Quantum error correction failure distributions: comparison of coherent and stochastic error models

Jeff P. Barnes,1,* Colin J. Trout,1,2 Dennis G. Lucarelli,1 and B.D. Clader1

1Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Rd., Laurel, MD, 20723, USA
2Schools of Chemistry and Biochemistry, Computational Science and Engineering, and Physics
Georgia Institute of Technology, Atlanta, GA 30332, USA

We compare failure distributions of quantum error correction circuits for stochastic errors and coherent errors. We utilize a fully coherent simulation of a fault tolerant quantum error correcting circuit for a $d = 3$ Steane and surface code. We find that the output distributions are markedly different for the two error models, showing that no simple mapping between the two error models exists. Coherent errors create very broad and heavy-tailed failure distributions. This suggests that they are susceptible to outlier events and that mean statistics, such as pseudo-threshold estimates, may not provide the key figure of merit. This provides further statistical insight into why coherent errors can be so harmful for quantum error correction. These output probability distributions may also provide a useful metric that can be utilized when optimizing quantum error correcting codes and decoding procedures for purely coherent errors.

I. INTRODUCTION

The theory of fault tolerance and the associated threshold theorem show that given an error rate on physical qubits below some threshold, one can perform a quantum computation with arbitrary accuracy with manageable overhead due to error correction (see e.g., [1–9]). It is well known that the Pauli matrices form a basis for any arbitrary qubit state. Therefore any arbitrary unitary evolution can be written as a linear combination of Pauli matrices. Thus, it is sufficient to correct only for these types of errors in quantum error correction [10].

This fact often leads to a logical leap that is not always justified. In particular, in numerical simulations of quantum error correction routines, it is standard practice to only consider stochastic Pauli errors. In some cases such a model is valid. For example, in models of open quantum systems [11–13], the system environment coupling causes the environment to constantly “measure” the qubits. This can cause qubit depolarization, which can be modeled as the mapping of the qubit density matrix

$$\rho \rightarrow (1-p_1-p_2-p_3)\rho + p_1 X \rho X + p_2 Y \rho Y + p_3 Z \rho Z,$$

where the $p_j$ terms are classical probabilities. Therefore a completely acceptable interpretation of Eq. (1) is one where with probability $(1-p_1-p_2-p_3)$ no error happens to the wavefunction, and with probability $p_j$ a Pauli $X$, $Y$, or $Z$ error occurs for $j = \{1, 2, 3\}$ respectively. This error model is referred to by a variety of different names in the literature. Here we refer to it as the “stochastic Pauli error model”. More generally, an error model can be generated with a larger set of gates than just Pauli gates, but where error gates are randomly inserted in a quantum circuit with classical probabilities. The key characteristic of all these stochastic error models is that the various error configurations all arise from classical probabilities, resulting in no interfering pathways between them.

Coherent errors, which can arise by over- or under-rotations of a quantum gate, were often called “inaccuracies” in the early literature on quantum error correction [14, 15]. They cannot map to any stochastic error model as they result in a fully coherent evolution with multiple interfering pathways. Proving the existence of threshold in the presence of coherent errors thus requires one to consider the fully coherent sum of interfering fault paths (see, e.g., Sec. 8 in Ref. [7]). The impact of coherent errors has received renewed attention due to recent results confirming that such errors can be a more severe hurdle for fault-tolerance than stochastic errors [16, 17].

A more thorough, physics-based justification for when a stochastic error model is an appropriate approximation requires one to consider the entire system-bath interaction. Only when the bath state, entangled with a given error, is orthogonal to all other bath states is the stochastic error model truly appropriate [5, 14]. Despite these fairly well-known results from the early literature on quantum error correction, the use of stochastic error models is still widely used in numerical simulations to calculate thresholds, including many results that have examined the accuracy of approximating various error channels by stochastic errors [18–23]. Of course, a significant reason for using this error model this is that these types of errors can be efficiently simulated classically via the Gottesman-Knill theorem [24, 25] if one restricts the set of gates to the Clifford group or a subset (often just Pauli operators are used).

Here, we report that a stochastic error model, has a drastically different logical failure distribution than a failure distribution resulting from coherent errors. To show this, we compare the output failure distributions of a quantum error correction (QEC) memory circuit correct-
ing stochastic Pauli errors to coherent errors. We utilize a numerical simulation of the entire encoded wavefunction, and we examine the failure distribution of the logical state for the Steane [7,1,3] code and a distance 3 surface code. We devise failure metrics that avoid the need to implement a final round of perfect error correction and decoding, commonly used when studying the failure characteristics of a logical channel [18, 22, 23, 26].

Our results show that the output failure distributions are markedly different for the two error models, with coherent errors lead to very broad failure distributions. This suggests that outlier events are much more likely to occur for coherent errors than for stochastic errors, which may provide insight into why they can be so harmful to QEC 

\[ \text{II. NUMERICAL SIMULATION APPROACH} \]

We numerically simulate two common QEC d = 3 codes in this paper: the Steane [7,1,3] code [2, 10], and the titled-17 surface code [27–29]. Insertions of errors codes in this paper: the Steane [7,1,3] code [2, 10], and the titled-17 surface code. We devise failure metrics that avoid the need to implement a final round of perfect error correction and decoding, commonly used when studying the failure characteristics of a logical channel [18, 22, 23, 26]. Our results show that the output failure distributions are markedly different for the two error models, with coherent errors lead to very broad failure distributions. This suggests that outlier events are much more likely to occur for coherent errors than for stochastic errors, which may provide insight into why they can be so harmful to QEC [16, 17].

