Tractable simulation of error correction with honest approximations to realistic fault models

Daniel Puzzuoli,[1,2] Christopher Granade,[2,3] Holger Haas,[2,3] Ben Criger,[4] Easwar Magesan,[5] and D. G. Cory[2,6,7]

[1]*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[2]*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[3]*Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[4]*Institute for Quantum Information, RWTH Aachen University, D-52056 Aachen, Germany*
[5]*IBM TJ Watson Research Center, Yorktown Heights, New York, USA*
[6]*Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[7]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada N2L 2Y5*

In previous work, we proposed a method for leveraging efficient classical simulation algorithms to aid in the analysis of large-scale fault-tolerant circuits implemented on hypothetical quantum information processors. Here, we extend those results by numerically studying the efficacy of this proposal as a tool for understanding the performance of an error-correction gadget implemented with fault models derived from physical simulations. Our approach is to approximate the arbitrary error maps that arise from realistic physical models with errors that are amenable to a particular classical simulation algorithm in an "honest" way; that is, such that we do not underestimate the faults introduced by our physical models. In all cases, our approximations provide an "honest representation" of the performance of the circuit composed of the original errors. This numerical evidence supports the use of our method as a way to understand the feasibility of an implementation of quantum information processing given a characterization of the underlying physical processes in experimentally accessible examples.

I. INTRODUCTION

Quantum computing is the theory and practice of using systems which exhibit the properties of quantum mechanics to store and process information, allowing certain computational problems to be solved with greater speed than any known classical algorithm [1–3]. This work addresses a gap in efforts towards developing a practical quantum computer, namely that between our understanding of the physics of existing small quantum systems and methods for reasoning about the performance of large fault-tolerant circuits. We provide a concrete algorithm that is tractable given currently available classical computational resources and that enables reasoning about the performance of large-scale quantum information processors using experimental evidence in small systems. Our method accomplishes this task by extending the applicability of simulation algorithms for limited error models with *honest* approximations of errors outside these models. This is of immediate importance given that several modalities have been proposed for the development of large-scale quantum information processors, including quantum dots [4] and superconducting qubits [5].

The development and implementation of fault-tolerant quantum error correction (QEC) is a key milestone towards the development of large-scale quantum computation. Therefore, understanding the thresholds, overhead, and resource requirements of fault-tolerance methods is a critical step towards evaluating the feasibility of proposals for large-scale quantum information processing. For some such fault-tolerance schemes, we can analytically find thresholds on the acceptable errors, such that for any error of a weaker rate than the threshold, we can arbitrarily reduce the logical error rate. However, analytic thresholds have not yet been proven for fault-tolerance proposals based on topological properties, such as

surface-code based implementations [6]. For these proposals, numerical simulations are performed based on restricted error models that admit efficient classical simulation algorithms. Connecting these numerically simulated thresholds to models based on the physics of a device is a pressing concern in the development and appraisal of quantum information processing (QIP) proposals.

In recent years, proof-of-principle experiments have provided a better understanding of the physics underlying candidate QIP devices by implementing and fully characterizing quantum algorithms on small systems [7]. Such characterization techniques, however, are based on quantum process tomography [8] and are thus exponentially expensive in the size of the system. Though techniques exist to improve the characterization of quantum systems by using classical [9] and quantum [10] simulation resources, the exponential cost is inevitable if we demand that the system be characterized in terms of a quantum channel, due to the number of parameters that must be estimated. Thus, we cannot practically hope to directly characterize intermediate or large systems to demonstrate feasibility of QIP proposals.

A more attractive option, then, is to apply *models* of noise derived from experimental characterization of small instances of, or from simulations of the physics underlying, a proposed device. Such realistic noise models are not in general efficient to simulate, however, such that the question naturally arises: How does one leverage the efficient simulation algorithms for limited noise models to reason about the performance of large quantum information processors, using the *realistic* physical models afforded by small experimental examples? Crucially, we demand that we extrapolate small models to simulations of large quantum information processors in an *honest* fashion; that is, such that errors are only ever exaggerated and are never underestimated.

Doing so would enable the analysis of the performance of circuits implemented on a particular device and would in turn provide insight into whether an implementation would likely succeed using a given fault-tolerance scheme. Additionally, an honest model provides a useful tool for other tasks, such as understanding resource costs of a device that is already well below the threshold. Potential applications such as these are motivated by already existent numerical studies of circuit performance that are not tied to any particular device. These studies depend on the existence of efficiently simulable *subtheories* of quantum mechanics.

The most commonly used subtheory stems from the Gottesman-Knill (GK) theorem [11], which provides an efficient classical simulation algorithm for acting Clifford gates and Pauli measurements on stabilizer states. Coupling this with Monte Carlo (MC) techniques, stabilizer circuits with faults modeled as the probabilistic application of stabilizer circuit elements can be simulated efficiently. While this report works exclusively with error models useful for GK-MC simulation, the general method described is independent of the set of efficiently simulable channels, making it possibly useful within the context of other efficient simulation results, such as Wigner function simulation [12], match-circuit simulation [13], quantum normalizer circuit simulation [14,15], or the nonadaptive strong-simulation algorithm for Clifford circuits [16]. GK-MC simulations have been used to numerically estimate threshold error rates in topological codes [6,17,18] and have been augmented with sequential Monte Carlo techniques to enable reasoning about overhead required even when well below threshold [19]. Given the existing applications of GK-MC simulation, and the potential utility of other efficient simulation algorithms, it is highly desirable to use these techniques to aid in the understanding of real device performance. Understanding how to do this could also provide insights into how to compare the performance of experimentally realized gates to analytically derived thresholds for limited error models.

The challenge of applying efficient simulation techniques to realistic systems lies in the simple fact that errors for physical systems will, in general, not fit into one of the efficient simulation formalisms. In Ref. [20], we proposed a solution to this problem via the concept of *honest* error approximations. The idea is to replace an arbitrary error $\mathcal{E}$ with a channel Λ from a restricted class S, such that Λ honestly represents the tendency of $\mathcal{E}$ to preserve or distort quantum information. If S is a subset of efficiently simulable channels, Λ will be an efficient, honest estimate of the error induced by $\mathcal{E}$. While we work exclusively with our definition of honesty, we stress the generality of this proposal.

We define *honesty* using the operationally motivated concept of *distinguishability*. Let $\| \cdot \|_1$ denote the Schatten 1-norm [21]. For two quantum states ρ_0 and ρ_1, the quantity $\frac{1}{2} + \frac{1}{4}\|\rho_0 - \rho_1\|_1$ is the optimal success probability of inferring the value of a bit $\alpha \in \{0, 1\}$, drawn with uniform probability, given the state ρ_α, using a single measurement [22]. The error that a noise map $\mathcal{E}$ induces on a state ρ is then quantified as $\|\rho - \mathcal{E}(\rho)\|_1$, which we term the input-output (IO) distinguishability. This provides a strong operationally motivated definition of error; given a 50% chance that $\mathcal{E}$ acts on ρ, the IO distinguishability tells us what the optimal

probability of "noticing" its action is. Using this quantification of error, the following optimization problem for finding an honest approximation to a given noise map $\mathcal{E}$ was given in Ref. [20] as follows:

$$\text{Minimize: } \|\Lambda - \mathcal{E}\|_\diamond$$
$$\text{Subject to: for every pure state } \rho, \quad\quad (1)$$
$$\|\rho - \Lambda(\rho)\|_1 \geqslant \|\rho - \mathcal{E}(\rho)\|_1,$$

where Λ ranges over some desirable subset of quantum channels (e.g., Pauli channels) and $\| \cdot \|_\diamond$ is the diamond norm, which provides an analog to distinguishability for channels [23].

The constraint of the above problem encodes our definition of honesty; if the constraint is satisfied for two maps Λ and $\mathcal{E}$, we say that Λ is an "*honest* representation" of $\mathcal{E}$. The main result of Ref. [20] is an easy-to-compute, state-independent condition on the maps Λ and $\mathcal{E}$ which is sufficient to ensure honesty in the qubit case. For higher dimensions, the property ensures something similar to honesty, where the Schatten 1-norm is replaced by the 2-norm, though, to date, all instances that we have generated have also been found to satisfy Eq. (1) when tested using random pure states. We use this simpler condition in an optimization problem detailed in Appendix A 1 to numerically find honest approximations in this work. Details on the actual implementation of the problem, as well as a definition of the diamond norm, can also be found in Appendix A 1. We note that work similar to Ref. [20] has highlighted the utility of efficiently simulable measurements for modeling nonunital errors [24].

While our method ensures that the individual gate errors are honestly represented, its utility as a means for determining overall circuit performance depends on the preservation of the honesty of the approximations under composition and tensor products. (For brevity, we will abuse terminology slightly by using the word "composition" to refer to both direct composition and tensor products of maps.) That is, we are concerned with the degree to which a composition of honest approximations of gate elements is itself an honest approximation of the circuit composed of the individual gates in question.

In this paper, we numerically demonstrate that honesty is preserved under composition for some well-motivated physical models when used in a typical QEC circuit, providing evidence for the general applicability of our method. The circuit we simulate is a stabilizer circuit, and thus we approximate gate errors as probabilistic applications of stabilizer circuit elements to make it efficiently simulable using GK-MC. Specifically, we consider approximating errors as Pauli channels and as mixed-Clifford channels (probabilistic application of Pauli gates and Clifford gates, respectively). Our analysis is carried out within a generic simulation schema, described in detail in the next section, which emulates analysis of a physical system's ability to perform a QIP task, as per Refs. [4,25]. This schema serves as a demonstration of how we imagine our approximation method can be used and tests its value within such a use case. We find that, in all tested cases, our approximations compose well; that is, the approximated circuits honestly represent the performance of the originals.

Additionally, we include performance statistics for "Pauli twirling," which is another way of generating a Pauli channel from an arbitrary error, within our simulation schema. In Ref. [20], we showed that Pauli-twirled approximations can underestimate the error induced by the original channel (as per our definition) and thus may not be useful for circuit performance evaluation. In Ref. [26], Geller and Zhou ask if the Pauli-twirled approximations are "sufficiently good." The motivation for their work is that the Pauli-twirled approximation is efficiently computable in the dimension of the system, given a full description of a map, and that twirled channels are in general easier to estimate experimentally than the original [27–29]. This is in contrast to our method, which is not necessarily efficient to compute in the dimension. They argue that, while the Pauli-twirled approximations sometimes underestimate the failure probability for the task they consider, the amount of underestimation is small. Our simulations, in conjunction with Geller and Zhou's work, demonstrate good and poor regimes of performance for the Pauli-twirled approximations, in terms of circuit performance evaluation, with Geller and Zhou's error models falling into the good regime. In Appendix C, we identify and examine these regimes in detail and argue that the poorly performing regime may be more representative of the types of errors typically found in experiment.

II. SIMULATION SCHEMA

Our simulation schema begins with a low-level physical model and ends with a high-level, efficient simulation of a QEC circuit, with our approximation method being a bridge between the two. A physical model is a description of the continuous-time dynamics of a candidate physical system for QIP, consisting of deterministic and stochastic parts to internal and control Hamiltonians, as well as dissipative open quantum-system dynamics, and includes constraints on control amplitudes. Using control techniques, a gate set is generated from a physical model, which consists of all elementary quantum logic gates required for the desired QEC circuit. Once a gate set is generated, the error on each gate is approximated using our method, yielding an *honest* representation of the original gate set, which is used in an efficient simulation of the desired QEC circuit.

The circuit that we simulate, shown in Fig. 1, is a gadget that performs one round of error correction in an $[[n,k,d]]$ stabilizer code, which we implement using the five-qubit perfect code (a $[[5,1,3]]$ code). The circuit was chosen by balancing the desire that it be representative of standard practices, with the

requirement that it be small enough to allow for fast simulation of arbitrary gate errors, thereby allowing for comparison of the efficiently simulable errors to the original. For us, a gate set consists of the gates $\{\mathbb{1}, X, Y, Z, H, \mathrm{CNOT}\}$, where the first four are standard single qubit Pauli gates, H is the Hadamard gate, and CNOT is the two-qubit controlled-NOT gate. Note that the circuit we ultimately simulate uses only the $\mathbb{1}$, H, and CNOT gates. The rest are included for further comparison of the gate approximations.

We implement this procedure for three gate sets, generated from two physical models with varying model parameters and control techniques. The physical models and control techniques are chosen to be representative of those found in experiment. Doing so allows us to encounter errors not typically considered in fault-tolerance research, despite naturally occurring in physical implementations. We emphasize, however, that neither the approximation method nor the procedure for testing it given here have been tailored for a particular outcome. The method is generic; it is independent of both the underlying physical model and gates, as well as the QEC circuit.

