Estimating the bias of CX gates via character randomized benchmarking

Jahan Claes and Shruti Puri

Department of Applied Physics, Yale University, New Haven, Connecticut 06511, USA

Yale Quantum Institute, Yale University, New Haven, Connecticut 06511, USA

(Dated: June 2, 2022)

Recent work has demonstrated that high-threshold quantum error correction is possible for biasednoise qubits, provided one can implement a controlled-not (CX) gate that preserves the bias. Biaspreserving CX gates have been proposed for several biased-noise qubit platforms, most notably Kerr cats. However, experimentally measuring the noise bias is challenging as it requires accurately estimating certain low-probability Pauli errors in the presence of much larger state preparation and measurement (SPAM) errors. In this paper, we introduce bias randomized benchmarking (BRB) as a technique for measuring bias in quantum gates. BRB, like all RB protocols, is highly accurate and immune to SPAM errors. Our first protocol, CX-dihedral BRB, is a straightforward method to measure the bias of the entire CX-dihedral group. Our second protocol, interleaved bias randomized benchmarking (IBRB), is a generalization of interleaved RB tailored to the experimental constraints biased-noise qubits; this is a more involved procedure that directly targets the bias of the CX gate alone. Our BRB procedures occupy a middle ground between classic RB protocols that only estimate the average fidelity, and tomographic RB protocols that provide more detailed characterization of noise but require more measurements as well as experimental capabilities that are not necessarily available in biased-noise qubits.

I. INTRODUCTION

Achieving scalable quantum computing will require efficient error-correction to counter the effects of noise. Recent work has revealed that there exist error correcting codes that tolerate much higher noise rates [1-7] when the noise is biased, that is, when errors that cause bit flips are suppressed compared to errors that cause only phase flips. In order to efficiently detect errors, these proposals rely on using controlled-NOT (CX) gates that are *bias-preserving* such that bit-flip noise remain suppressed to leading order [2–10]. Without bias-preserving CX gates, error correction is possible but requires more complex circuits, reducing the effectiveness of the underlying code [11]. Recently, CX gates that preserve the noise bias have been theoretically proposed in Kerr cat [12] and other qubit platforms [13, 14] (see also [15] for improvements to bias-preserving CX gates in Kerr cat qubits). Experimental efforts towards realizing these proposals are rapidly growing. In order to determine the effectiveness of the bias-tailored codes implemented with the experimentally realized CX gates, it is necessary to estimate the amount of noise-asymmetry or bias along with the total error probability of these gates. However, it is challenging to precisely measure the rate of bit flip errors of highly-biased noise channels, as such rates are extremely low.

A common method for precisely estimating low error probabilities in quantum gates is randomized benchmarking (RB) [16–19] and its derivatives [20–36]. In RB, one randomly generates circuits from some set of elementary gates and estimates the average error probability as a function of the circuit depth; the decay rate of the error probability with circuit depth then gives information about the error rate in the circuit. RB protocols typically boast two main advantages over other characterization methods. Primarily, RB protocols decouple state preparation and measurement (SPAM) errors from gate errors. In addition, because RB involves evaluating the error rate of circuits composed of many elementary gates, small error probabilities are magnified and may be precisely measured. The accuracy of RB experiments to estimate properties of the error channel can be rigorously guaranteed in a variety of settings [23, 24, 37–41].

There exists a zoo of different RB protocols (see [25] for a taxonomy) which differ in whether they characterize a group of gates or a single gate, whether they measure only the average fidelity or additional properties of the noise, and whether they are tailored to specific experimental hardware. The original RB measures the fidelity averaged over elements of the Clifford group [16, 17], but there also exist extensions of RB which measure the average fidelity of other efficientlysimulatable groups [22, 26, 27], most notably character RB [21, 23, 24] which uses techniques from representation theory to significantly simplify the RB decay functions. In contrast to RB methods characterizing groups of gates, interleaved RB [20] and its variants [23, 28-31, 42] instead measure the average fidelity of a single gate or layer of gates. There also exist RB protocols that measure properties of the noise channel beyond the average fidelity for either groups of gates or specific gates [32–35, 43, 44]. Moreover, efforts have been directed towards designing both group and interleaved RB protocols that are specifically tailored to the gate-set available in a given hardware [24, 36, 42, 45].