For a single qubit, the most general evolution is given by an interaction Hamiltonian

\[ H = \hbar \omega (u_x X + u_y Y + u_z Z) \]

where \( X, Y \) and \( Z \) are Pauli operators [10], the rotation axis is \((u_x, u_y, u_z)\), and couplings with field amplitudes are absorbed into \( \omega \). Evolution is given by \( |\Psi\rangle \rightarrow U|\Psi\rangle \), with the propagator:

\[ U = \exp(-iHt/\hbar) = \cos(\omega t)I - i \sin(\omega t) (u_x X + u_y Y + u_z Z). \]

In these simulations, the shortest path in Hilbert space is used to evolve a gate. Thus, the Hadamard gate \( W = (X - Z)/\sqrt{2} \) is created with an axis \((1,0,-1)/\sqrt{2}\) and a duration \( t = \pi/(2\omega) \).

The CNOT\((n \rightarrow m)\) gate is implemented using the Hamiltonian:

\[ H = \hbar \omega (1 - |\psi_n\rangle\langle 1_n| + |1_n\rangle\langle 0_n| \otimes X_m). \]

The projectors ensure that no phase difference accumulates for \(|0_n\rangle\). This same structure extends to more than one controlling qubit, such as the Toffoli gate. For the controlled-Z gate, replace \( X \) with \( Z \).

A measurement of qubit \( n \) as implemented as

\[ |\Psi\rangle \rightarrow \left\{ \begin{array}{ll} |0_n\rangle\langle 0_n|\Psi\rangle/||0_n|\Psi\rangle|| & \text{rand()} \leq ||0_n|\Psi\rangle|^2 \\ |1_n\rangle\langle 1_n|\Psi\rangle/||1_n|\Psi\rangle|| & \text{otherwise.} \end{array} \]  \hspace{1cm} (6) \]

Measurements are instantaneous, and their records are perfect. These are standard assumptions in QEC.

\[ \text{B. Error Models} \]

1. Pauli Error Model

A stochastic Pauli error model is implemented by selecting a set of fault locations and an error rate, \( p \). A detailed discussion can be found in [30]. The choice here is conservative: errors are restricted to being only on the codeword qubits, and only before each syndrome measurement. While conservative, this is not unrealistic for this error model as the Pauli errors can be efficiently commuted through a QEC circuit. When the code encounters a fault location, it randomly applies an \( X \) and \( Z \) error each with probability \( p \). Since \( XZ = -iY \), all Pauli errors can occur.

2. Pulse-Area Error Model

For coherent errors, we utilize an error model that we call the pulse-area error model. Coherent manipulations of quantum systems almost universally involve the application of an electromagnetic field, and these are imprecise. Static fields are not independent degrees of freedom, and their fluctuations are tied back to their sources [31, 32]. Free fields, which include RF pulses in magnetic resonance [33] and laser pulses in time domain optical spectroscopy [34], are usually well represented by coherent states [35], and their quantum back-action is quantifiable [36]. Commonly, though, these terms can be neglected, and the field amplitudes are c-functions in the Hamiltonian that describes a qubit gate. A perfect gate requires the integral of the field amplitude over time (in a rotating frame) to be a fixed angle. The pulse-area error model puts a distribution on this angle. The mean is zero, since systematic errors can be removed with calibration [37].

The pulse-area error model consists of replacing \( \omega \) in Eq. (4) with \( \omega(1 + \sigma r) \), with \(-1 < r < 1\) uniformly and independently per gate.

\[ \exp \left( -\frac{i}{\hbar} \omega G \frac{\pi}{2\omega} \right) \rightarrow \exp \left( -\frac{i}{\hbar} (1 + \sigma r) G \frac{\pi}{2} \right) \]

All gates in the QEC circuit have \( \omega = 1.0 \) and duration \( \pi/2 \). Thus, \( \sigma \) is related to the fractional jitter in the field
amplitudes. For example, for a commercially available pulsed laser sources, \( \sigma \approx 0.005 \) [38].

### III. METRIC FOR QEC FAILURE

This section derives a metric to quantify QEC circuit failure, but adapted to a wavefunction simulation. The prescription here is for the Steane \([7,1,3]\) code with a few special considerations required for the surface code noted in Sec. III B. The focus here is on developing a metric for failure that does not require any perfect measurements, QEC rounds, or decoding, but rather a metric that can be applied to the wavefunction itself.

#### A. Steane Failure Metric

The following definition, from [6], is considered a standard: a QEC circuit that takes any input state with

\[
\text{repair} \quad \text{error} \quad \text{encode} \quad \text{ancilla}
\]

\[
\begin{align*}
T(2,3,1)C(1,2)C(1,3)X(1)C(1,3)C(1,2)P(2,3) &= X(2)X(3)P(2,3) \\
T(2,3,1)C(1,2)C(1,3)X(2)C(1,3)C(1,2)P(2,3) &= X(2)P(2,3) \\
T(2,3,1)C(1,2)C(1,3)X(3)C(1,3)C(1,2)P(2,3) &= X(3)P(2,3)
\end{align*}
\]

(8)

\[
S_L = \{ |0_L\rangle, |1_L\rangle \}
\]

\[
S_{L+1} = S_L \cup \{ X_q |0_L\rangle, X_q |1_L\rangle, Y_q |0_L\rangle, Y_q |1_L\rangle, Z_q |0_L\rangle, Z_q |1_L\rangle \}
\]

\[
S_{L+2} = S_L \cup \{ X_q |0_L\rangle, X_q |1_L\rangle, Z_q |0_L\rangle, Z_q |1_L\rangle, X_q Z_q' |0_L\rangle, X_q Z_q' |1_L\rangle \} \quad \forall q, q'
\]

(9)

These kets form orthonormal states, with \(|S_L| = 2, |S_{L+1}| = 44, \) and \(|S_{L+2}| = 128\) for the particular case of Steane's code. Further, since \( X_q Z_q = -i Y_q \), one can show \( S_L \subseteq S_{L+1} \subseteq S_{L+2} \). But note that the total Hilbert space of Steane's code has \( 2^7 = 128 \) dimensions. Thus, \( S_{L+2} \) completely spans it. Since no state could ever leave \( S_{L+2} \), it cannot represent a QEC failure criteria, but \( S_{L+1} \) can.