III. ERROR COMPOSITION

Before walking through our implementation of this scheme, we construct a simple example in which honest approximations compose dishonestly. This example demonstrates the inherent limitations of approximating an error with another error that composes in a fundamentally different way. Consider the two single-qubit maps

$$\Gamma(\rho) = U(\theta)\rho U^{\dagger}(\theta), \tag{2}$$

$$\Lambda(\rho) = (1 - p)\rho + pZ\rho Z, \tag{3}$$

where $U(\theta) = \exp(-i\frac{\theta}{2}Z)$ and Z is the Pauli z operator. In Ref. [20], it was shown that if $p = |\sin(\theta/2)|$, then Λ and Γ have identical IO distinguishability properties; that is,

$$\|\rho - U(\theta)\rho U(\theta)^{\dagger}\|_1 = \|\rho - \Lambda(\rho)\|_1, \tag{4}$$

for all ρ and therefore Λ is an honest representation of Γ. Consider the state $\rho_+ = |+\rangle\langle+|$, where $X|+\rangle = |+\rangle$, and X is the Pauli x operator, which is chosen as it is maximally sensitive to both $U(\theta)$ and Λ. One can check that

$$\|\rho_+ - \Gamma(\rho_+)\|_1 = \|\rho_+ - \Lambda(\rho_+)\|_1 = 2|\sin(\theta/2)| = 2p. \tag{5}$$
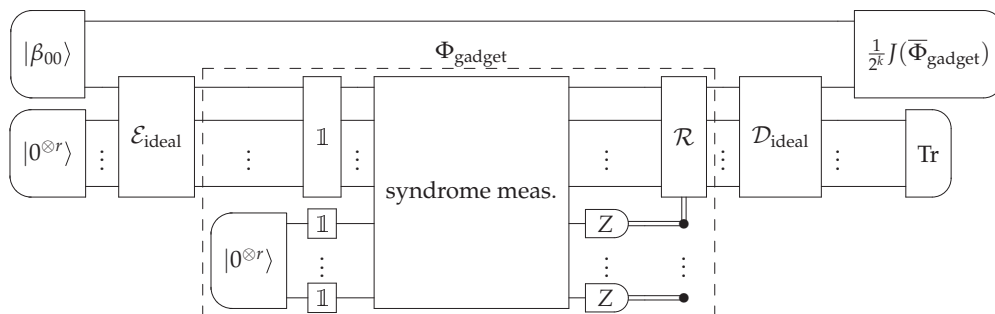

FIG. 1. Circuit to produce a Choi state for the logical action $\overline{\Phi}_{\mathrm{gadget}}$ of a QEC gadget acting on an $[[n,k,d]]$ stabilizer code, where $r \equiv n - k$.

Now, consider the circuit composed of two applications of Γ, $C = \Gamma \circ \Gamma$, and the approximate circuit $C^{(a)} = \Lambda \circ \Lambda$, where

$$C^{(a)}(\rho) = [(1 - p)^2 + p^2]\rho + 2p(1 - p)Z\rho Z. \qquad (6)$$

For $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, it can be checked that

$$\|\rho_+ - C(\rho_+)\|_1 > \|\rho_+ - C^{(a)}(\rho_+)\|_1, \qquad (7)$$

and thus the composition of approximations is not an honest representation of the original circuit.

This example can be understood by looking at how repetitive application of these channels affects the state. Each application of Γ, a unitary error, deterministically rotates the state by a small angle, whereas Λ, a dephasing error, rotates the state by $180°$ but with a small probability. In the former case, the distance that each application moves the current state remains constant, whereas in the latter case, this distance decreases exponentially, resulting in an underestimation of errors after only two applications. We encounter this situation in our simulations; one of the gate sets we consider has an identity gate error that is essentially a unitary about the z axis, and the honest Pauli approximation is the dephasing channel that reproduces its IO distinguishability properties. Given this discussion and the frequency with which the gate occurs, we expected that our approximations might underestimate the overall circuit error. However, even in this case, our approximations perform as desired, providing strong numerical evidence for the value of this method in QEC circuits.

Despite these potential difficulties with error composition, it remains possible that, after a QEC protocol is applied, the resulting effective errors might compose more desirably. As an example, the first step of QEC in stabilizer codes is measuring the *error syndrome*. This consists of measuring a generating set of stabilizer elements $\{Q_i\}_{i=1}^k$, which produces a k-bit string b with $b_i = 1$ if the outcome from measuring Q_i is -1 and $b_i = 0$ if it is $+1$. If a particular string b is measured, then the system is projected onto the subspace defined by the projector

$$\Pi_b = 2^{-k}[\mathbb{1} + (-1)^{b_1}Q_1] \cdots [\mathbb{1} + (-1)^{b_k}Q_k]. \qquad (8)$$

For a Pauli operator P and codeword $|\psi\rangle$, $\Pi_b P |\psi\rangle = 0$ if P does not produce syndrome b and $\Pi_b P |\psi\rangle = P |\psi\rangle$ if P produces syndrome b. Thus, indexing the Pauli operators as $\{P_i\}$ and denoting S_b as the set of indices for Pauli operators that produce syndrome b, if a particular syndrome b is measured after an error Λ, having χ matrix χ_{ij} in the Pauli basis [8], acts on an arbitrary codeword $|\psi\rangle$, the state will be (ignoring normalization)

$$\Pi_b \Lambda(|\psi\rangle \langle\psi|) \Pi_b = \sum_{ij} \chi_{ij} \Pi_b P_i |\psi\rangle \langle\psi| P_j \Pi_b$$

$$= \sum_{ij \in S_b} \chi_{ij} P_i |\psi\rangle \langle\psi| P_j.$$

Given this form, it is clear that if P_i and P_j have different syndromes, then χ_{ij} can play no part in the postsyndrome measurement state and is therefore effectively truncated by syndrome measurement. In this way, errors become more "incoherent" and this, at least superficially, makes errors "more like" Pauli channels (which have diagonal χ matrices in the Pauli basis). Thus, whatever the form of Λ, after a correction step is enacted, the effective error may compose more like

a Pauli channel. In practice, this argument may fail due to various aspects of imperfect syndrome measurement, such as limited visibility measurements, and the time it takes to perform measurement protocols like ancilla-assisted syndrome measurement.

IV. IMPLEMENTING THE SCHEMA

A. Physical models and gate set generation

We consider two physical models, PM1 and PM2. PM1 is motivated by a double quantum dot system and PM2 represents an archetypal two-level system (see Appendix B for a description and full details). Gate Set 1 (GS1) is built on PM1, and Gate Sets 2 and 3 (GS2 and GS3) are built on PM2 (using different model parameters). GS1 and GS2 use noise refocusing techniques [30,31], which mitigate errors induced by stochastic Hamiltonians. GS2 and GS3 implement gates via hard pulses; that is, the pulse sequences used to generate the gates are manually specified by choosing control amplitudes. Due to the complicated structure of the Hamiltonian in PM1, optimal control theory (OCT) was used to find pulse sequences that implement the gates in GS1 with high fidelity [32,33]. Every gate in a set is made to be the same length in time, as our circuit simulation proceeds in discrete time steps in which a single gate acts on every register qubit. A full description of how the gates are simulated is given in Appendix A 2.

The different combinations of physical model and control techniques give rise to different types of gate errors. A detailed account on the form of the errors for each gate set is given in Appendix C, as it has particular relevance within the context of that discussion. We do, however, wish to highlight that some gates in GS1 have largely unitary errors, resulting from the use of OCT pulse finding in gate implementation. Thus, given the discussion on error composition, GS1 provides a strong test for our method.

B. Gate set approximations and statistics

For each gate set we generate three efficiently simulable approximate gate sets:

Pauli twirled: The Pauli-twirled errors.

Pauli: The honest Pauli channel approximation.

Clifford: The honest mixed-Clifford channel approximation. Note that only the single-qubit errors are approximated as mixed Cliffords (we allowed the algorithm to search over all mixed Cliffords). Due to the large number of two-qubit Cliffords, the honest Pauli approximation for the CNOT gate is reused.

Several metrics are used to compare how each approximation performs on individual gates. In what follows we denote the noisy implementation of some ideal operation U_{ideal} by $\Lambda_{\text{Original}}$, and use Λ as a place holder for the various approximations. The first few metrics are well-known quantities.

(i) χ_{00}: The first entry of the χ-process matrix of the error in the Pauli basis [8]. This quantity is reported due to its relation to the average gate fidelity [34] and, for Pauli channels, is the probability that no fault occurs.

(ii) $\|\Lambda - U_{\text{Ideal}}\|_\diamond$ and $\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$: The distance of the approximation Λ to the ideal gate and original error, respectively.

TABLE I. Statistics for the various approximations of the identity gate for each gate set, approximated from $N = 10^6$ random pure states.

	Statistics	Original	Pauli twirled	Pauli	Clifford
	χ_{00}	0.999994	0.999994	0.997618	0.998314
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.76×10^{-3}	1.20×10^{-5}	4.76×10^{-3}	4.77×10^{-3}
GS1	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.76×10^{-3}	6.73×10^{-3}	3.64×10^{-3}
	$\bar{h}$		-3.73×10^{-3}	1.14×10^{-7}	1.64×10^{-6}
	p_{viol}		1.0	0.0	0.0
	χ_{00}	0.999087	0.999087	0.999085	0.999086
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}
GS2	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.48×10^{-5}	2.50×10^{-5}	2.06×10^{-6}
	$\bar{h}$		-1.63×10^{-7}	2.65×10^{-6}	8.34×10^{-7}
	p_{viol}		0.49861	0.0	0.0
	χ_{00}	0.998751	0.998751	0.996501	0.996501
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.99×10^{-3}	2.50×10^{-3}	7.00×10^{-3}	7.00×10^{-3}
GS3	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.50×10^{-3}	5.28×10^{-3}	5.28×10^{-3}
	$\bar{h}$		-1.03×10^{-3}	1.91×10^{-3}	1.91×10^{-3}
	p_{viol}		0.74978	0.0	0.0

The rest stem from our definition of honesty. Using the function

$$h(\Gamma, \mathcal{E}, \rho) \equiv \|\rho - \Gamma(\rho)\|_1 - \|\rho - \mathcal{E}(\rho)\|_1, \qquad (9)$$

which we call the *hedging* of the channel Γ relative to $\mathcal{E}$ for the state ρ, the statement that Λ honestly represents the error of $\Lambda_{\text{Original}}$ can be restated as $h(\Lambda, \Lambda_{\text{Original}}, \rho) \geqslant 0$ for all pure states ρ. We calculate three quantities related to the hedging, which we approximate by randomly sampling N pure states $\{|\psi_i\rangle\}_{i=1}^N$.

　(a) $\bar{h}(\Lambda, \Lambda_{\text{Original}}) \equiv \int d\psi \, h(\Lambda, \Lambda_{\text{Original}}, |\psi\rangle\langle\psi|)$
$\approx \frac{1}{N} \sum_{i=1}^N h(\Lambda, \Lambda_{\text{Original}}, |\psi_i\rangle\langle\psi_i|)$: The average of the hedging function over pure states.

　(b) $p_{\text{viol}} \approx \frac{N_{\text{viol}}}{N}$: The ratio of pure states $|\psi\rangle$ for which $h(\Lambda, \Lambda_{\text{Original}}, |\psi\rangle\langle\psi|) < 0$, where $N_{\text{viol}} = |\{|\psi\rangle \in \{|\psi_i\rangle\}_{i=1}^N : h(\Lambda, \Lambda_{\text{Original}}, |\psi\rangle\langle\psi|) < 0\}|$.

Table I presents the statistics for the identity gate from each gate set, using $N = 10^6$ uniformly sampled pure states. See Appendix D for tables containing statistics on all gates.

Looking at the various diamond norm distances among the Original, Ideal, Pauli twirled (Λ_{PT}), and Pauli errors (Λ_{P}), a simple ordering can be seen to hold for every gate,

$$\|\Lambda_{\text{PT}} - U_{\text{Ideal}}\|_\diamond \leqslant \|\Lambda_{\text{Original}} - U_{\text{Ideal}}\|_\diamond \leqslant \|\Lambda_{\text{P}} - U_{\text{Ideal}}\|_\diamond,$$

$$\|\Lambda_{\text{PT}} - \Lambda_{\text{Original}}\|_\diamond \leqslant \|\Lambda_{\text{P}} - \Lambda_{\text{Original}}\|_\diamond.$$

Thus, while the Pauli-twirled error is always closer than the Pauli to the Original, it is also always closer to the Ideal than the Original and is therefore a less noticeable error than the Original. The location of the Clifford approximation in the second inequality chain varies; for some gates, it is an order of magnitude closer to the Original than both the Pauli and Pauli-twirled approximations, and, for others, it is the same distance to the Original as the Pauli, indicating that the best Pauli is also the best Clifford approximation, and, for the rest, it is between the two.

Two other important and connected observations can be made. In some cases, the χ_{00} element of the honest approximations is much lower than that of the Original, and in others it does not appreciably differ. In the former cases,

the Pauli-twirled approximations tend to be much closer to the Ideal, and have worse hedging performance, than in the latter cases. These are demonstrations of channels with different average fidelities but similar IO distinguishability properties and channels with identical average fidelities but very different IO distinguishability properties. This observation is connected to the different regimes of performance for Pauli twirling, as well as how "coherent" or "unitary" an error is, and is explained in detail in Appendix C.