In this spirit, here we introduce group and interleaved RB procedures tailored to extract information about the total noise as well as noise-asymmetry in biased-noise hardware. Our first procedure is CX-dihedral bias RB (BRB), which is a straightforward modification of CXdihedral RB [22, 23]. This procedure measures the noise bias and average fidelity of the entire CX-dihedral group. Our second procedure, interleaved bias RB (IBRB), is a generalization of interleaved RB [20] and the recently introduced 2-for-1 interleaved RB [23, 42]. In particular, the IBRB procedure is tailored to the available gate-set of biased-noise qubits [12, 46, 47]. This procedure uses interleaved CX and Z gates to estimate the bias and fidelity of the CX. Both the BRB and IBRB protocols make heavy use of the recently introduced framework of character RB [23, 24].

Our approach to biased-noise benchmarking is notably different from Pauli channel estimation [34, 35, 44]. Pauli channel estimation is an interleaved RB procedure that uses character RB and randomized compiling to measure the full set of Pauli-diagonal elements of the noise channel associated with any Clifford operation. While Pauli channel estimation would appear to be sufficient for measuring the bias of CX gates, it has two drawbacks that make it unsuitable for estimating bias in biasednoise qubits. First, Pauli channel estimation demands considerable experimental overhead, as it estimates the probability of each Pauli error rather than the probability of sets of Pauli errors such as bit-flips and phase-flips. More importantly, Pauli channel estimation requires interleaving the full Pauli group between the Clifford operators, and is only guaranteed to measure the Clifford's noise when the Pauli group is high-fidelity. This is not suitable for biased-noise qubits, where we generically expect X and Y gates to be of comparable fidelity to CXgates if implemented in a bias-preserving manner [12]. On the other hand, if X and Y gates are not implemented in a bias-preserving manner then it won't be possible to accurately estimate the suppressed bit-flip rate in the CX gates. In contrast to Pauli channel estimation, our interleaved bias RB is designed specifically to use only interleaved Z gates, which are trivially bias-preserving and can be implemented with high-fidelity [8, 12].

Our paper is organized as follows. In Section II, we define the bias and fidelity of a multi-qubit gate in terms of the Pauli-diagonal part of its noise channel. In Section III, we introduce the Pauli, Z, and CX-dihedral groups that we will use to benchmark the bias. In Section IV, we give the step-by-step instructions for our bias RB procedures and illustrate them with simulated experiments. The derivations of these procedures, as well as technical details about biased-noise error channels, are relegated to appendices.

II. DEFINING THE BIAS

Intuitively, the bias of a noise channel Λ is a measure of the likelihood that an error will not flip a bit and will instead only apply an erroneous phase. We refer to errors that only apply an erroneous phase as **dephasing errors**, while we refer to errors that include a bit-flip as **non-dephasing errors**, even if they also apply erroneous phases. For example, an error $\mathbb{1} \otimes Z$ is dephasing, while $Y \otimes Z$ is non-dephasing.