From the vantage point of Monte Carlo wavefunction simulations, it seems more natural to check if two or more Pauli operators have acted on the starting state \( |\Psi_0\rangle \) for a \( d = 3 \) code. These spaces are built up as

\[
S_\psi = \{ |\Psi_0\rangle \}
\]

\[
S_{\psi +1} = S_\psi \cup \{ X_q |\Psi_0\rangle, Y_q |\Psi_0\rangle, Z_q |\Psi_0\rangle \}
\]

\[
S_{\psi +2} = S_\psi \cup \{ X_q |\Psi_0\rangle, Z_q |\Psi_0\rangle, X_q Z_q' |\Psi_0\rangle \}.
\]

(10)

They are orthonormal sets, with \(|S_{\psi +1}| = 22\) and \(|S_{\psi +2}| = 64\) for the particular case of Steane's code. Since Steane's code and the surface code are both CSS codes, they correct errors of the form \( X_q Z_{q'} \) for all \( q \) and \( q' \). Checking if \( |\Psi(t)\rangle \) has ventured beyond either \( S_{\psi +1} \) or \( S_{\psi +2} \) is reasonable.

To summarize, we have three metrics to detect a QEC circuit failure, in Eq. (11). For each simulation, when they are 1, the circuit has failed; otherwise, we expect them to be 0.

\[
P_{\text{fail}}^{(\psi+1)} = 1 - \sum_{s \in S_{\psi+1}} |\langle s |\Psi(t)\rangle|^2
\]

\[
P_{\text{fail}}^{(\psi+2)} = 1 - \sum_{s \in S_{\psi+2}} |\langle s |\Psi(t)\rangle|^2
\]

\[
P_{\text{fail}}^{(\psi+1)} = 1 - \sum_{s \in S_{\psi+1}} |\langle s |\Psi(t)\rangle|^2.
\]

(11)

Technically the sums in Eq. (11) should also extend over the ancilla qubits. However, we assume that measurements are perfect, therefore when we analyze the failure metrics, we know that the ancilla will be in a product state with the data qubits, and thus the sum over them is independent and can be excluded.

Two additional metrics are useful: the projection into
\[ \mathcal{S}_L, \text{ given in Eq. (12), and the fidelity } F, \text{ in Eq. (13).} \]
The result \( F = 1.0 \) implies a perfect repair.
\[
P_{\text{code}} = \sum_{s \in \mathcal{S}_L} |\langle s | \Psi(t) \rangle|^2 \tag{12}\]
\[
F^2 = |\langle \Psi(t) | (\alpha |0_L \rangle + \beta |1_L \rangle) |^2. \tag{13}\]

Several useful bounds are shown in Eq. (14). Physically, a low \( P_{\text{code}} \) implies a low \( F \), since \( |\Psi(t)\rangle \) is outside of \( \mathcal{S}_L \). The inverse does not hold, since a logical error such as \( |1_L\rangle = \mathbf{X}_L |0_L\rangle \) on an encoded \( |0_L\rangle \) leaves \( P_{\text{code}} = 1 \) but \( F = 0 \). It is also seen that \( P_{\text{fail}}^{(\psi+1)} \) is the most stringent criteria for circuit failure.
\[
P_{\text{fail}}^{(L+1)} \leq P_{\text{fail}}^{(\psi+1)}
\]
\[
P_{\text{fail}}^{(\psi+2)} \leq P_{\text{fail}}^{(\psi+1)}
\]
\[
F^2 \leq P_{\text{code}} \leq 1 - P_{\text{fail}}^{(L+1)}. \tag{14}\]

Several computational tools to Monte Carlo estimate a pseudo–threshold apparently restrict \( |\Psi_0\rangle \) to be one of the six stabilizer states
\[
\{ |0_L\rangle, |1_L\rangle, (|0_L\rangle \pm |1_L\rangle)/\sqrt{2}, (|0_L\rangle \pm i|1_L\rangle)/\sqrt{2} \}. \tag{15}\]
This makes drawing distinctions between the metrics in Eq. (11) problematic, as discussed later in this paper.

### B. Surface Code Failure Metric

The failure metrics for the Steane code was relatively straightforward since all errors map to a unique syndrome. There is a degeneracy in the surface code that complicates this somewhat. We show here how to define a failure metric analogous to that for the Steane code that takes this into account.

The logical state is generated utilizing the parity-check matrix constructed from the \( \mathbf{X} \) stabilizers for the surface code shown later in equation 22. It results in the basis states:
\[
|0\rangle_L = \frac{1}{4} \left( |000000000\rangle + |110110000\rangle + |011000000\rangle + |101110000\rangle \right)
\]
\[
+ |101110000\rangle + |000001110\rangle + |110110110\rangle + |011010111\rangle + |101101011\rangle + |000111101\rangle + |110101101\rangle + |011011101\rangle
\]
\[
+ |101101101\rangle \right) \tag{16}\]

\[
|1\rangle_L = \hat{X}^{\otimes 9} |0\rangle_L
\]