C. Circuit design and simulation results

We simulate the gadget Φ_{gadget} that performs one round of error correction on one block of an error-correcting code, as per Fig. 1. We isolate the action $\overline{\Phi}_{\text{gadget}}$ of this gadget on the encoded state by preceding and following it with perfect encoding and decoding operations, $\mathcal{E}_{\text{ideal}}$ and $\mathcal{D}_{\text{ideal}}$. The circuit is simulated by computing its Choi state; for a code that encodes k logical qubits into n physical qubits, we take the state

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2^k}} \sum_{i=1}^{2^k} |i\rangle \otimes |i\rangle, \qquad (10)$$

where $\{|i\rangle\}_{i=1}^{2^k}$ is an orthonormal basis for the k-qubit Hilbert space, and compute

$$(C \otimes \mathbb{1}_{L(\mathbb{C}^k)})(|\beta_{00}\rangle\langle\beta_{00}|) = \frac{1}{2^k} J(C), \qquad (11)$$

where C represents the entire circuit and $J(C)$ is the Choi-Jamiołkowski matrix for the circuit. (This simulation method is chosen due to the efficiency with which maps of the form we consider can be applied to states. See Appendix A 3 for details.) Due to the perfect encoding and decoding operations, $J(C) = J(\overline{\Phi}_{\text{gadget}})$. Explicitly, the circuit performs the following operations:

(1) The state $|\beta_{00}\rangle$ [Eq. (10)] is prepared, half of which is perfectly encoded into an $[\![n, k, d]\!]$ stabilizer code.

(2) The gadget Φ_{gadget} is applied to the encoded physical qubits, consisting of the following:

(a) Original circuit

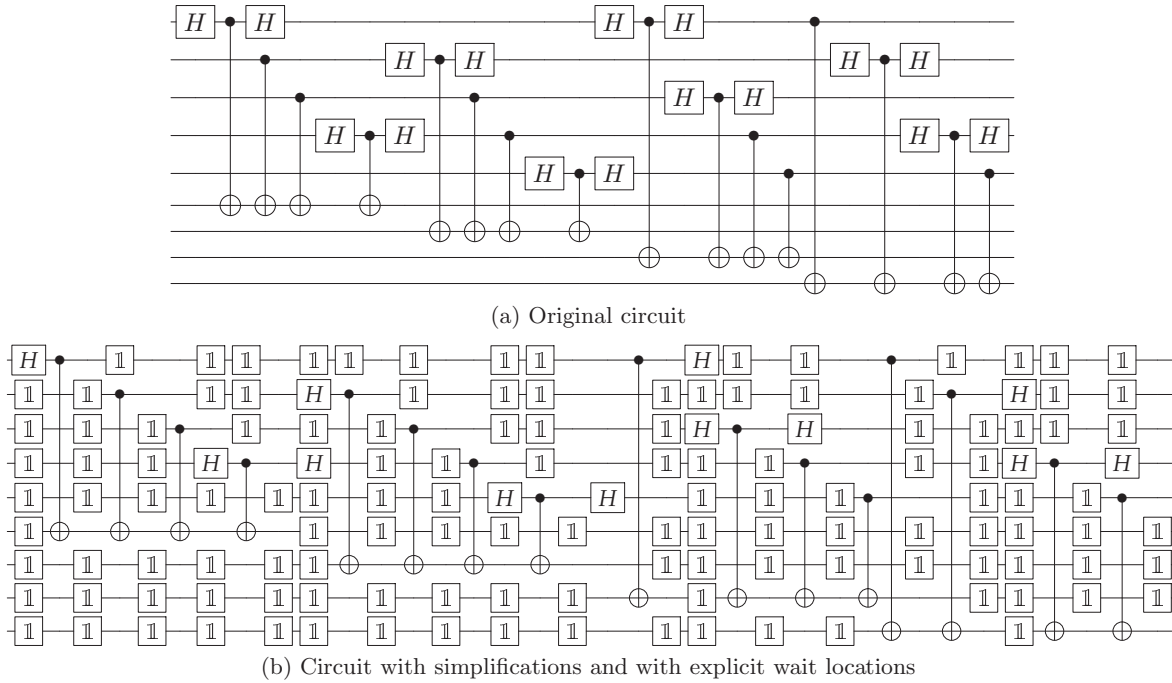(b) Circuit with simplifications and with explicit wait locations

FIG. 2. Syndrome measurement circuit for the five-qubit perfect code.

(i) One imperfect wait location on all of the data qubits. This is a placeholder for possible nontrivial operations in gadgets meant to perform logical operations.

(ii) Simultaneously, a register of ancillas for ancilla-assisted syndrome measurement is prepared. An imperfect identity operation acts on each ancilla to represent imperfect ancilla preparation.

(iii) Imperfect ancilla-assisted syndrome measurement is performed (see Fig. 2). Measurement of the physical ancillas is taken to be perfect, with errors represented by identity gates that precede the measurement.

(3) A perfect recovery operation is performed by classical feed-forward of syndrome measurement details (see Fig. 3).

(4) Once Φ_{gadget} is done, the resultant state is then perfectly decoded and the physical ancillas are discarded.

The recovery operation is chosen to be perfect as, in practice, it isn't always necessary to physically perform the recovery; errors can be tracked and taken into account when further operations are performed on the block [35].
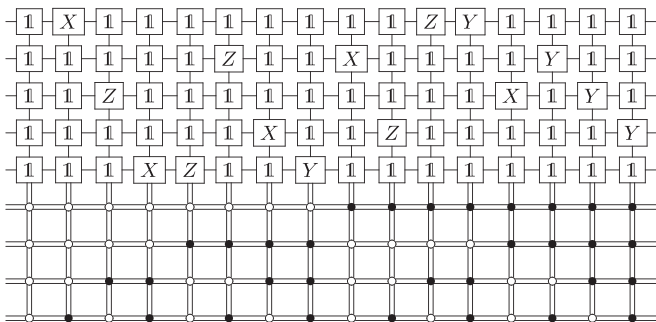
We implement this simulation schema using the five-qubit perfect code, with the syndrome measurement and recovery circuits designed by the PYTHON package QUAEC [36]. Note that we choose to perform the syndrome measurement in a non-fault-tolerant way, as a fault-tolerant gadget for a code with n physical qubits would require at least $2n$ ancilla qubits for the Steane or Knill fault-tolerant error correction (FTEC) gadgets or strictly more than $\sum_i \text{wt}(S_i) = 16$ ancillae for the Shor FTEC gadget. Thus, at least 10 ancillae are needed for the perfect code, requiring simulation of at least 16 qubits, putting us outside the range of quickly simulable circuits with arbitrary errors.

With four stabilizer generators, this code requires four ancillas for encoding and four for syndrome measurement using the circuit shown in Fig. 2. Any redundant Hadamard gates have been removed. The recovery operation is shown in Fig. 3. This type of non-fault-tolerant syndrome measurement is similar to gadgets proposed for use in topological QEC codes, such as the surface code. In particular, the syndrome measurement gadget used by Fowler *et al.* [6], shown in Fig. 4, relies on CNOT gates between each data qubit in the support of a stabilizer generator and a common ancilla qubit. We emphasize that we are not concerned with *absolute* circuit performance. Rather, the task at hand is the comparison of *relative* performance of efficiently simulable error approximations, so

FIG. 3. Recovery circuit for the five-qubit perfect code.
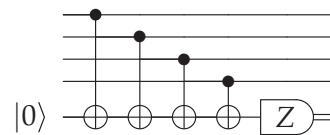

FIG. 4. Circuit to measure the stabilizer generator $Z^{\otimes 4}$, proposed by Fowler *et al.* [6] for use in surface codes.

TABLE II. Statistics for $\overline{\Phi}_{\text{gadget}}$ using $N = 10^6$ randomly sampled pure states.

	Statistics	Original	Pauli twirled	Pauli	Clifford
	χ_{00}	0.999964	0.999964	0.985820	0.989930
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.76×10^{-3}	7.28×10^{-5}	2.84×10^{-2}	2.04×10^{-2}
GS1	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.76×10^{-3}	2.87×10^{-2}	2.01×10^{-2}
	$\overline{h}$		-3.69×10^{-3}	1.85×10^{-2}	1.23×10^{-2}
	p_{viol}		1.0	0.0	0.0
	χ_{00}	0.991372	0.991372	0.991355	0.991367
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.73×10^{-2}	1.73×10^{-2}	1.73×10^{-2}	1.73×10^{-2}
GS2	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.45×10^{-5}	4.29×10^{-5}	1.14×10^{-5}
	$\overline{h}$		-1.63×10^{-8}	2.24×10^{-5}	7.66×10^{-6}
	p_{viol}		0.55566	0.0	0.0
	χ_{00}	0.992495	0.987594	0.969499	0.969499
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.51×10^{-2}	2.48×10^{-2}	6.10×10^{-2}	6.10×10^{-2}
GS3	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.03×10^{-2}	4.60×10^{-2}	4.60×10^{-2}
	$\overline{h}$		6.36×10^{-3}	3.04×10^{-2}	3.04×10^{-2}
	p_{viol}		0.0	0.0	0.0

it suffices that this circuit contains all of the typical elements and procedures for QEC, regardless of fault tolerance.

Table II gives the simulation statistics for $\overline{\Phi}_{\text{gadget}}$, again using $N = 10^6$ sample pure states to compute the hedging parameters. For each gate set, the Pauli and Clifford approximations compose well; the approximated circuit honestly represents the error of the original. This is especially encouraging for GS1, given its unitary identity error. For the Pauli-twirled errors, we see that in GS1 they fail the honesty condition for every tested pure state. For GS2, they fail the honesty condition for just over half of the pure states tested but by an arguably small degree. Interestingly, for GS3, the Pauli-twirled approximations provide an honest representation of the circuit performance.

V. CONCLUSION

In all examined cases, the honest approximations led to honest representations of circuit performance, providing confidence in our method as a tool for evaluating the performance of typical QEC circuits with realistic gate errors. By starting from continuous-time physical models, and building gates using common control techniques, we tested our method against errors typical of those found in experiment. The details of the physical models and control techniques were not tailor-made for any desired outcome. The strongest test of our method came from errors with strong unitary parts, arising from OCT designed pulses, a regime not typically considered in fault-tolerance research. Additionally, our results, in conjunction with the recent work by Geller and Zhou [26], demonstrate two regimes of performance for Pauli-twirled error approximations. In one regime, their performance can be considered "sufficiently good," while in the other, Pauli twirling results in systematic underestimation of the IO distinguishability notion of error (see Appendix C).

Our work is motivated by the desire for the simulations to be pessimistic. We want to be reasonably assured that, if the simulation with the approximated errors performs well according to some metric, then the actual implementation will

perform well also. Currently, experimental implementations of QIP are limited to small system sizes. Extrapolating their performance to hypothetical large-scale systems requires caution. Quantum processors will require constant application of error-correction protocols like the one we consider, and it is imaginable that in large systems, consisting of hundreds of qubits or more, even a small underestimation of the effect of physical-level errors may dramatically compound, resulting in false expectations of overall performance.

This work provides hope that our method can be used as a tool for extrapolating performance from small systems in an honest way. This is key to understanding the feasibility of various proposals for large-scale quantum devices, which in turn will aid in planning the way forward for feasible experimental implementations of QIP.

APPENDIX A: METHODS

1. Channel approximation and diamond norm computation

To measure the distance between two maps Λ and $\mathcal{E}$, we use the diamond norm distance $\|\Lambda - \mathcal{E}\|_\diamond$, which provides an analog to distinguishability for channels [23]. For any map

$\Phi : \mathrm{L}(\mathcal{H}_1) \to \mathrm{L}(\mathcal{H}_2)$ (mapping the linear operators acting on one Hilbert space to the linear operators acting on another), the diamond norm can be defined as follows:

$$\|\Phi\|_{\diamond} \equiv \max\{\|(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{H}_1)})(X)\|_1$$
$$: X \in \mathrm{L}(\mathcal{H}_1 \otimes \mathcal{H}_1), \|X\|_1 \leqslant 1\}, \qquad (A1)$$

where $\mathbb{1}_{\mathrm{L}(\mathcal{H}_1)}$ is the identity channel acting on $\mathrm{L}(\mathcal{H}_1)$. Given a set of channels S, to find an honest approximation of an error map $\mathcal{E}$, we seek to minimize $\|\Lambda - \mathcal{E}\|_{\diamond}$ for $\Lambda \in S$ such that Λ satisfies the constraint given in Ref. [20]. To implement the optimization, we require that S has a parametrization. For our purposes, it is natural to consider sets of the form $S = \{\sum_{i=1}^{n} p_i \Lambda_i : p_i \geqslant 0 \text{ and } \sum_{i=1}^{n} p_i = 1\}$. Given this, we have the following optimization problem:

Input: Finite set of channels $\{\Lambda_i\}_{i=1}^n$ and channel $\mathcal{E}$

Minimize: $f(p_1,...,p_n) = \left\| \sum_{i=1}^{n} p_i \Lambda_i - \mathcal{E} \right\|_{\diamond}$

s.t. $\sum_{i=1}^{n} p_i = 1, p_i \geqslant 0,$ and $A \geqslant 0,$ where

$$A = (\mathbb{1} - M_\Lambda)^T (\mathbb{1} - M_\Lambda) - (\mathbb{1} - M_\mathcal{E})^T (\mathbb{1} - M_\mathcal{E})$$
$$+ (\|\vec{t}_\Lambda\|_2^2 - \|\vec{t}_\mathcal{E}\|_2^2 - 2\|(\mathbb{1} - M_\Lambda)^T \vec{t}_\Lambda$$
$$- (\mathbb{1} - M_\mathcal{E})^T \vec{t}_\mathcal{E}\|_2)\mathbb{1}, \qquad (A2)$$

where $M_\Lambda = \sum_{i=1}^{n} p_i M_i, \vec{t}_\Lambda = \sum_{i=1}^{n} p_i \vec{t}_i, (M_i, \vec{t}_i)$ is the Bloch representation of channel Λ_i, $(M_\mathcal{E}, \vec{t}_\mathcal{E})$ is the Bloch representation of $\mathcal{E}$, and $\mathbb{1}$ is the identity matrix of appropriate size. Note that the constraint given here is slightly more general than the one in Ref. [20], as it includes the possibility for S to contain nonunital channels. We note again that if a qubit map Λ satisfies the constraint in the above optimization problem, then it is an *honest* representation of the error $\mathcal{E}$, according to the definition in the introduction. For higher dimensional channels, the constraint is sufficient to ensure something similar to honesty, where the Schatten 1-norm is replaced by the Schatten 2-norm, in which case the approximation Λ is still guaranteed to be somehow globally worse than $\mathcal{E}$, though not in an operationally motivated way. To date, however, all higher dimensional approximations that we have generated using this algorithm have also been found to be honest when tested using random pure states.

