485 (2016).
- [19] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the Coherence of Noise, *New J. Phys.* **17**, 113020 (2015).



- [20] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, Comparing Experiments to the Fault-Tolerance Threshold, *Phys. Rev. Lett.* **117**, 170502 (2016).
- [21] J. J. Wallman, Bounding experimental quantum error rates relative to fault-tolerant thresholds, [arXiv:1511.00727](https://arxiv.org/abs/1511.00727).
- [22] J. J. Wallman and S. T. Flammia, Randomized benchmarking with confidence, *New J. Phys.* **16**, 103032 (2014).
- [23] J. Combes, C. Granade, C. Ferrie, and S. T. Flammia, Logical Randomized Benchmarking, [arXiv:1702.03688](https://arxiv.org/abs/1702.03688).
- [24] P. Iyer and D. Poulin, A small quantum computer is needed to optimize fault-tolerant protocols, *Quantum Sci. Technol.* **3**, 030504 (2018).
- [25] B. Rahn, A. C. Doherty, and H. Mabuchi, Exact Performance of Concatenated Quantum Codes, *Phys. Rev. A* **66**, 032304 (2002).
- [26] 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).
- [27] M. Gutiérrez, L. Svec, A. Vargo, and K. R. Brown, Approximation of realistic errors by Clifford channels and Pauli measurements, *Phys. Rev. A* **87**, 030302 (2013).
- [28] M. Gutiérrez and K. R. Brown, Comparison of a quantum error-correction threshold for exact and approximate errors, *Phys. Rev. A* **91**, 022335 (2015).
- [29] 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).
- [30] M. N. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
- [31] J. Fern, J. Kempe, S. N. Simic, and S. Sastry, Generalized performance of concatenated quantum codes — a dynamical systems approach, *IEEE Trans. Autom. Control* **51**, 448 (2006).
- [32] J. Watrous, Simpler semidefinite programs for completely bounded norms, *Chicago J. Comp. Sci.* **2013**, 8 (2013).
- [33] S. Beale, J. Wallman, M. Gutiérrez, K. R. Brown, and R. Laflamme, Quantum Error Correction Decohere Noise, *Phys. Rev. Lett.* **121**, 190501 (2018).
- [34] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. A. Sloane, Quantum error correction via codes over GF(4), *IEEE Trans. Inf. Theory* **44**, 1369 (1998).
- [35] E. M. Rains, Quantum codes of minimum distance two, *IEEE Trans. Inf. Theory* **45**, 266 (1999).
- [36] M. Grassl and M. Roetteler, Leveraging automorphisms of quantum codes for fault-tolerant quantum computation, in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium* (IEEE, Piscataway, NJ, 2013), pp. 534–538.
- [37] B. Zeng, A. Cross, and I. L. Chuang, Transversality versus universality for additive quantum codes, *IEEE Trans. Inf. Theory* **57**, 6272 (2011).
- [38] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Perfect Quantum Error Correcting Code, *Phys. Rev. Lett.* **77**, 198 (1996).
- [39] A. M. Steane, Error Correcting Codes in Quantum Theory, *Phys. Rev. Lett.* **77**, 793 (1996).
- [40] E. Magesan, D. Puzzioli, C. E. Granade, and D. G. Cory, Modeling quantum noise for efficient testing of fault-tolerant circuits, *Phys. Rev. A* **87**, 012324 (2013).
- [41] M. Li, M. Gutiérrez, S. E. David, A. Hernandez, and K. R. Brown, Fault tolerance with bare ancillary qubits for a  $[[7, 1, 3]]$  code, *Phys. Rev. A* **96**, 032341 (2017).
- [42] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [43] S. B. Bravyi and A. Yu. Kitaev, Quantum codes on a lattice with boundary, [arXiv:quant-ph/9811052](https://arxiv.org/abs/quant-ph/9811052).
- [44] N. H. Nickerson, Error correcting power of small topological codes, [arXiv:1609.01753](https://arxiv.org/abs/1609.01753).
- [45] D. Poulin, Optimal and efficient decoding of concatenated quantum block codes, *Phys. Rev. A* **74**, 052333 (2006).
- [46] A. S. Fletcher, P. W. Shor, and M. Z. Win, Structured near-optimal channel-adapted quantum error correction, *Phys. Rev. A* **77**, 012320 (2008).