that we can use to construct our logical code space as in equation 9. Note that state \( |1\rangle_L \) is the \( |0\rangle_L \) state under the action of the bitwise-NOT on every binary string within it’s sum. Due to this relation, there is an equivalence in the action of single-qubit errors to both \( |0\rangle_L \) and \( |1\rangle_L \) with respect to the degeneracy of the single-error spaces. For single-qubit errors, there is equivalence of the action of \( \hat{X} \) errors, \( \hat{X}_1 |0\rangle_L = \hat{X}_2 |0\rangle_L \) and \( \hat{X}_6 |0\rangle_L = \hat{X}_7 |0\rangle_L \), and also degeneracies for \( \hat{Z} \) errors, \( \hat{Z}_0 |0\rangle_L = \hat{Z}_1 |0\rangle_L \) and \( \hat{Z}_5 |0\rangle_L = \hat{Z}_8 |0\rangle_L \). \( \hat{Y} \) errors are completely non-degenerate. We can therefore construct the failure criteria pertaining to the logical codeword space in the following manner:

\[
\mathcal{S}_L = \{ |0_L\rangle, |1_L\rangle \}
\]
\[
\mathcal{S}_{L+1} = \mathcal{S}_L \cup \{ \hat{X}_i |0_L\rangle, \hat{Y}_j |1_L\rangle, \hat{Y}_j |0_L\rangle, \hat{Z}_k |0_L\rangle, \hat{Z}_k |1_L\rangle \} \text{ s.t. } i \neq 2, 7; \ k \neq 3, 8 \tag{17}\]

where the indices \( i, j, k \) run over all data qubit indices unless otherwise specified. Similarly, for the criteria more natural for the purpose of wavefunction simulations the spaces can be constructed as:
\[
\mathcal{S}_\psi = \{ |\Psi_0\rangle \}
\]
\[
\mathcal{S}_{\psi+1} = \mathcal{S}_\psi \cup \{ \hat{X}_i |\Psi_0\rangle, \hat{Y}_j |\Psi_0\rangle, \hat{Z}_k |\Psi_0\rangle \} \tag{18}\]
\[
\mathcal{S}_{\psi+2} = \mathcal{S}_\psi \cup \{ \hat{X}_i |\Psi_0\rangle, \hat{Z}_k |\Psi_0\rangle, \hat{X}_i \hat{Z}_k |\Psi_0\rangle \} \text{ such that } i \neq 2, 7; \ k \neq 3, 8 \]

where, again, the indices run over all qubits except for special cases. While it appears that some single-qubit \( \hat{Y} \) errors have been omitted in the space \( \mathcal{S}_{\psi+2} \), the symmetry of the error space ensures that these states have been taken into account. For instance, the state \( \hat{Y}_2 |\Psi_0\rangle = i \hat{X}_2 \hat{Z}_2 |\Psi_0\rangle \) which, through the degeneracy of the \( \hat{X} \) errors, can be represented by the state \( i \hat{X}_1 \hat{Z}_2 |\Psi_0\rangle \). The assessment of the surface code will incorporate computing the overlap of output wavefunctions with the aforementioned correctable error spaces. Because the surface
code is an error correcting code, the success criteria $P_{\text{code}}$ and fidelity ($F^2$) given in Eqs. (12) and (13) will be of interest as well.

IV. STEANE SIMULATION RESULTS

With failure criteria now defined, we proceed to simulate the error models discussed in Sec. II B for Steane's [7,1,3] code. The baseline QEC trial begins by encoding a random qubit. The syndromes, and the errors they detect, are given in Table I. To ensure fault-tolerance, the three syndromes that check for $Z$ errors are measured, and then measured again. A loop back occurs if the three sets of syndrome bits do not match. Otherwise, any $Z$ errors are repaired, and the circuit repeats this for the $X$ error detecting syndromes.

The syndrome measurement procedure uses Shor style ancilla It begins by creating a cat state in the 4 ancilla qubits, which is always error free. The test and rejection steps seen in Fig. 6 of [6] are thus not required. After entanglement, the cat state is rotated and read out in the style discussed in Ref. [39]. Each ancilla is then perfectly re-initialized allowing them to be re-used for the next syndrome.

![FIG. 1. Histograms of the Steane QEC failure metrics, from $3 \times 10^6$ trials of the Steane QEC circuit, using the Pauli error model with $p = 0.00005, 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02$. $P_{\text{fail}}(L+1)$ (top) is always binomially distributed, but $P_{\text{fail}}(\psi+1)$ (middle) and $P_{\text{fail}}(\psi+2)$ (bottom) never are.

A. Pauli Error Model

The Steane QEC circuit was run $3 \times 10^6$ times at nine different $p$ values. Histograms of each failure criteria are shown in Fig. 1. The metric $P_{\text{fail}}(L+1)$ is binomially distributed (within numerical accuracy), and has the obvious interpretation. The fraction of trials with $P_{\text{fail}}(L+1)(p) = 1$ for a given $p$ is shown by the red curve in Fig. 2. A pseudo-threshold of $\approx 0.005$ can be observed. The curve appears similar to examples provided in [40], and gives confidence that the wavefunction simulations can reproduce the standard pseudo-threshold.

The difficulty with $P_{\text{fail}}(\psi+1)$ and $P_{\text{fail}}(\psi+2)$ is that the spaces of Eq. (10) are not aligned with the 6 stabilizer states of Eq. (15). When the code is run with the $|\Psi_0\rangle$ restricted to be from the set of Eq. (15), then all three metrics in Eq. (11) are binomially distributed. We note here that all Gottesman-Knill and error propagation simulations either require or implicitly assume that input states are Stabilizer states. Even for Pauli errors, this restriction already changes the output failure distribution away from what will actually happen when arbitrarily encoded states must be protected. Since, arbitrary encoded states cannot be abandoned, so how can this result be understood?