We implement the approximation optimization problem in MATLAB using the built in fmincon function. The diamond norm is computed using a semidefinite program given by Watrous [37] and implemented using the CVX package [38]. Linear constraints ensure that the vector $(p_1, \ldots, p_n)$ is a probability vector and nonlinear constraints check that the eigenvalues of the matrix A are non-negative with a tolerance of 10^{-15}. The SQP algorithm is used for the optimization. Due to the nonconvexity of this problem it is necessary to run many local solvers and then choose the best result. This is done using the MultiStart function which instantiates the local solver many times over randomly chosen starting points that satisfy the constraints. For each approximation, we used 72 starting points.

2. Cumulant simulation

To simulate quantum logic gates using our noise models, a method for the simulation of stochastic quantum evolution is required. This has been considered in the context of analyzing the fidelity with which decoherence-free subsystems can be implemented [39]. In that case, the cumulant expansion [40–42] was applied to model the effects of stochastic dynamics on a quantum system.

Following that approach, we will consider that, conditioned on a particular realization of noise, our system evolves according to the Liouville-von Neumann equation

$$\frac{\partial}{\partial t} \rho(t) = -i[H(t), \rho(t)] + D[\rho(t)], \qquad (A3)$$

where $\rho(t)$ is the density operator describing our system at time t, H is the Hamiltonian of the system, and $D \in \mathrm{T}(\mathcal{H})$ is a linear transformation describing the decoherence of the system. We assume that $H(t)$ can be decomposed into deterministic and stochastic parts,

$$H(t) = H_{\mathrm{det}}(t) + H_{\mathrm{st}}(t). \qquad (A4)$$

We then further decompose $H_{\mathrm{st}}(t)$ such that all of the stochasticity is encapsulated in a set of scalar-valued functions $\{\omega_1(t), \ldots, \omega_k(t)\}$. Thus,

$$H_{\mathrm{st}}(t) = \sum_i \omega_i(t) A_i(t) \qquad (A5)$$

for some set of deterministic operator-valued functions $\{A_i(t)\}$.

To analyze the dissipation transformation D, we assume that it can be written in Lindblad form,

$$D[\rho(t)] = \sum_i L_i \rho(t) L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho(t)\}, \qquad (A6)$$

where $\{L_i\}$ are called the Lindblad operators of the system.

Both the Liouvillian operator $\mathcal{L}[\rho(t)] := [H, \rho(t)]$ and the dissipation operator D act linearly on density operators and thus may be represented by superoperators $\hat{\mathcal{L}}, \hat{D} \in \mathrm{L}(\mathrm{L}(\mathcal{H}))$, where $\mathrm{L}(\mathcal{H})$ marks the set of all linear operators acting on Hilbert space $\mathcal{H}$. Using the isomorphism that $\mathrm{L}(\mathcal{H}) \cong \mathcal{H} \otimes \mathcal{H}$, we shall use the column-stacking basis for $\mathcal{H} \otimes \mathcal{H}$, such that $||i\rangle\langle j|\rangle\rangle = |j\rangle \otimes |i\rangle$. Therefore, one can rewrite Eq. (A3) as

$$\frac{\partial}{\partial t} |\rho(t)\rangle\rangle = (-i[\hat{\mathcal{L}}_{\mathrm{det}}(t) + \hat{\mathcal{L}}_{\mathrm{st}}(t)] + \hat{D})|\rho(t)\rangle\rangle. \qquad (A7)$$

Now we go to the rotating frame of the deterministic superoperator $\hat{\mathcal{L}}_{\mathrm{det}}(t)$, i.e., we define a unitary $\mathcal{U}(t) = \mathrm{T}\exp(-i\int_0^t \hat{\mathcal{L}}_{\mathrm{det}}(t')dt')$ such that

$$\frac{\partial}{\partial t}|\tilde{\rho}(t)\rangle\rangle = (-i\mathcal{U}^\dagger(t)\hat{\mathcal{L}}_{\mathrm{st}}(t)\mathcal{U}(t) + \mathcal{U}^\dagger(t)\hat{D}\mathcal{U}(t))|\tilde{\rho}(t)\rangle\rangle, \quad (A8)$$

where $|\tilde{\rho}(t)\rangle\rangle = \mathcal{U}^\dagger(t)|\rho(t)\rangle\rangle$.

The formal solution to Eq. (A8), then, for a single realization of the trajectories $\{\omega(t)\}$ is given by

$$|\tilde{\rho}(t)\rangle\rangle = \mathrm{T}\exp\left(-i\int_0^t \hat{\hat{G}}(t')dt'\right)|\rho(0)\rangle\rangle, \qquad (A9)$$

with $\hat{\hat{G}}(t) := \mathcal{U}^\dagger(t)\hat{\hat{\mathcal{L}}}_{\mathrm{st}}(t)\mathcal{U}(t) + i\mathcal{U}^\dagger(t)\hat{\hat{D}}\mathcal{U}(t)$.

For our purposes, we are interested in the average evolution $\hat{\hat{S}}$ over the ensemble of control trajectories,

$$\hat{\hat{S}}(t) = \left\langle \mathrm{T}\exp\left(-i\int_0^t \hat{\hat{G}}(t')dt'\right)\right\rangle. \qquad (A10)$$

The cumulant expansion gives us that $\hat{\hat{S}}(t) = \exp(\hat{\hat{K}}(t))$, where

$$\hat{\hat{K}}(t) = \sum_{n=1}^\infty \frac{(-it)^n}{n!}K_n = -it\hat{\hat{K}}_1 - \frac{t^2}{2}\hat{\hat{K}}_2 + \cdots, \qquad (A11)$$

$$\hat{\hat{K}}_1 = \frac{1}{t}\int_0^t dt_1\langle\hat{\hat{G}}(t_1)\rangle, \qquad (A12)$$

$$\hat{\hat{K}}_2 = \frac{1}{t^2}\mathrm{T}\int_0^t dt_1\int_0^t dt_2\langle\hat{\hat{G}}(t_1)\hat{\hat{G}}(t_2)\rangle - \hat{\hat{K}}_1^2. \qquad (A13)$$

To simplify this, we assume that each control parameter $\omega_i(t)$ is a trajectory of a stationary zero-mean process (The zero-mean assumption technically is not an assumption; the mean of each random process can be absorbed into the deterministic part of the Hamiltonian.) That is, that $\vec{\omega} \sim \mathrm{GP}(0, \vec{\vec{\phi}})$, where $\vec{\vec{\phi}}$ is the matrix-valued autocorrelation function for $\vec{\omega}(t)$, such that $\phi_{i,j}(t_1 - t_2) = \langle\omega_i(t_1)\omega_j(t_2)\rangle$.

Then $\hat{\hat{K}}_1$ becomes simply

$$\hat{\hat{K}}_1 = \frac{i}{t}\int_0^t dt_1\mathcal{U}^\dagger(t_1)\hat{\hat{D}}\mathcal{U}(t_1), \qquad (A14)$$

whereas we can then rewrite $\hat{\hat{K}}_2$ in terms of the autocorrelation function,

$$\hat{\hat{K}}_2 = \frac{2}{t^2}\int_0^t dt_1\int_0^{t_1} dt_2\sum_{i,j=1}^k \phi_{i,j}(t_1 - t_2)$$
$$\times \mathcal{U}^\dagger(t_1)\hat{\hat{A}}_i(t_1)\mathcal{U}(t_1)\mathcal{U}^\dagger(t_2)\hat{\hat{A}}_j(t_2)\mathcal{U}(t_2), \qquad (A15)$$

$$-\frac{2}{t^2}\int_0^t dt_1\int_0^{t_1} dt_2\mathcal{U}^\dagger(t_1)\hat{\hat{D}}\mathcal{U}(t_1)\mathcal{U}^\dagger(t_2)\hat{\hat{D}}\mathcal{U}(t_2) - \hat{\hat{K}}_1^2, \qquad (A16)$$

where $\hat{\hat{A}}_i(t) = -A_i^*(t) \otimes \mathbb{1} + \mathbb{1} \otimes A_i(t)$. In this way, we note that the cumulant expansion generalizes the Magnus expansion to account for stochastically varying fields. The motivation for using cumulants instead of expanding the time-ordered exponential in terms of moments of the stochastic process stems from the fact that cumulant averages enter in the exponential, reducing the risk of truncation artifacts.

To numerically simulate the gate action, we discretize $\hat{\hat{\mathcal{L}}}_{\mathrm{det}}(t')$ along our gate length t at N points, with equal time intervals Δt between these points, i.e., we evaluate $\{\hat{\hat{\mathcal{L}}}_{\mathrm{det}}(m\Delta t)\}$, with $m = 1, \ldots, N$ while $t = N\Delta t$. Next we approximate $\mathcal{U}(n\Delta)$ by

$$\mathcal{U}(n\Delta) \approx \exp(-i\hat{\hat{\mathcal{L}}}_{\mathrm{det}}(n\Delta t)\Delta t)\ldots\exp(-i\hat{\hat{\mathcal{L}}}_{\mathrm{det}}(\Delta t)\Delta t)$$
$$\times \exp(-i\hat{\hat{\mathcal{L}}}_{\mathrm{det}}(0)\Delta t). \qquad (A17)$$

Finally, we turn turn the integral in line (A14) into a sum

$$\hat{\hat{K}}_1 \approx \frac{i}{N}\sum_{n=0}^{N-1}\mathcal{U}^\dagger(n\Delta t)\hat{\hat{D}}\mathcal{U}(n\Delta t), \qquad (A18)$$

and the double integral in line (A15) into a double sum,

$$\hat{\hat{K}}_2 \approx \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{i,j=1}^k \phi_{i,j}(0)\mathcal{U}^\dagger(n\Delta t)\hat{\hat{A}}_i(n\Delta t)\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(n\Delta t)\hat{\hat{A}}_j(n\Delta t)\mathcal{U}(n\Delta t)$$

$$+ \frac{2}{N^2}\sum_{n=1}^{N-1}\sum_{m=0}^{n-1}\sum_{i,j=1}^k \phi_{i,j}((n-m)\Delta t)\mathcal{U}^\dagger(n\Delta t)\hat{\hat{A}}_i(n\Delta t)\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(m\Delta t)\hat{\hat{A}}_j(m\Delta t)\mathcal{U}(m\Delta t)$$

$$- \frac{1}{N^2}\sum_{n=0}^{N-1}\mathcal{U}^\dagger(n\Delta t)\hat{\hat{D}}\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(n\Delta t)\hat{\hat{D}}\mathcal{U}(n\Delta t) - \frac{2}{N^2}\sum_{n=1}^{N-1}\sum_{m=0}^{n-1}\mathcal{U}^\dagger(n\Delta t)\hat{\hat{D}}\mathcal{U}(n\Delta t)\mathcal{U}^\dagger(m\Delta t)\hat{\hat{D}}\mathcal{U}(m\Delta t) - \hat{\hat{K}}_1^2. \qquad (A19)$$

To simulate the gates, we truncated $\hat{\hat{K}}(t)$ in Eq. (A11) at second order, which can be partially justified with the following. If we have no dissipator term in Eq. (A3), then, due to statistical independence, the mth-order cumulant disappears if, for a set of times $\{t_1, t_2, \ldots, t_n\}$, any of the time gaps $|t_1 - t_2|$, $|t_2 - t_3|, \ldots, |t_{n-1} - t_n|$ are larger than the correlation time τ_c of the stochastic process [40]. Since cumulants at every order vanish once the gap between the set of time points exceeds τ_c, then if $t \gg \tau_c$, the mth-order cumulant $\hat{\hat{K}}_m$ is effectively an integral over an $(m-1)$ dimensional sphere with radius τ_c,

integrated over t. Therefore, $\hat{\hat{K}}_m$ scales roughly as $\tau_c^{m-1}A^m t$, where A is the maximum norm of $\hat{\hat{A}}_i(t)$. Comparing the second- and fourth-order cumulants, $\hat{\hat{K}}_2$ and $\hat{\hat{K}}_4$, reveals that $\frac{\tau_c^3 A^4 t}{\tau_c A^2 t} = \tau_c^2 A^2$, meaning that if $\tau_c A \ll 1$ and $t \gg \tau_c$, we have a justification for truncating the cumulant expansion at second order. For the physical models considered in this work, the dissipator terms in the Liouville-von Neumann equation were considerably smaller in their norm than the noise Hamiltonian terms, so we assume that the arguments above are still applicable.