To formally define the bias, we introduce the χ -matrix of a quantum channel [48]. Since the Pauli group on N qubits forms a basis for the set of all operators on N qubits, we can write the action of an arbitrary noise channel as

$$\Lambda(\rho) = \sum_{\substack{\vec{\alpha}_1, \vec{\beta}_1 \\ \vec{\alpha}_2, \vec{\beta}_2}} \chi_{\vec{\alpha}_1 \vec{\beta}_1, \vec{\alpha}_2 \vec{\beta}_2} \mathbf{X}(\vec{\alpha}_1) \mathbf{Z}(\vec{\beta}_1) \rho \mathbf{Z}(\vec{\beta}_2) \mathbf{X}(\vec{\alpha}_2)$$
(1)

for some coefficients $\chi_{\vec{\alpha}_1\vec{\beta}_1,\vec{\alpha}_2\vec{\beta}_2}$ [49]. Here $\vec{\alpha}_i, \vec{\beta}_i \in \mathbb{Z}_2^N$ are vectors of 0s and 1s that index the Pauli group, and for any single qubit operator O we use the notation

$$\mathcal{O}(\vec{v}) := \mathcal{O}_1^{v_1} \otimes \dots \otimes \mathcal{O}_N^{v_N}.$$
⁽²⁾

The diagonal of the χ -matrix is non-negative and sums to one. As a Pauli error is dephasing if and only if $\vec{\alpha} = 0$, we define the probabilities of dephasing and nondephasing errors, $p_{\rm D}$ and $p_{\rm ND}$, by

$$p_{\mathrm{D}} := \sum_{\vec{\beta} \neq \vec{0}} \chi_{0\vec{\beta},0\vec{\beta}} \tag{3}$$

$$p_{\rm ND} := \sum_{\substack{\vec{\alpha} \neq \vec{0} \\ \vec{\beta}}} \chi_{\vec{\alpha}\vec{\beta},\vec{\alpha}\vec{\beta}}.$$
 (4)

Finally, we define the **bias** η as the ratio of the error probabilities,

$$\eta := p_{\rm D}/p_{\rm ND}.\tag{5}$$

This definition of the bias for a multi-qubit gate was originally given in [12].

The probability of no error is given by $\chi_{\vec{00},\vec{00}}$, which is related to $p_{\rm D}$ and $p_{\rm ND}$ in the obvious way

$$\chi_{\vec{0}\vec{0},\vec{0}\vec{0}} = 1 - p_{\rm D} - p_{\rm ND}.$$
 (6)

The usual measure of gate quality is the average fidelity F_{Λ} , which in can be written in terms of the χ -matrix as [50, 51]

$$F_{\Lambda} = \frac{2^N \chi_{\vec{0}\vec{0},\vec{0}\vec{0}} + 1}{2^N + 1} \tag{7}$$

where N is the number of qubits. There exist numerous RB procedures to measure the average fidelity of a group of gates [16, 17, 21–25] or a specific gate [20, 23, 28, 29, 31, 42] and therefore determine $\chi_{\vec{n}\vec{0}}$ $\vec{n}\vec{0}$.

Note that while $p_{\rm D}$ and $p_{\rm ND}$ only directly give information about the diagonal elements of the χ -matrix, complete-positivity of the noise channel Λ implies the offdiagonal elements of the χ -matrix are bounded by the diagonal elements as [33, Appendix D]

$$|\chi_{\vec{\alpha}_1\vec{\beta}_1,\vec{\alpha}_2\vec{\beta}_2}| \le \sqrt{\chi_{\vec{\alpha}_1\vec{\beta}_1,\vec{\alpha}_1\vec{\beta}_1}\chi_{\vec{\alpha}_2\vec{\beta}_2,\vec{\alpha}_2\vec{\beta}_2}}.$$
 (8)

One can use this bound to show that if two biased noise channels Λ_A and Λ_B have non-dephasing error probabilities p_{ND}^A and p_{ND}^B , their composition will still have

$$p_{\rm ND} \lesssim p_{\rm ND}^A + p_{\rm ND}^B + 2\sqrt{p_{\rm ND}^A p_{\rm ND}^B} \tag{9}$$

and similar for $p_{\rm D}$ (see Appendix B). Thus, $p_{\rm ND}$ as defined in Eq. 4 is a useful characterization of the channel's behavior under composition even if the off-diagonal elements are not negligible. We also note that previous work on quantum error correction has demonstrated that measurement of the stabilizers of an error-correcting code causes the off-diagonal elements of the χ -matrix to rapidly decay [52–55].