First ask: does every non-zero value of $P_{\text{fail}}(\psi+1)$ or

| Operator | $Z_1$ | $Z_2$ | $Z_3$ | $Z_4$ | $Z_5$ | $Z_6$ | $Z_7$
|----------|------|------|------|------|------|------|------|
| $X_2X_3X_5X_7$ | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| $X_2X_3X_6X_7$ | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| $X_3X_4X_5X_7$ | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| $Z_2Z_3Z_5Z_7$ | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| $Z_2Z_3Z_6Z_7$ | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| $Z_3Z_4Z_5Z_7$ | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

FIG. 2. The Monte–Carlo simulation of the failure rate for the Steane QEC circuit, using the Pauli error model with $p = 0.00005, 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02$. $P_{\text{fail}}(L+1)$ (top) is always binomially distributed, but $P_{\text{fail}}(\psi+1)$ (middle) and $P_{\text{fail}}(\psi+2)$ (bottom) never are.
$P_{\text{fail}}^{(\psi+2)}$ indicate a circuit failure? One way to examine this is to run chains of QEC circuits, where the input of next QEC cycle is the output of the previous one. In such simulations, $P_{\text{code}}$ and $F$ transiently drop from 1.0, but then they recover, as seen in Fig. 3 (a). This is a case where weight-1 errors escape a QEC cycle, only to be repaired in the next cycle. By chance, this can occur sequentially, but the $P_{\text{fail}}$ metrics remain 0.0, indicating the errors were not fatal ones.

In contrast, if $F^2 < 1$ and $P_{\text{code}} = 1$, then the encoded qubit was corrupted, and clearly the QEC circuit failed. These situations are seen in Fig. 3 (b) to (e). Scatterplots of large samples with $p = 0.02$ did not find any other distinct behaviors, so we consider how to explain each case.

1. Case b

Within a single cycle, $P_{\text{code}} = 1$ and $F < 1$, indicating a corrupted but encoded qubit. There are many insertions of 3 Pauli errors that caused this. An example is $X_4$, $X_5$, and then $X_6$, all during the measurement of the $Z$ error detecting syndromes. Within Steane’s code, $X_4X_5 \equiv X_5X_6$, so the $X$ repair returns $X_L|\Psi_0\rangle$, a properly encoded but corrupted state.

In the failed QEC cycle, it is found that $P_{\text{fail}}^{(L+1)} = 0$, yet $P_{\text{fail}}^{(\psi+2)} = P_{\text{fail}}^{(\psi+1)} = 0.9597$ and $F = 0.2008$. The explanation comes by considering the starting state of Eq. (2). For a $X_L$ error, $P_{\text{fail}}^{(\psi+1)} = 1 - \sin^4 \theta \cos^2 \phi$. If the success criteria is set by demanding a threshold, say $P_{\text{fail}}^{(\psi+1)} > 10^{-6}$, then irreparable errors are allowed when encoded states are “lucky” enough to be nearly an eigenstate of that error. For a threshold of $10^{-6}$, errors are allowed if they shift $|\Psi_0\rangle$ by $< 0.05^\circ$ on the Bloch sphere. Whether this suffices will depend on the needs of the quantum algorithm.

2. Case c

In this trial, two fault locations had errors. Referring to Table I, the $Z$ error detection syndromes finished without errors, but a $Z_q$ error occurs after the first two $X$ syndrome measurements. Since these are checking for $X$ errors, the syndrome is $(0,0,0)$. In the second round, a $X_7$ error occurs before the last syndrome. This error should return $(1,1,0)$, but we are on the last syndrome, so $(0,0,0)$ results. As the consistency check is passed, no errors are detected. For this $p^2$ process, an error of $X_7Z_6$ escapes. These combination of $X_qZ_{q'}$ for $q \neq q'$ are the only errors that distinguish $P_{\text{fail}}^{(\psi+1)}$ from $P_{\text{fail}}^{(\psi+2)}$.

In the spirit of Ref. [6], the more stringent weight-1 error criteria of $P_{\text{fail}}^{(\psi+1)} > 10^{-6}$ is favored.

3. Cases d and e

This appears to show cases where $P_{\text{fail}}^{(\psi+1)} > 10^{-6}$ cannot distinguish all errors. Both (d) and (e) would be detected as errors, yet the permanent loss of $F$ in (d) is not observed in (e). An examination of several QEC chain trials suggests the answer.

In one case, during the second measurement of the syndromes to detect $X$ errors, but just before the consistency check, the following Pauli operators were inserted: $S, X_3, S, X_6, S$, where $S$ implies a syndrome measurement. The first syndrome returns 0. An $X_2$ error should give $(1,0,0)$, so the second syndrome returns 0. At the next error, $X_6X_5 \equiv X_6X_6$, and the syndrome for an $X_4$ error is $(0,1,0)$. Thus, the last measurement yields $(0,0,0)$, consistent with the first syndrome measurements. This $p^2$ process allowed a corrupted qubit with a weight-1 error to pass.

In another case, in the second measurement of the $X$ error detecting syndromes, this sequence was found: $S, X_3, S, X_2, S$. As before, this dances around the detection table, and $X_6X_6$ escapes. In the next QEC cycle, however, a single $X_2$ error occurs within the $Z$ detecting syndromes. Only a $X_2$ error is left for the QEC circuit to repair. By random chance, a $p$ process has fixed an unrepairable $p^2$ error.