3. Circuit simulation

Each gate in the gate set acts on either one or two qubits, subjecting the rest to identical, uncorrelated noise (the noisy identity gate). The action of a noisy process on a quantum register can be calculated in the Kraus representation,

$$\Lambda(\rho) = \sum_{j=1}^{4^n} A_j \rho A_j^\dagger. \tag{A20}$$

For a generic noisy process, using naïve matrix multiplication, this calculation involves $\sim 2^{5n}$ operations and requires the storage of $\sim 2^{4n}$ complex parameters. These costs can be reduced dramatically by exploiting the fact that the noise is independent and that the gate set acts identically on different qubits. Noise maps that act independently commute and can be applied in sequence as a result,

$$\lambda^{\otimes n} = \lambda_1(\lambda_2(\dots \lambda_n(\rho))). \tag{A21}$$

Thus, the amount of storage is reduced to that required to store the gate set and the current state, $\sim 2^{2n}$ complex parameters, and the number of operations required now scales as $n2^{3n}$. This can be further reduced by noting that each channel λ_j is equivalent to the perfect identity on $n-1$ qubits, and its effect can be precalculated to reduce the total number of operations to $n2^{2n}$. The extension to two-qubit gates is straightforward; for further information, see Ref. [43].

APPENDIX B: PHYSICAL MODELS AND GATE PROTOCOL DETAILS

This Appendix describes the physical models, noise refocusing techniques, and the subsequent gate sets generated from these. It is assumed that the density matrix $\rho(t)$ describing a physical system evolves according to

$$\frac{\partial}{\partial t}\rho(t) = -i\hbar[H(t),\rho(t)]$$
$$+ \sum_i \left(L_i \rho(t) L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho(t)\} \right), \tag{B1}$$

where $H(t)$ is the Hamiltonian for the system and $\{L_i\}$ is a set of Lindblad operators generating nonunitary dynamics [44–46]. A physical model must specify all deterministic and stochastic Hamiltonians (both internal and control) and specify any Lindblad operators that the system is subject to.

1. Physical Model 1

Physical Model 1 is motivated by a double quantum dot physical system. A double quantum dot is a pair of electrons contained in a double potential well. The spatial and spin states of the electrons encode logical states $|0\rangle$ and $|1\rangle$,

$$|0\rangle = \left|\Phi_{11}^T\right\rangle \otimes (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/\sqrt{2}, \tag{B2}$$

$$|1\rangle = \left(a\left|\Phi_{11}^S\right\rangle + b\left|\Phi_{02}\right\rangle\right) \otimes (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}, \tag{B3}$$

where $|\Phi_{11}^S\rangle$ and $|\Phi_{11}^T\rangle$ are symmetric and antisymmetric spatial states with one electron in each of the potential wells and $|\Phi_{02}^S\rangle$ is a symmetric spatial state having two electrons in one particular well.

The electron state can be controlled by varying the voltage detuning $B(t)$ and Zeeman splitting difference $A(t)$, described below.

Voltage detuning introduces a potential energy difference B between the quantum wells. $B > 0$ favors the $|\Phi_{02}\rangle$ spatial state over $|\Phi_{11}^S\rangle$ and $|\Phi_{11}^T\rangle$, because $|\Phi_{02}\rangle$ allows for both electrons to minimize their potential energy. The parameters a and b in Eq. (B3) are therefore B dependent and given by Fermi-Dirac statistics [47] such that the probability $p_{11} = |a|^2$ of having an electron with potential energy B is given by $p_{11} = \frac{1}{1+e^{B/B_1-B_2}}$, whereby

$$a(B) = \sqrt{\frac{1}{1+e^{B/B_1-B_2}}}$$

$$b(B) = \sqrt{\frac{1}{1+e^{-(B/B_1-B_2)}}}.$$

The detuning Hamiltonian $H(B)$ is diagonal in the spatial states $\{|\Phi_{11}^S\rangle, |\Phi_{11}^T\rangle, |\Phi_{02}\rangle\}$ and takes a form

$$H(B) = B\left(\left|\Phi_{11}^S\right\rangle\left\langle\Phi_{11}^S\right| + \left|\Phi_{11}^T\right\rangle\left\langle\Phi_{11}^T\right|\right)$$
$$+ B_0 \left|\Phi_{02}\right\rangle\left\langle\Phi_{02}\right|, \tag{B4}$$

where B_0 is the energy eigenvalue of $|\Phi_{02}\rangle$ at zero detuning. Up to a constant identity contribution, this yields the following Hamiltonian for logical states:

$$H(B) = \begin{pmatrix} \langle 0| H(B) |0\rangle & \langle 0| H(B) |1\rangle \\ \langle 1| H(B) |0\rangle & \langle 1| H(B) |1\rangle \end{pmatrix}$$
$$= \frac{1}{2}\begin{pmatrix} |b|^2(B-B_0) & 0 \\ 0 & -|b|^2(B-B_0) \end{pmatrix}.$$

Zeeman splitting is related to the energy difference between electron spin-up and spin-down states in the presence of an external magnetic field. A magnetic field gradient across the potential wells introduces an energy splitting A between $|\Phi_{11}\rangle \otimes |\uparrow\downarrow\rangle$ and $|\Phi_{11}\rangle \otimes |\downarrow\uparrow\rangle$, where $|\Phi_{11}\rangle = |\Phi_{11}^S\rangle + |\Phi_{11}^T\rangle$, leading to a Hamiltonian $H(A) = \frac{A}{2}(|\Phi_{11}\rangle\langle\Phi_{11}| \otimes |\uparrow\downarrow\rangle\langle\uparrow\downarrow| - |\Phi_{11}\rangle\langle\Phi_{11}| \otimes |\downarrow\uparrow\rangle\langle\downarrow\uparrow|)$. The resulting matrix $H(A)$ in the logical basis is

$$H(A) = \begin{pmatrix} \langle 0| H(A) |0\rangle & \langle 0| H(A) |1\rangle \\ \langle 1| H(A) |0\rangle & \langle 1| H(A) |1\rangle \end{pmatrix}$$
$$= \frac{1}{2}\begin{pmatrix} 0 & aA \\ a^*A & 0 \end{pmatrix}.$$

Substituting the parameters $a(B)$ and $b(B)$ into the above expressions and summing them results in the effective logical single qubit Hamiltonian

$$H(t) = \frac{1}{2}\frac{A(t)+\alpha(t)}{\sqrt{1+\exp\left(\frac{B(t)}{B_1}-B_2\right)}}X$$
$$+ \frac{1}{2}\frac{B(t)-B_0+\beta(t)}{1+\exp\left[-\left(\frac{B(t)}{B_1}-B_2\right)\right]}Z, \tag{B5}$$

where the parameters $\alpha(t)$ and $\beta(t)$ encapsulate the stochastic behavior of the parameters $A(t)$ and $B(t)$. As the logical state $|0\rangle$ corresponds to the ground state of the physical Hamiltonian

at zero detuning ($B = 0$) and in zero magnetic field gradient ($A = 0$), we make an assumption that relaxation acts on the logical state simply as a Lindblad operator $L = \frac{1}{2\sqrt{T_1}}(X + iY)$.

The parameters $A(t)$, $B(t)$, B_i for $i \in \{0,1,2\}$, and T_1, define the deterministic evolution of the system, with the first two representing the single-qubit controls and the latter being the T_1 time constant of the system. Both control scalars are specified for intervals of length δt, over which the controls remain constant, whereas the rate of change of these control scalars between adjacent intervals is bounded with $|\frac{dA(t)}{dt}| \leqslant \Delta A_{\max}$ and $|\frac{dB(t)}{dt}| \leqslant \Delta B_{\max}$. Additionally, the controls are bounded by some maximum value; that is, $|A(t)| \leqslant A_{\max}$ and $0 \leqslant B(t) \leqslant B_{\max}$ for some $A_{\max}, B_{\max} \geqslant 0$.

The parameters specified by α and β are independent, zero-mean, stationary Gaussian processes [48] such that $\langle \alpha(t) \rangle = \langle \beta(t) \rangle = 0$ and $\langle \alpha(t_1)\beta(t_2) \rangle = 0$. The autocorrelation functions are given by

$$\langle \alpha(t_1)\alpha(t_2) \rangle = \Gamma_{\alpha_1}^2 \delta(|t_1 - t_2|) + \Gamma_{\alpha_2}^2 e^{-(\frac{|t_1-t_2|}{\tau_1})^2 + (\frac{|t_1-t_2|}{\tau_2})^4 - (\frac{|t_1-t_2|}{\tau_3})^6}$$

$$\langle \beta(t_1)\beta(t_2) \rangle = \Gamma_{\beta_1}^2 + \Gamma_{\beta_2}^2 \delta(|t_1 - t_2|), \tag{B6}$$

where $\delta(t)$ is the Dirac δ function. Parameters labeled with the letter Γ represent the noise strengths, and those labeled with τ represent various correlation times.

The Hamiltonian for simulating two-qubit gates is given by

$$H(t) = H^{(1)}(t) \otimes \mathbb{1} + \mathbb{1} \otimes H^{(2)}(t) + H_{zz}(t)$$

$$H_{zz}(t) = \frac{1}{4} \frac{C(t)(1 + \gamma(t))}{\left(1 + \exp\left[-\left(\frac{B^{(1)}(t)}{B_1} - B_2\right)\right]\right)}$$

$$\times \frac{Z \otimes Z - Z \otimes \mathbb{1} - \mathbb{1} \otimes Z}{\left(1 + \exp\left[-\left(\frac{B^{(2)}(t)}{B_1} - B_2\right)\right]\right)},$$

with two Lindblad operators $L_1 = \frac{1}{2\sqrt{T_1}}(X + iY) \otimes \mathbb{1}$ and $L_2 = \frac{1}{2\sqrt{T_1}}\mathbb{1} \otimes (X + iY)$. Any parameters or Hamiltonians denoted by superscript (i) mark either the first ($i = 1$) or the second ($i = 2$) qubit and are identical to the Hamiltonian in line (B5). The stochastic parts for single-qubit Hamiltonians on different qubits are taken to be independent. The two-qubit control parameter $C(t)$ can only take two values, $C(t) \in \{0, C_{\max}\}$, and the noise parameter $\gamma(t)$ is an independent zero-mean stationary Gaussian process with autocorrelation function

$$\langle \gamma(t_1)\gamma(t_2) \rangle = \Gamma_\gamma^2 \delta(|t_1 - t_2|). \tag{B7}$$

2. Physical Model 2

Physical Model 2 is an archetypal two-level system. For a single qubit, the Hamiltonian is given by

$$H(t) = \frac{1}{2}[B(t)(1 + \beta_1(t)) + \beta_2(t)]Z$$

$$+ \frac{1}{2}A(t)(1 + \alpha(t))[\cos(\phi(t))X + \sin(\phi(t))Y], \tag{B8}$$

with the only Lindblad operator given by $L = \frac{1}{2\sqrt{T_1}}(X + iY)$.

The parameters $A(t)$, $B(t)$, $\phi(t)$, and T_1 define the deterministic evolution of the system, with the first three representing the single-qubit controls. Every control value is specified for intervals of length δt, over which the controls remain constant. Each control scalar is bounded by some maximum value; that

is, $|A(t)| \leqslant A_{\max}$ and $|B(t)| \leqslant B_{\max}$ for some $A_{\max}, B_{\max} \geqslant 0$ but is not limited by any control rates.

The parameters specified using the letters α and β are all stationary Gaussian processes. All are zero-mean and independent. That is, $\langle \alpha(t) \rangle = \langle \beta_i(t) \rangle = 0$ for $i = 1, 2$, and $\langle \alpha(t_1)\beta_i(t_2) \rangle = \langle \beta_1(t_1)\beta_2(t_2) \rangle = 0$ for $i = 1, 2$. The autocorrelation functions are given as

$$\langle \alpha(t_1)\alpha(t_2) \rangle = \Gamma_\alpha^2 g_{1/f}\left(\Lambda_\alpha^{(l)}, \Lambda_\alpha^{(u)}, |t_1 - t_2|\right), \tag{B9}$$

$$\langle \beta_1(t_1)\beta_1(t_2) \rangle = \Gamma_{\beta_1}^2 g_{1/f}\left(\Lambda_{\beta_1}^{(l)}, \Lambda_{\beta_1}^{(u)}, |t_1 - t_2|\right), \tag{B10}$$

$$\langle \beta_2(t_1)\beta_2(t_2) \rangle = \Gamma_{\beta_2}^2 g_{1/f}\left(\Lambda_{\beta_2}^{(l)}, \Lambda_{\beta_2}^{(u)}, |t_1 - t_2|\right), \tag{B11}$$

where the parameters labeled with the letter Γ are the noise strengths and those labeled with Λ represent upper and lower cutoffs for $1/f$ noise. The autocorrelation function $g_{1/f}$ for $1/f$ noise is defined as the Fourier transform of $1/f$ spectral density with smooth cutoffs [49],

$$g_{1/f}(\Lambda_1, \Lambda_2, \Delta t)$$

$$= \int_{-\infty}^{\infty} \frac{2}{\pi\omega}\left(\arctan\left(\frac{\omega}{\Lambda_1}\right) - \arctan\left(\frac{\omega}{\Lambda_2}\right)\right) e^{-i\omega\Delta t} \, d\omega. \tag{B12}$$

Notice that $\lim_{\Lambda_1 \to 0, \Lambda_2 \to \infty} \frac{2}{\pi\omega}(\arctan(\frac{\omega}{\Lambda_1}) - \arctan(\frac{\omega}{\Lambda_2})) = \frac{1}{|\omega|}$.

The two-qubit Hamiltonian for this model is given by

$$H(t) = H^{(1)}(t) \otimes \mathbb{1} + \mathbb{1} \otimes H^{(2)}(t) + H_{zz}(t)$$

$$H_{zz}(t) = -\frac{1}{2}C(t)(1 + \gamma(t))Z \otimes Z, \tag{B13}$$

with two Lindblad operators $L_1 = \frac{1}{2\sqrt{T_1}}(X + iY) \otimes \mathbb{1}$ and $L_2 = \frac{1}{2\sqrt{T_1}}\mathbb{1} \otimes (X + iY)$.

Single-qubit Hamiltonians denoted by $H^{(i)}$ acting either on the first ($i = 1$) or the second ($i = 2$) qubit have identical parameters to the Hamiltonian in Eq. (B8), and stochastic parts for single-qubit Hamiltonians are taken to be independent. The two-qubit control parameter $C(t)$ is bounded in its maximum value $|C(t)| \leqslant C_{\max}$ but is otherwise unconstrained. $\gamma(t)$ is an independent zero-mean stationary Gaussian process, its autocorrelation function being given by

$$\langle \gamma(t_1)\gamma(t_2) \rangle = \Gamma_\gamma^2 g_{1/f}\left(\Lambda_\gamma^{(l)}, \Lambda_\gamma^{(u)}, |t_1 - t_2|\right). \tag{B14}$$

3. XY sequence gate protocol

Suppose we have a dynamical decoupling sequence which is given as a list of unitary operations $\{A_i\}$, $i = 1, \ldots, N$, where i denotes the temporal order of these operations. We demand that

$$A_N \ldots A_2 A_1 = e^{i\phi}\mathbb{1}, \tag{B15}$$

where $e^{i\phi}$ is an arbitrary global phase. If we want to spread a unitary gate U across the sequence $\{A_i\}$, we first

find

$$U_1 = A_1 U^{1/N} A_1^{-1}, \tag{B16}$$

$$U_2 = A_2 A_1 U^{1/N} A_1^{-1} A_2^{-1}, \tag{B17}$$

$$\vdots \tag{B18}$$

$$U_N = A_N ... A_2 A_1 U^{1/N} A_1^{-1} A_2^{-1} ... A_N^{-1}, \tag{B19}$$

where $(U^{1/N})^N = U$, and then implement the sequence A_1, U_1, A_2, U_2, ..., A_N, U_N resulting in $U_N A_N ... U_2 A_2 U_1 A_1 = e^{i\phi} U$, which follows from direct substitution and Eq. (B15).

a. XY8 sequence

The XY8 sequence [30] is an eight-unitary decoupling sequence where, following the notation above, $A_1 = A_3 = A_6 = A_8 = X$ and $A_2 = A_4 = A_5 = A_7 = Y$. In the limit of perfect control and infinitesimally short (delta) pulses, the sequence refocuses noise along any direction that varies slower than the sequence is implemented. To implement a unitary gate U within the sequence, we split it into eight parts,

$$U_1 = U_7 = X U^{\frac{1}{8}} X, \quad U_2 = U_6 = Y X U^{\frac{1}{8}} X Y,$$
$$U_3 = U_5 = X Y X U^{\frac{1}{8}} X Y X, \quad U_4 = U_8 = U^{\frac{1}{8}}, \tag{B20}$$

where we simplify the expression using $XYXY = YXYX = -\mathbb{1}$, and the fact that the global phase of the desired unitary is irrelevant.

b. XY4 sequence

The XY4 sequence [31] is a four-unitary decoupling sequence where, following the notation above, $A_1 = A_3 = X$ and $A_2 = A_4 = Y$. Like the XY8 sequence, the XY4 sequence refocuses noise along any direction that varies slower than the sequence is implemented, given that the pulses are ideal and infinitesimally short. We spread a unitary gate U across the sequence by breaking it into four parts as follows:

$$U_1 = X U^{\frac{1}{4}} X, \quad U_2 = Y X U^{\frac{1}{4}} X Y,$$
$$U_3 = X Y X U^{\frac{1}{4}} X Y X, \quad U_4 = U^{\frac{1}{4}}. \tag{B21}$$

4. Gate sets

Gate Set 1 is built on Physical Model 1 using the parameters in Table III. This gate set was built from an XY8 pulse sequence [30], with single-qubit gates being implemented within this sequence according to the XY sequence gate protocol. Each pulse piece in the sequence was found via the GRAPE algorithm [32,33] with control constraints from Table III incorporated into the algorithm. All gates are 199.2 ns long and the discretization step for cumulant simulations was chosen to be 0.1 ns.

Gate Set 2 (GS2) is built on Physical Model 2 using the parameters in Table IV. This gate set was built from an XY4 pulse sequence [31], again, with single-qubit gates being implemented within this sequence according to the XY sequence gate protocol. All pulse pieces are performed using hard pulses. All gates are 168 ns long and the discretization step

TABLE III. Parameters used for Physical Model 1, Gate Set 1.

Control parameter	Value	Noise parameter	Value
B_0	1.5193×10^{13} Hz	T_1	1 s
B_1	1.5193×10^{11} Hz	Γ_{α_1}	4.804 Hz
B_2	120	Γ_{α_2}	1.519×10^8 Hz
A_{max}	3.798×10^8 Hz	τ_1	10^{-2} s
B_{max}	3.0385×10^{13} Hz	τ_2	10^{-3} s
ΔA_{max}	0.7596×10^{18} Hz/s	τ_3	10^{-4} s
ΔB_{max}	1.215×10^{23} Hz/s	Γ_{β_1}	1.519×10^9 Hz
C_{max}	8.73568×10^{12} Hz	Γ_{β_2}	4.804×10^6 Hz
δt	10^{-10} s	Γ_γ	10^3 Hz

for cumulant simulations was chosen to be 0.25 ns. (As GS2 uses hard pulses, the cumulant simulation can be discretized more coarsely, as the pulse amplitudes and phases remain constant for longer periods of time. The same can be said for GS3, though, given that the gates are so short, a smaller time step was used anyway).

Gate set 3 (GS3) is also built on Physical Model 2 but uses the noise parameters in Table V to provide variety in the resultant gate errors. No refocusing pulse sequences are used; all gates are generated from simple hard pulses. All gates are 25 ns long and the discretization step for cumulant simulations was chosen to be 0.1 ns.

For each gate set, the two-qubit CNOT gate is implemented using the identity

$$(-1)^{\frac{3}{4}} U_{CNOT} = e^{i\frac{\pi}{2}\mathbb{1} \otimes \frac{X}{2}} e^{-i\frac{\pi}{2}\mathbb{1} \otimes \frac{Y}{2}} e^{-i\pi \frac{Z \otimes Z}{4}} e^{i\frac{\pi}{2}\mathbb{1} \otimes \frac{Y}{2}} e^{i\frac{\pi}{2}\frac{Z}{2} \otimes \mathbb{1}}. \tag{B22}$$

For GS1 and GS2, as for single-qubit gates, the gate is broken into parts that are interspersed into their respective XY sequences. In this case, however, the first two single-qubit rotations are done during the first half of the XY sequence, and the last two single-qubit rotations are done during the second half of the XY sequence, in a way similar to the single-qubit

TABLE IV. Parameters used for Physical Model 2, Gate Set 2.

Control parameter	Value	Noise parameter	Value
A_{max}	$2\pi \times 10^8$ Hz	T_1	10^{-4} s
B_{max}	$2\pi \times 10^9$ Hz	Γ_α	3×10^4 Hz
C_{max}	$2\pi \times 10^8$ Hz	Γ_{β_1}	3×10^4 Hz
δt	10^{-9} s	Γ_{β_2}	$10^6/2\pi$ Hz
		$\Lambda_\alpha^{(l)}$	$1/2\pi$ Hz
		$\Lambda_\alpha^{(u)}$	10^9 Hz
		$\Lambda_{\beta_1}^{(l)}$	$1/2\pi$ Hz
		$\Lambda_{\beta_1}^{(u)}$	10^9 Hz
		$\Lambda_{\beta_2}^{(l)}$	$1/2\pi$ Hz
		$\Lambda_{\beta_2}^{(u)}$	10^9 Hz
		Γ_γ	$1.2 \times 10^3/2\pi$ Hz
		$\Lambda_\gamma^{(l)}$	$1/2\pi$ Hz
		$\Lambda_\gamma^{(u)}$	10^9 Hz

TABLE V. Parameters used for Physical Model 2, Gate Set 3.

Control parameter	Value	Noise parameter	Value
$A_{\max}$	$2\pi \times 10^8$ Hz	T_1	10^{-5} s
$B_{\max}$	$2\pi \times 10^9$ Hz	Γ_α	0 Hz
$C_{\max}$	$2\pi \times 10^8$ Hz	Γ_{β_1}	10^4 Hz
δt	10^{-9} s	Γ_{β_2}	10^4 Hz
		$\Lambda_\alpha^{(l)}$	$1/2\pi$ Hz
		$\Lambda_\alpha^{(u)}$	10^9 Hz
		$\Lambda_{\beta_1}^{(l)}$	$1/2\pi$ Hz
		$\Lambda_{\beta_1}^{(u)}$	10^9 Hz
		$\Lambda_{\beta_2}^{(l)}$	$1/2\pi$ Hz
		$\Lambda_{\beta_2}^{(u)}$	10^9 Hz
		Γ_γ	$1.2 \times 10^3/2\pi$ Hz
		$\Lambda_\gamma^{(l)}$	$1/2\pi$ Hz
		$\Lambda_\gamma^{(u)}$	10^9 Hz

gates. The two-qubit coupling operation is implemented in the middle of the XY sequence. For GS3, the CNOT gate is implemented according to the above identity, using hard pulses.

See Appendix A 2 for details on the procedure used to simulate the gates.

APPENDIX C: SECONDARY ANALYSIS

This Appendix contains analysis that is secondary to the main point of the paper but that we consider important in its own right. We identify the good and poor regimes of performance for the "Pauli-twirled" errors and classify our gate sets into these regimes. This analysis also aids in the understanding of the behavior of the approximations generated by our own method; in particular, the observation made in the main body that in some cases, the average fidelity of our approximations is much less than that of the original, whereas in some cases they are very similar. Given the importance on the classification of errors to this discussion, we conclude by explaining why the errors for each gate set take the form that they do.

The identification of these regimes requires analysis on what happens to the IO distinguishability properties of a channel when it is twirled. Generally, twirling a map Λ by a set of unitaries $\{U_k\}_{k=1}^N$ is the action of mapping $\Lambda \rightarrow \frac{1}{N}\sum_{k=1}^N U_k^\dagger \circ \Lambda \circ U_k$. The twirled map results from choosing a unitary operator from the twirling set with uniform probability, applying it, applying Λ, and then inverting the twirling operator. If the twirling set is chosen to be the Pauli operators, it is called a Pauli twirl and, if perfectly implemented, will transform any map into the Pauli channel that results from mathematically truncating the off-diagonal elements of the process (χ) matrix [8].

Before analyzing the effects of Pauli twirling specifically, we can look at the general effect of twirling on the diamond norm distance of an arbitrary channel to the identity operation. Let $\mathcal{H}$ denote a finite-dimensional Hilbert space,

and $L(\mathcal{H})$ the set of linear operators from $\mathcal{H} \rightarrow \mathcal{H}$. One property of the diamond norm is that for $\Phi : L(\mathcal{H}) \rightarrow L(\mathcal{H})$, and any unitary operators $U, V \in L(\mathcal{H})$, it holds that $\|U \circ \Phi \circ V\|_\diamond = \|\Phi\|_\diamond$ [21]. Thus, for any finite set of unitaries $\{U_k\}_{k=1}^N \subset L(\mathcal{H})$, and quantum channel $\Lambda : L(\mathcal{H}) \rightarrow L(\mathcal{H})$, it holds by straightforward application of the triangle inequality that

$$\left\| \mathbb{1}_{L(\mathcal{H})} - \frac{1}{N}\sum_{k=1}^N U_k^\dagger \circ \Lambda \circ U_k \right\|_\diamond \leqslant \left\| \mathbb{1}_{L(\mathcal{H})} - \Lambda \right\|_\diamond, \quad (C1)$$

where $\mathbb{1}_{L(\mathcal{H})}$ is the identity channel. In other words, the distance of a twirled error to the identity operation is always bounded above by that of the original error and so, in a worst-case sense, twirling typically acts to make an error harder to detect.

To specify the regimes of performance of twirling, we look at the Bloch representation of quantum channels. Any qubit map Λ can be represented as a matrix M and vector $\vec{t}$ that acts on Bloch vectors as $\vec{r} \rightarrow M\vec{r} + \vec{t}$. M can be written in terms of its polar decomposition $M = OP$, where O and P are an orthogonal and positive semidefinite matrix, respectively. Thus, the action on the Bloch sphere can be represented as $\vec{r} \rightarrow O(P\vec{r} + O^T\vec{t})$ [50]. That is, as a possibly nonunital channel followed by an orthogonal rotation of the Bloch sphere, which corresponds to a unitary rotation for qubits [51]. We say that a channel is in the "unitary regime" if the effect of O is relatively large compared to P and $\vec{t}$, and we say the channel is in the "deforming regime" if the opposite is true. To quantify the "effect" of a Bloch matrix M, we use the quantity $\|\mathbb{1} - M\|_2$, and use $\|\vec{t}\|_2$ to quantify the effect of the nonunital part. We use $\|\cdot\|_2$ to denote both the Hilbert-Schmidt norm on matrices and the Euclidean norm on vectors, where context and notation will make clear which is meant. In the following paragraphs, we examine how twirling affects worst-case IO distinguishability properties of different errors in these regimes.