Three cases involving $p^2$ processes with outcomes like in (e) were examined. They all had the same result: by random chance, the next QEC cycle had a $p$ process that converted the weight-2 error into a weight-1 error. As this cannot be assigned to the role of the QEC circuit, we favor the criteria $P_{\text{fail}}^{(\psi+1)} > 10^{-6}$ to be a good detector of QEC circuit failure.
B. Pulse-Area Error Model

We now use the pulse area error model outlined in Eq. (7). The reader should keep in mind the following argument in support of a stochastic error model. Consider the QEC circuit as an operator expansion:

$$\cdots \Omega_3 \left( \cos \frac{\pi \sigma r_2}{2} U_2 - i \sin \frac{\pi \sigma r_2}{2} 1 \right) \left( \cos \frac{\pi \sigma r_1}{2} U_1 - i \sin \frac{\pi \sigma r_1}{2} 1 \right) \left| \Psi_0 \right\rangle$$

(19)

There are measurements $\Omega$ and unitary gates $U$. Following Eq. (7), and using $U_n^2 = 1$, this becomes

$$\cdots \Omega_3 \left( \cos \frac{\pi \sigma r_2}{2} U_2 - i \sin \frac{\pi \sigma r_2}{2} 1 \right) \left| \Psi_0 \right\rangle$$

(20)

Now suppose $\sigma \ll 1$. In that case, the linear terms in $\sigma$ simply have each gate $U_n \rightarrow 1$. This superficially resembles the Pauli error model: insert a Pauli operator, which cancels the gate, with amplitude $\sigma r_n$ at gate $n$. The sum of these terms in Eq. (20) creates $|\Psi(t)\rangle$. Using the argument in support of the stochastic error model, each syndrome measurement $\Omega$ selects one of these terms at a time, so the incoherent sum only needs to be considered. That is very useful, because even if $\sigma^2 \ll \sigma$, with the 144 gates of the Steane syndrome extraction circuitry, there are 10,296 dual gate $O(\sigma^2)$ terms to add up, which would be difficult. Yet, in order to fully select one error, 6 syndromes must be measured, even while errors continue to accumulate. The numerical simulation can be used to check this approximation.

To see how the Steane code handles these pulse area errors, $10^6$ trials with 0.001 $\leq \sigma \leq 0.100$ are shown in Fig. 4. All three failure metrics are shown for completeness. First, note that the distributions are very broad. A curve fit to the distribution of $P_{\text{fail}}^{(\psi+1)}(\sigma = 0.01)$ showed that roughly 75% of the trials could be described by the log-normal distribution of the form $\exp(-a \log^2(bP_{\text{fail}}))$.

The maximum likelihood scales as $\text{ML}(P_{\text{fail}}^{(\psi+1)}) \approx 5\sigma^4$. Employing the $P_{\text{fail}}^{(\psi+1)} > 10^{-6}$ criteria for failure results in a failure-versus-$\sigma$ curve in Fig. 5. For small $\sigma$, it varies as $\approx 200\sigma^{2.5}$, but comparison with Fig. 2 reveals a much more sigmoidal shape.

In the Pauli error model, a successful QEC cycle completely restores the encoded qubit. What about the pulse area error model? Fig. 6 shows log-log plots for the distributions of $F$ and $P_{\text{code}}$. Perfect outcomes ($F = P_{\text{code}} = 1$) are not apparent, but it may be that the part of $|\Psi(t)\rangle$ outside of $S_L$ is still repairable. What about the part inside? Consider the metric in Eq. (21).

$$F^2/P_{\text{code}} = \frac{|\langle \Psi_0 |\Psi(t)\rangle|^2}{|\langle 0|\Psi(t)\rangle|^2 + |\langle 1|\Psi(t)\rangle|^2}$$

(21)

It normalizes the part of $|\Psi(t)\rangle$ still in $S_L$. If it is $< 1.0$, then $|\Psi(t)\rangle$ has rotated from $|\Psi_0\rangle$ within the codeword space. This is an unrepairable error. Histograms of this metric are shown at the bottom of Fig. 6. For $\sigma = 0.005$, the most likely value of $F^2/P_{\text{code}} \approx 0.0001$. Thus, a quantum calculation that encodes digital numbers into amplitudes will only be accurate to approximately 4 digits. This distribution scales like $4.0\sigma^4$, so to achieve float64 precision (16 decimal digits) requires $\sigma \approx 0.0007$. This is analogous to round-off errors in finite precision classical computers. We note that this error is likely to be reduced as code distance is increased, but lacking a distance 5 simulation we cannot show this.

We finally note that, like the Pauli error model, we have simulated strings of QEC circuits to examine the long term behavior (not shown). This small errors noted here did not grow over multiple cycles, but remained bounded. The QEC circuit is working, but the behavior is much more complicated than the Pauli error model suggests.

V. SURFACE CODE SIMULATION RESULTS

We have demonstrated that the failure distributions for the pulse area error model are qualitatively different than those generated by purely Pauli errors. For completeness we run a similar analysis on a surface code simulation. The particular variant we choose is the tilted-17 surface code, which is a variation on the distance 3 surface code that requires less data and ancilla qubits compared to
FIG. 5. Using the criteria $P_{\text{fail}}^{(\psi+1)} > 10^{-6}$ on the $10^6$ trials of the Steane QEC circuit, under the pulse-area error model, gives rise to this failure rate curve. The failure rate does not follow the expected linear slope. The black line is the unit line.

conventional distance 3 surface code [27–29]. A diagram of the qubit and stabilizer layout of the code is shown in Fig. 7. A list of the eight stabilizers are shown below:

$$
\begin{align*}
\hat{X}_0\hat{X}_1\hat{X}_3\hat{X}_4 & \quad \hat{Z}_0\hat{Z}_3 \\
\hat{X}_1\hat{X}_2 & \quad \hat{Z}_1\hat{Z}_2\hat{Z}_4\hat{Z}_5 \\
\hat{X}_6\hat{X}_7 & \quad \hat{Z}_3\hat{Z}_4\hat{Z}_6\hat{Z}_7 \\
\hat{X}_4\hat{X}_5\hat{X}_7\hat{X}_8 & \quad \hat{Z}_5\hat{Z}_8
\end{align*}
$$

The measurement of the weight-4 stabilizers are scheduled in such a way that single qubit errors on the ancilla qubits propagate to, at most, two-qubit errors on the data in a manner such that the propagation direction is perpendicular to the direction of the logical operator of the surface code resulting in fault-tolerance [41]. An additional requirement for fault-tolerance is the repetitive measurement (typically for $d$ measurement rounds) of the stabilizers before applying a correction operation.