First, for a unital deforming qubit channel Λ, the Bloch representation is simply a positive semidefinite matrix P. Diagonalize P as $P = UDU^\dagger$ for some orthogonal matrix U and diagonal non-negative matrix D. Then, as diagonal Bloch matrices are realizable as Pauli channels, we can write Λ as $\Lambda(\rho) = \sum_{i=0}^3 p_i V P_i V^\dagger \rho V P_i V^\dagger$, where V is the unitary corresponding to U. Channels of this form are called generalized Pauli channels, and the diamond norm distance between channels of this form with different probability vectors $\vec{p}$ and $\vec{q}$ is given by $\sum_{i=0}^3 |p_i - q_i|$ [52,53]. For these channels, as $V P_i V^\dagger$ has no identity part for $i \geqslant 1$, the χ_{00} element in the Pauli basis is identical to the probability assigned to the identity operator. As this quantity is conserved in Pauli twirling, and as the identity map on qubits $\mathbb{1}_{L(\mathbb{C}^2)}$ is a generalized Pauli channel, it follows that, for a deforming unital qubit channel Λ and its Pauli twirl Λ_{PT}, $\|\mathbb{1}_{L(\mathbb{C}^2)} - \Lambda\|_\diamond = (1 - p_0) + p_1 + p_2 + p_3 = 2(1 - p_0) = \|\mathbb{1}_{L(\mathbb{C}^2)} - \Lambda_{PT}\|_\diamond$. Thus, diamond norm distance to the identity for these channels is unaffected by Pauli twirling.

To examine the qualitative behavior of Pauli twirling on a nonunital deforming channel, we examine the special case of

TABLE VI. We denote the Bloch representation of each gate error as $(OP, \vec{t})$, where O is an orthogonal matrix, P is positive semidefinite, and $\vec{t}$ is the "nonunital part." Norms of size less than 10^{-10} are displayed as 0, as at this size they are irrelevant compared to the dominant parts of the error.

	Statistics	$\mathbb{1}$	X	Y	Z	H	CNOT
	$\\|\mathbb{1} - O\\|_2$	6.74×10^{-3}	1.83×10^{-2}	1.83×10^{-2}	5.06×10^{-3}	1.63×10^{-2}	5.78×10^{-2}
GS1	$\\|\mathbb{1} - P\\|_2$	8.22×10^{-7}	1.06×10^{-2}	4.43×10^{-3}	1.33×10^{-2}	6.24×10^{-3}	3.56×10^{-2}
	$\\|\vec{t}\\|_2$	0.0	5.51×10^{-9}	5.56×10^{-9}	0.0	1.36×10^{-8}	2.19×10^{-8}
	$\\|\mathbb{1} - O\\|_2$	3.35×10^{-5}	5.52×10^{-4}	5.52×10^{-4}	2.09×10^{-4}	1.59×10^{-3}	1.73×10^{-3}
GS2	$\\|\mathbb{1} - P\\|_2$	2.16×10^{-3}	2.15×10^{-3}	2.15×10^{-3}	2.32×10^{-3}	2.19×10^{-3}	8.13×10^{-3}
	$\\|\vec{t}\\|_2$	4.01×10^{-8}	5.65×10^{-5}	4.95×10^{-5}	1.66×10^{-4}	9.91×10^{-5}	2.14×10^{-4}
	$\\|\mathbb{1} - O\\|_2$	0.0	1.09×10^{-5}	1.50×10^{-5}	1.04×10^{-5}	2.83×10^{-6}	4.45×10^{-2}
GS3	$\\|\mathbb{1} - P\\|_2$	3.06×10^{-3}	2.93×10^{-3}	2.93×10^{-3}	3.06×10^{-3}	2.96×10^{-3}	1.12×10^{-2}
	$\\|\vec{t}\\|_2$	2.50×10^{-3}	1.37×10^{-3}	1.37×10^{-3}	2.50×10^{-3}	2.11×10^{-3}	2.95×10^{-3}

an amplitude damping channel, which has Kraus operators

$$K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \quad \text{(C2)}$$

for some parameter γ and Bloch representation

$$M = \begin{pmatrix} \sqrt{1-\gamma} & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 \\ 0 & 0 & 1-\gamma \end{pmatrix}, \quad \vec{t} = \begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix}. \quad \text{(C3)}$$

In the Bloch representation, the effect of Pauli twirling is to remove the off-diagonal elements of M and set $\vec{t} \to 0$. Thus, the only effect that twirling has on this channel is to remove the nonunital part. Note that the dominant error considered by Geller and Zhou in Ref. [26] is of this form; the Bloch matrix is diagonal, and so, in some sense, Pauli twirling has a minimal effect. We consider the worst-case performance of this channel on a pure qubit state. It is clear that the state most affected by Λ, and by its Pauli twirl Λ_{PT}, is the -1 eigenstate of Z, which we denote as ρ_-. It can be easily checked that $\\|\rho_- - \Lambda(\rho_-)\\|_1 = 2\gamma$, and $\\|\rho_- - \Lambda_{\mathrm{PT}}(\rho_-)\\|_1 = \gamma$. Thus, while the worst-case performance in this case is lessened by Pauli twirling, it is only by a factor of 2.

Last, we look at purely unitary channels. For two unitary operators U, V in $L(\mathcal{H})$, there exists a pure state $|\psi\rangle \in \mathcal{H}$ for which

$$\\|U \cdot U^\dagger - V \cdot V^\dagger\\|_\diamond = \\|U |\psi\rangle \langle\psi| U^\dagger - V |\psi\rangle \langle\psi| V^\dagger\\|_1$$
$$= 2\sqrt{1 - |\langle\psi| U^\dagger V |\psi\rangle|^2} \quad \text{(C4)}$$

(see Ref. [21]). From this form, it is clear that any state that maximizes the distinguishability between the identity operation and a unitary U will be orthogonal to the unitary rotation axis. Thus, for a rotation U by an angle θ, $\\|\mathbb{1}_{L(\mathbb{C}^2)} - U \cdot U^\dagger\\|_\diamond = 2|\sin(\theta/2)|$. As the χ_{00} element of a qubit rotation by angle θ is $\cos^2(\theta/2)$, it follows from the preceding discussion that, for the Pauli-twirled error Λ_{PT}, $\\|\mathbb{1}_{L(\mathbb{C}^2)} - \Lambda_{\mathrm{PT}}\\|_\diamond = 2(1 - \cos^2(\theta/2)) = 2\sin^2(\theta/2)$. Thus, for small values of θ, the twirled channel can be orders of magnitude closer to the identity than the original.

With this qualitative analysis in hand, we examine the form of the Bloch representation of the errors considered in this paper. The Bloch representation of each error is decomposed

into the three pieces, O, P, and $\vec{t}$, and the size of each piece is reported in Table VI. As the identity gate occurs most frequently in the circuit, its properties are the most important.

We see that GS1 has the largest unitary component to its errors (in terms of the ratio to the other components). The error in the identity gate is almost entirely unitary and, therefore, falls neatly into the unitary regime. Indeed, this is consistent with the fact that a single-qubit unitary rotation by an angle θ will have a χ_{00} element of $\cos^2(\theta/2) \approx 0.99999$, whereas the Pauli channel that *exactly* matches its IO distinguishably has a χ_{00} element of $1 - |\sin(\theta/2)| \approx 0.99762$. We see that these numbers correspond exactly to the given number of digits in Table I, where a two-order-of-magnitude decrease in the distance of the Pauli-twirled error to the identity channel is shown. While the other errors in the set have unitary parts, the nonunitary parts are comparable in size. As such, character from both the unitary and unital deforming regimes is observed; the diamond norm distance of the Pauli-twirled error to the identity decreases, but the decrease is not as impressive as for the identity gate. For the Y and H errors, the unitary part is 2 or 3 times larger than the nonunitary, and thus the decrease in diamond norm distance to the ideal channel for the Pauli-twirled errors is the greatest of the nonidentity gates.

For every gate in GS2, $\\|\vec{t}\\|_2$ is relatively small and $\\|\mathbb{1} - O\\|_2$ is in most cases an order of magnitude smaller than $\\|\mathbb{1} - P\\|_2$, putting GS2 into the unital deforming regime. When looking at the diamond norm distances of the Pauli-twirled channels to the identity, we see that, as expected in this regime, there is very little decrease. Indeed, while the Pauli-twirled channels generally underestimate the error in the hedging metrics, the underestimation is small.

For GS3, $\\|\mathbb{1} - O\\|_2$ is usually of comparatively negligible size, and $\\|\vec{t}\\|_2$ is of the same order of magnitude as $\\|\mathbb{1} - P\\|_2$, putting this gate set into the nonunital deforming regime. Looking at the decrease of the diamond norm distance to the ideal operations for the Pauli-twirled errors, we see roughly what is expected; a decrease on the order of a factor of 2. As in the case of GS2, while the Pauli-twirled errors show underestimation in the hedging statistics, it is by a small degree.

This discussion supports and illuminates the idea of two regimes of behavior for Pauli-twirled approximations. In the unitary case, we observe large underestimation of IO distinguishability properties. For the circuit we have considered, this underestimation propagates upwards; the gadget simulated with Pauli-twirled errors is much harder to distinguish from the perfect gadget than the original errors. In the deformation regime, the Pauli-twirled errors can result in underestimation. One might argue, however, as Geller and Zhou do, that this level of underestimation is acceptable. Indeed, while the hedging is negative for the Pauli-twirled gates and circuit for GS2, it is small in magnitude. For GS3, the Pauli-twirled circuit even has positive hedging. This is perhaps not surprising on an intuitive level. Pauli channels are a subset of deforming channels, and while not all deforming errors are Pauli channels, they are *like* Pauli channels. All deforming channels contract the Bloch sphere inwards (with possible nonunital shifts), so Pauli twirling simply maps one deforming channel to another, which does not dramatically change the IO distinguishability properties. This stands in contrast to the unitary regime where, in regards to IO distinguishability, something fundamental can be lost when a dominantly unitary error is mapped to a Pauli channel via Pauli twirling.

This discussion also demonstrates why our approximations behave the way they do in regards to average gate fidelity. As we have seen from the Pauli twirl analysis, for a fixed average fidelity, channels in the unitary regime are much more distinguishable from the identity operation than those in the deforming regime. Thus, if the original error is in the unitary regime, our approximations, which fall into the deforming regime, will sometimes have a much worse average fidelity than the original to allow them to honestly represent its IO distinguishability properties.

As a final point, we can ask why the errors for each gate set have the form that they do. To explain this, it is necessary to examine the physical models and the pulse design techniques used, which are described in detail in Appendix B.

The simplest case is GS3, where hard pulses with no refocusing sequences were used. As the T_1 time was the shortest of all of the gate sets, and no refocusing pulses were used, the strong nonunital effects were allowed to continue building in their natural direction throughout each gate resulting in primarily nonunital errors. In GS2, which used the same physical model as GS3, the T_1 time was an order of magnitude longer. In addition, the refocusing pulse sequence used has the effect of "flipping" the nonunital shift between the positive and negative z direction, and so, when averaged over the whole sequence, the effective nonunital shift tends to zero. As a result, the size of the nonunital piece tends to be at least an order of magnitude smaller than the other pieces, putting this gate set into the unital deformation regime. Finally, as GS1 had the longest T_1 time, and the refocusing sequence was performed at a rate far faster than the T_1 time, the nonunital error components are negligible. Aside from this, the prime difference between GS1 and the other gate sets is that the pulses used were found using the GRAPE algorithm, rather than being hard, which accounts for the dominant unitary components of the errors. For hard pulses, in the limit of no noise, the implemented unitary should be perfect, aside from numerical imprecision in the field amplitudes, which can be made arbitrarily small with low overhead. For pulses found using the GRAPE algorithm this is not the case; even a noiseless implementation will not produce the perfect unitary. Arguably, this situation is more prevalent in experimental implementations as pulse-finding algorithms like GRAPE are necessary to obtain high-fidelity control over systems with complicated Hamiltonians. In addition, these types of algorithms can be used to achieve high-fidelity control across a range of internal Hamiltonian parameters, even when the Hamiltonian is simple [33].

APPENDIX D: GATE STATISTICS

Tables VII–IX contain statistics on all gates in Gate Sets 1-3.

TABLE VII. Statistics for the various approximations of the gates in GS1, approximated from $N = 10^6$ random pure states.

Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
𝟙	χ_{00}	0.999994	0.999994	0.997618	0.998314
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.76×10^{-3}	1.20×10^{-5}	4.76×10^{-3}	4.77×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.76×10^{-3}	6.73×10^{-3}	3.64×10^{-3}
	$\bar{h}$		-3.73×10^{-3}	1.14×10^{-7}	1.64×10^{-6}
	p_{viol}		1.0	0.0	0.0
X	χ_{00}	0.996147	0.996147	0.991234	0.994179
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.51×10^{-2}	7.71×10^{-3}	1.75×10^{-2}	1.59×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.29×10^{-2}	1.67×10^{-2}	4.58×10^{-3}
	$\bar{h}$		-5.76×10^{-3}	1.12×10^{-3}	3.90×10^{-4}
	p_{viol}		0.99892	0.0	0.0
Y	χ_{00}	0.998330	0.998330	0.992021	0.994711
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.35×10^{-2}	3.34×10^{-3}	1.60×10^{-2}	1.43×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.30×10^{-2}	1.86×10^{-2}	8.15×10^{-3}
	$\bar{h}$		-7.96×10^{-3}	9.94×10^{-4}	3.09×10^{-4}
	p_{viol}		0.99953	0.0	0.0

TABLE VII. (*Continued*).

Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
	χ_{00}	0.995289	0.995289	0.992055	0.993663
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.06×10^{-2}	9.42×10^{-3}	1.59×10^{-2}	1.31×10^{-2}
Z	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.59×10^{-3}	7.68×10^{-3}	3.46×10^{-3}
	$\overline{h}$		-5.71×10^{-4}	3.01×10^{-3}	1.29×10^{-3}
	p_{viol}		0.60369	0.0	0.0
	χ_{00}	0.997642	0.997642	0.991031	0.993638
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.35×10^{-2}	4.72×10^{-3}	1.79×10^{-2}	1.54×10^{-2}
H	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.27×10^{-2}	1.85×10^{-2}	8.83×10^{-3}
	$\overline{h}$		-6.42×10^{-3}	2.51×10^{-3}	1.10×10^{-3}
	p_{viol}		1.0	0.0	0.0
	χ_{00}	0.993152	0.993152	0.976600	0.976600
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.09×10^{-2}	1.37×10^{-2}	4.68×10^{-2}	4.68×10^{-2}
CNOT	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.82×10^{-2}	4.52×10^{-2}	4.52×10^{-2}
	$\overline{h}$		-1.17×10^{-2}	1.52×10^{-2}	1.52×10^{-2}
	p_{viol}		1.0	0.0	0.0

TABLE VIII. Statistics for the various approximations of the gates in GS2, approximated from $N = 10^6$ random pure states.

Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
	χ_{00}	0.999087	0.999087	0.999085	0.999086
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}	1.83×10^{-3}
$\mathbb{1}$	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.48×10^{-5}	2.50×10^{-5}	2.06×10^{-6}
	$\overline{h}$		-1.63×10^{-7}	2.65×10^{-6}	8.34×10^{-7}
	p_{viol}		0.49861	0.0	0.0
	χ_{00}	0.999074	0.999074	0.998972	0.999029
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.91×10^{-3}	1.85×10^{-3}	2.06×10^{-3}	2.00×10^{-3}
X	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.38×10^{-4}	4.85×10^{-4}	1.16×10^{-4}
	$\overline{h}$		-4.04×10^{-5}	9.61×10^{-5}	6.12×10^{-5}
	p_{viol}		0.7897	0.0	0.0
	χ_{00}	0.999075	0.999075	0.998990	0.999033
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.91×10^{-3}	1.85×10^{-3}	2.02×10^{-3}	1.99×10^{-3}
Y	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		4.32×10^{-4}	4.72×10^{-4}	1.03×10^{-4}
	$\overline{h}$		-4.02×10^{-5}	7.32×10^{-5}	5.58×10^{-5}
	p_{viol}		0.8674	0.0	0.0
	χ_{00}	0.999023	0.999023	0.998869	0.998901
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.97×10^{-3}	1.95×10^{-3}	2.26×10^{-3}	2.21×10^{-3}
Z	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.67×10^{-4}	4.27×10^{-4}	3.33×10^{-4}
	$\overline{h}$		-1.37×10^{-5}	1.87×10^{-4}	1.52×10^{-4}
	p_{viol}		0.61392	0.0	0.0
	χ_{00}	0.999060	0.999060	0.998684	0.998878
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	2.33×10^{-3}	1.88×10^{-3}	2.63×10^{-3}	2.53×10^{-3}
H	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.25×10^{-3}	1.45×10^{-3}	4.35×10^{-4}
	$\overline{h}$		-3.05×10^{-4}	1.98×10^{-4}	1.32×10^{-4}
	p_{viol}		1.0	0.0	0.0
	χ_{00}	0.998071	0.998071	0.997683	0.997683
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.99×10^{-3}	3.86×10^{-3}	4.63×10^{-3}	4.63×10^{-3}
CNOT	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.01×10^{-3}	1.33×10^{-3}	1.33×10^{-3}
	$\overline{h}$		-7.38×10^{-5}	5.44×10^{-4}	5.44×10^{-4}
	p_{viol}		0.76905	0.0	0.0

TABLE IX. Statistics for the various approximations of the gates in GS3, approximated from $N = 10^6$ random pure states.

Gate	Statistics	Original	Pauli twirled	Pauli	Clifford
$\mathbb{1}$	χ_{00}	0.998751	0.998751	0.996501	0.996501
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.99×10^{-3}	2.50×10^{-3}	7.00×10^{-3}	7.00×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.50×10^{-3}	5.28×10^{-3}	5.28×10^{-3}
	$\overline{h}$		-1.03×10^{-3}	1.91×10^{-3}	1.91×10^{-3}
	p_{viol}		0.74978	0.0	0.0
X	χ_{00}	0.998751	0.998751	0.997447	0.997674
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.17×10^{-3}	2.50×10^{-3}	5.11×10^{-3}	4.65×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.37×10^{-3}	3.12×10^{-3}	2.75×10^{-3}
	$\overline{h}$		-3.64×10^{-4}	1.39×10^{-3}	1.06×10^{-3}
	p_{viol}		0.69216	0.0	0.0
Y	χ_{00}	0.998751	0.998751	0.997669	0.997669
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	3.17×10^{-3}	2.50×10^{-3}	4.66×10^{-3}	4.66×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.38×10^{-3}	2.76×10^{-3}	2.76×10^{-3}
	$\overline{h}$		-3.63×10^{-4}	1.06×10^{-3}	1.06×10^{-3}
	p_{viol}		0.69113	0.0	0.0
Z	χ_{00}	0.998751	0.998751	0.996501	0.996501
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.99×10^{-3}	2.50×10^{-3}	7.00×10^{-3}	7.00×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.50×10^{-3}	5.28×10^{-3}	5.28×10^{-3}
	$\overline{h}$		-1.03×10^{-3}	1.92×10^{-3}	1.91×10^{-3}
	p_{viol}		0.75055	0.0	0.0
H	χ_{00}	0.998751	0.998751	0.996883	0.996940
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	4.31×10^{-3}	2.50×10^{-3}	6.23×10^{-3}	6.12×10^{-3}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		2.27×10^{-3}	4.43×10^{-3}	4.36×10^{-3}
	$\overline{h}$		-7.92×10^{-4}	1.68×10^{-3}	1.60×10^{-3}
	p_{viol}		0.74018	0.0	0.0
CNOT	χ_{00}	0.997441	0.997441	0.981266	0.981266
	$\|\Lambda - U_{\text{Ideal}}\|_\diamond$	1.98×10^{-2}	5.12×10^{-3}	3.75×10^{-2}	3.75×10^{-2}
	$\|\Lambda - \Lambda_{\text{Original}}\|_\diamond$		1.83×10^{-2}	3.90×10^{-2}	3.90×10^{-2}
	$\overline{h}$		-1.15×10^{-2}	1.48×10^{-2}	1.48×10^{-2}
	p_{viol}		0.99988	0.0	0.0

[1] Peter W. Shor, SIAM J. Comput. **26**, 1484 (1997).

[2] Lov K. Grover, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, *STOC '96, New York, NY, USA, 1996* (ACM, New York, 1996), pp. 212–219.

[3] Nathan Wiebe, Daniel Braun, and Seth Lloyd, Phys. Rev. Lett. **109**, 050505 (2012).

[4] N. C. Jones, R. Van Meter, A. G. Fowler, P. L. McMahon, J. Kim, T. D. Ladd, and Y. Yamamoto, Phys. Rev. X **2**, 031007 (2012).

[5] M. H. Devoret and R. J. Schoelkopf, Science **339**, 1169 (2013).

[6] Austin G. Fowler, Ashley M. Stephens, and Peter Groszkowski, arXiv:0803.0272 (2008); Phys. Rev. A **80**, 052312 (2009).

[7] Yaakov S. Weinstein, Timothy F. Havel, Joseph Emerson, Nicolas Boulant, Marcos Saraceno, Seth Lloyd, and David G. Cory, J. Chem. Phys. **121**, 6117 (2004).

[8] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).

[9] Christopher E. Granade, Christopher Ferrie, Nathan Wiebe, and D. G. Cory, New J. Phys. **14**, 103013 (2012).

[10] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and D. G. Cory, arXiv:1309.0876 (2013).

[11] Scott Aaronson and Daniel Gottesman, Phys. Rev. A **70**, 052328 (2004).

[12] Victor Veitch, Christopher Ferrie, David Gross, and Joseph Emerson, New J. Phys. **14**, 113011 (2012).

[13] Leslie G. Valiant, SIAM J. Comput. **31**, 1229 (2002).

[14] M. Van den Nest, arXiv:1201.4867 (2012).

[15] Juan Bermejo-Vega and Maarten Van den Nest, arXiv:1210.3637 (2012).

[16] Richard Jozsa and Maarten Van den Nest, arXiv:1305.6190 (2013).

[17] Sergey Bravyi, Guillaume Duclos-Cianci, David Poulin, and Martin Suchara, arXiv:1207.1443 (2012).

[18] D. S. Wang, A. G. Fowler, A. M. Stephens, and L. C. L. Hollenberg, Quantum Inf. Comput. **10**, 456 (2010).

[19] Sergey Bravyi and Alexander Vargo, arXiv:1308.6270 (2013).

[20] Easwar Magesan, Daniel Puzzuoli, Christopher E. Granade, and David G. Cory, Phys. Rev. A **87**, 012324 (2013).

[21] John Watrous, Theory of Quantum Information, 2011 (unpublished course notes).

[22] C. A. Fuchs and J. van de Graaf, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[23] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Providence, RI, 2002).

[24] Mauricio Gutiérrez, Lukas Svec, Alexander Vargo, and Kenneth R. Brown, Phys. Rev. A **87**, 030302(R) (2013).

[25] Yu Tomita, Mauricio Gutiérrez, Chingiz Kabytayev, Kenneth R. Brown, M. R. Hutsel, A. P. Morris, Kelly E. Stevens, and G. Mohler, arXiv:1305.0349 (2013).

[26] Michael R. Geller and Zhongyuan Zhou, Phys. Rev. A **88**, 012314 (2013).

[27] Joseph Emerson, Marcus Silva, Osama Moussa, Colm Ryan, Martin Laforest, Jonathan Baugh, David G. Cory, and Raymond Laflamme, Science **317**, 1893 (2007).

[28] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Phys. Rev. A **78**, 012347 (2008).

[29] Cecilia C. López, Ariel Bendersky, Juan Pablo Paz, and David G. Cory, Phys. Rev. A **81**, 062113 (2010).

[30] T. Gullion, D. B. Baker, and M. S. Conradi, J. Magn. Reson. **89**, 479 (1990).

[31] A. A. Maudsley, J. Magn. Reson. **69**, 488 (1986).

[32] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J. Glaser, J. Magn. Reson. **172**, 296 (2005).

[33] Troy W. Borneman, Martin D. Hürlimann, and David G. Cory, J. Magn. Reson. **207**, 220 (2010).

[34] Michael A Nielsen, Phys. Lett. A **303**, 249 (2002).

[35] E. Knill, Nature **434**, 39 (2005).

[36] C. Granade and B. Criger, computer code QuaEC, University of Waterloo, Waterloo, ON, 2012.

[37] John Watrous, Theor. Comput. **5**, 217 (2009).

[38] Michael Grant and Stephen Boyd, CVX: *Matlab Software for Disciplined Convex Programming*, version 1.22, http://cvxr.com/cvx.

[39] P. Cappellaro, J. S. Hodges, T. F. Havel, and D. G. Cory, J. Chem. Phys. **125**, 044514 (2006).

[40] Ryogo Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962).

[41] N. G. van Kampen, Physica **74**, 215 (1974).

[42] N. G. van Kampen, Physica **74**, 239 (1974).

[43] J. R. Johansson, P. D. Nation, and Franco Nori, Comput. Phys. Commun. **184**, 1234 (2013).

[44] G. Lindblad, Commun. Math. Phys. **48**, 119 (1976).

[45] Vittorio Gorini, Andrzej Kossakowski, and E. C. G. Sudarshan, J. Math. Phys. **17**, 821 (1976).

[46] Heinz-Peter Breuer and Francesco Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2002).

[47] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Elsevier Butterworth-Heinemann, Oxford, 1951), Part 1.

[48] Carl Edward Rasmussen and Christopher K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning (The MIT Press, Cambridge, MA, 2005).

[49] F. N. Hooge and P. A. Bobbert, Physica B: Condensed Matter **239**, 223 (1997).

[50] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information*, 1st ed., Cambridge Series on Information and the Natural Sciences (Cambridge University Press, Cambridge, 2004).

[51] C. King and M. B. Ruskai, IEEE T. Inform. Theor. **47**, 192 (2001).

[52] M. F. Sacchi, J. Opt. B Quant. S. O. **7**, S333 (2005).

[53] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson, Phys. Rev. A **85**, 042311 (2012).