For this study, a lookup table decoder was implemented for error correction [41]. This method utilizes a small set of syndrome processing rules that is equivalent to applying a minimum weight perfect matching algorithm [42–44] to only nearest neighbor syndrome pairs [41] which, for a distance 3 code, is sufficient for error correction. There is some freedom to how one can schedule the syndrome extraction routine. For this manuscript we use a “single-shot” detection and correction cycle where we perform $X$ followed by $Z$ stabilizer measurements, repeat them three times to ensure fault tolerance, then decoding/correction is performed once.

A. Pauli Error Model

The “single-shot” error correction routines were each run $1 \times 10^6$ times at each error rate $p = 0.000001$, 0.0000025, 0.000005, 0.0000075, 0.00001, 0.000025, 0.00005, 0.000075, 0.0001, 0.0002, 0.0004, 0.0006, 0.0008. We show the failure histograms generated from the surface code simulations, in Fig. 8. Besides slightly shifted values for the failure histograms there is no qualitative difference between the surface code results compared to the Steane code shown in Fig. 1. This is not a surprising or unexpected result.

We observe a pseudo-threshold of $\sim 3 \times 10^{-5}$ from these surface code simulations. Note that our reported pseudo-thresholds appear to be below the value reported in [41] of $3 \times 10^{-4}$. The simulations in [41] implement $|0 \rangle_L$ as in input wavefunction while our simulations randomly sample a vector in the logical code word space $1/\sqrt{2} (|0 \rangle_L + \beta |1 \rangle_L)$. Also, there is a discrepancy between the labeling of the error rates between this study and the study in [41] where our recorded values of $p$ is equivalent to $p/3$ in [41]. By fixing our initial state to just $|0 \rangle_L$ and using the $P_{\text{fail}}^{(L+1)}$ criteria, we obtain a pseudo-
threshold of $\approx 4 \times 10^{-5}$ which, in the language of [41], is reported as $p_{th} \approx 1.2 \times 10^{-4}$; a comparable value.

![Histograms](image1)

**FIG. 8.** Histograms of the failure criteria for the tilted surface code with the Pauli error model. $1 \times 10^6$ samples were accumulated per error rate. Pauli error rates: $p = 0.000001, 0.0000025, 0.000005, 0.0000075, 0.00001, 0.000025, 0.00005, 0.000075, 0.0001, 0.0002, 0.0004, 0.0006, 0.0008$ generate the curves from left to right respectively.

![Histograms](image2)

**FIG. 9.** Histograms of the failure criteria for the tilted surface code with the pulse area error model. $1 \times 10^6$ samples were accumulated per error rate. Pulse area error strengths: $\sigma = 0.0025, 0.005, 0.0075, 0.01, 0.02, 0.03, 0.04, 0.05$ generate the curves from left to right.

### B. Pulse-Area Error Model

To compare the failure criteria of the surface code between the Pauli error model and pulse-area error model, we perform surface code simulations using the exact same circuit as for the Pauli error model just shown. We use the following error strengths, $\sigma = 0.0025, 0.005, 0.0075, 0.01, 0.02, 0.03, 0.04, 0.05$ ($1 \times 10^6$ samples per error rate). Broad, heavy-tailed distributions for the failure criteria are again observed as shown in Fig. 9 and appear to follow a log-normal distribution in a majority of the cases. Once again, just as the failure metrics for the Pauli error model and pulse area error model varied drastically with the Steane code, the surface code displays similar behavior. There is no simple map between the Pauli error model and the pulse area error model, with the pulse-area error model leading to quite broad and heavy-tailed distributions.

### VI. CONCLUSIONS

We have presented a wavefunction simulation of Steane’s [[7,1,3]] quantum error correcting code and the tilted-17 surface code for both the stochastic Pauli error model and the coherent pulse-area error model for randomly encoded states. The usual failure criteria had to be modified in order to account for the different frame that non-stabilizer states occupy. A comparison of Fig. 1 and Fig. 4 shows that the two error models result in markedly different failure distributions. Pauli errors tend to draw $P_{\text{fail}}(\psi_{+1})$ down from 1.0 with increasing $p$. These tails empirically vary as $\sqrt{P_{\text{fail}}(\psi_{+1})}$. In contrast, for increasing $\sigma$ in the pulse-area error model, the distribution rises from small $P_{\text{fail}}(\psi_{+1})$, with the most likely values scaling as $\sigma^4$. We conjecture that the heavy tails in the pulse-area error model distribution are an indicator of the strongly negative impact that purely coherent errors can have on QEC [16, 17].

These results also demonstrate that any result attempting to approximate arbitrary physical error channels with stochastic Pauli error channels must do so only when modeling certain system-bath type models. We have shown that one cannot rely on the assumption that syndrome measurements cut off the interfering pathways. A stochastic error model is only appropriate when the source of noise is due only to errors that are entangled with bath states that are mutually orthogonal to all other bath states. Of course, this is not true in general with coherent errors being the extreme counter-example. A simulation employing such an approximation, while it may be able to bound or approximate a mean threshold, will not be able to reproduce the output distribution that contains information from the fully coherent sum of all fault paths. As we have shown here, even for very small perturbations, the coherent addition of these fault paths cause marked differences in the output statistics. This also implies that any analysis that contains a perfect QEC measurement cycle and correction will, by design, cut off these coherent pathways, also corrupting the output statistics.

Finally, we note that these output distributions can be used as a more robust metric for QEC performance than just considering the logical error rate. They would allow one to examine whether ideas to randomize coherent er-
rors, such as randomized compiling [45] do indeed help by making the errors look more “Pauli” like to QEC. The authors do not expect to see any benefit from randomized compiling as the evolution for a single trajectory would still be completely coherent, but this is a question for further investigation.

ACKNOWLEDGMENTS

We acknowledge useful conversations with Joan Hoffmann and Ken Brown. This project was supported by the Intelligence Advanced Research Projects Activity via Department of Interior National Business Center contract number 2012-12050800010. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

[1] P. W. Shor, Phys. Rev. A 52, R2493 (1995).
[2] A. Steane, Physical Review Letters 77, 793 (1996).
[3] A. Calderbank and P. Shor, Physical Review A 54, 1098 (1996).
[4] D. Aharonov and M. Ben-Or, in Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing, STOC ’97 (ACM, New York, NY, USA, 1997) pp. 176–188.
[5] E. Knill, R. Laflamme, and W. H. Zurek, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 454, 365 (1998).
[6] P. Aliferis, D. Gottesman, and J. Preskill, Quantum Information and Computation 6, 97 (2006).
[7] D. Aharonov and M. Ben-Or, “Fault-tolerant quantum computation with constant error rate,” (2008) pp. 1207–1282.
[8] D. Gottesman, in Quantum Information Science and Its Contributions to Mathematics (Amer. Math. Soc., 2010) pp. 13–58.
[9] B. M. Terhal, Rev. Mod. Phys. 87, 307 (2015).
[10] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, 10th ed. (Cambridge University Press, 2010).
[11] F. Haake, Quantum Signatures of Chaos, 2nd ed. (Springer, 2001).
[12] H.-P. Breuer and F. Petruccione, The Theory of Open Quantum Systems (Oxford, 2002).
[13] U. Weiss, Quantum Dissipative Systems, 4th ed. (World Scientific, 2012).
[14] E. Knill and R. Laflamme, quant-ph/9608012 , 1 (1996).
[15] P. W. Shor, in Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on (IEEE, 1996) pp. 56–65.
[16] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, New Journal of Physics 18, 012002 (2016).
[17] R. Kueng, D. M. Long, A. Vohra, and S. T. Flammia, Phys. Rev. Lett. 117, 170502 (2016).
[18] M. Gutiérrez, L. Svec, A. Vargo, and K. R. Brown, Physical Review A 87, 030302(R) (2013).
[19] M. R. Geller and Z. Zhou, Physical Review A 88, 012314 (2013).
[20] D. Puzzuoli, C. Granade, H. Haas, B. Criger, E. Magee, and D. G. Cory, Physical Review A 89, 022306 (2014).
[21] W. M. Witzel, K. Young, and S. D. Sarma, Physical Review B 90, 115431 (2014).
[22] A. S. Darmawan and D. Poulin, arXiv preprint arXiv:1607.06460 (2016).
[23] M. Gutiérrez, C. Smith, L. Lulushi, S. Janardan, and K. R. Brown, Phys. Rev. A 94, 042338 (2016).
[24] D. Gottesman, arXiv preprint quant-ph/9807006 (1998).
[25] S. Aaronson and D. Gottesman, Physical Review A 70, 052328 (2004).
[26] B. Rahn, A. C. Doherty, and H. Mabuchi, Phys. Rev. A 66, 032304 (2002).
[27] H. Bombin and M. A. Martin-Delgado, Phys. Rev. A 76, 012305 (2007).
[28] C. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, New Journal of Physics 14, 123011 (2012).
[29] A. Kitaev, Annals of Physics 303, 2 (2003).
[30] A. Cross, D. P. DiVincenzo, and B. M. Terhal, quant-ph/0711.1556v2 x, 1 (2009).
[31] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, Photons and Atoms (Wiley, 1989).
[32] W. Heitler, The Quantum Theory of Radiation, 3rd ed. (Dover, 1954).
[33] R. Ernst, G. Bodenhausen, and A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions (Oxford, 1987).
[34] A. Yariv, Optical Electronics in Modern Communications, 5th ed. (Oxford, 1996).
[35] L. Mandel and E. Wolf, Optical Coherence and Quantum Optics (Cambridge University Press, 1995).
[36] J. Barnes and W. Warren, Physical Review A 60, 4363 (1999).
[37] J. T. Merrill and K. R. Brown, “Progress in compensating pulse sequences for quantum computation,” in Quantum Information and Computation for Chemistry (John Wiley and Sons, 2014) pp. 241–294.
[38] Scientific, 2012).
[39] D. Gottesman, arXiv preprint quant-ph/9807006 (1998).
[40] K. M. Svore, A. W. Cross, I. L. Chuang, and A. V. Aho, Quantum Information and Computation 6, 193 (2006).
[41] Y. Tomita and K. M. Svore, Phys. Rev. A 90, 062320 (2014).
[42] A. G. Fowler, A. C. Whiteside, and L. C. L. Hollenberg,
Phys. Rev. Lett. 108, 180501 (2012).
[43] J. Edmonds, Canad. J. Math. 17, 449 (1965).
[44] J. Edmonds, J. Res. Nat. Bur. Standards 69B, 125 (1965).
[45] J. J. Wallman and J. Emerson, Phys. Rev. A 94, 052325 (2016).