III. THE PAULI, Z, AND CX-DIHEDRAL QUANTUM GROUPS

We will focus specifically on the Pauli, Z, and CXdihedral [22, 56] groups. These are all finite subgroups of the full unitary group that can be efficiently simulated. We will consider these groups to be defined modulo overall phases for convenience.

The N-qubit Pauli group \mathcal{P}_N is the group generated by the Pauli X and Z operators on all N qubits. As XZ = ZX up to a phase, we may assume all X operators appear to the left of all Z operators. We then have

$$\mathcal{P}_N := \langle \mathbf{X}_1, \mathbf{Z}_1, \dots, \mathbf{X}_N, \mathbf{Z}_N \rangle / U(1) \tag{10}$$

$$= \left\{ \mathbf{X}(\vec{\alpha})\mathbf{Z}(\vec{\beta}) : \vec{\alpha} \in \mathbb{Z}_2^N, \vec{\beta} \in \mathbb{Z}_2^N \right\}.$$
(11)

Since $X(\vec{\alpha}_1)Z(\vec{\beta}_1)X(\vec{\alpha}_2)Z(\vec{\beta}_2) = X(\vec{\alpha}_1 + \vec{\alpha}_2)Z(\vec{\beta}_1 + \vec{\beta}_2)$ up to a phase, this group is isomorphic to \mathbb{Z}_2^{2N}

The N-qubit Z group \mathcal{Z}_N is the commutative subgroup of the Pauli group generated by all single-qubit Z:

$$\mathcal{Z}_N := \langle \mathbf{Z}_i : 1 \le i \le N \rangle / U(1) \tag{12}$$

$$= \left\{ \mathbf{Z}(\vec{\beta}) : \vec{\beta} \in \mathbb{Z}_2^N \right\}.$$
(13)

This group is clearly isomorphic to \mathbb{Z}_2^N .

Finally, the N-qubit CX-Dihedral group \mathcal{D}_N is the group generated by the N-qubit dihedral group along with all CX gates between any pair of qubits [22]:

$$\mathcal{D}_N := \langle \mathbf{X}_i, \mathbf{T}_i, \mathbf{C}\mathbf{X}_{i,j} : 1 \le i, j \le N \rangle / U(1).$$
(14)

Here, $T_i := \exp\{i\pi Z_i/4\}$ is the T-gate, while $CX_{i,j}$ denotes the CX gate with the first index, *i*, the control and the second index, *j*, the target. Simple presentations of \mathcal{D}_N were introduced in [22, 57], and efficient decompositions of elements of \mathcal{D}_N which minimize the number of two-qubit gates are given in [56].

IV. THE BIAS RB PROCEDURES

In this section, we outline our two bias RB procedures. The first is a simple and scalable protocol for measuring

b	$ ho_b$	E_b	$\chi_b \Big(\mathbf{X}(\vec{\alpha}) \mathbf{Z}(\vec{\beta}) \Big)$	$S_b(n)$
1	$\bigotimes^N 0\rangle \langle 0 $	$\bigotimes^N \mathbf{Z}$	$(-1)^{ \vec{\alpha} }$	$A_1\lambda_1^n$
2	$\bigotimes^{N} +\rangle\langle+ $	$\bigotimes^N \mathbf{X}$	$(-1)^{ \vec{\beta} }$	$A_2\lambda_2^n$

TABLE I. Initial states, measurements, weighting, and fitting functions for CX-dihedral bias RB. Here, $|\vec{\alpha}|$ denotes the \mathbb{Z}_{2} -valued sum of all elements of $\vec{\alpha}$, and similarly for $|\vec{\beta}|$.

 $p_{\rm D}$ and $p_{\rm ND}$ for the entire CX-dihedral group, which may be a useful proxy for the performance of the CX gate when the X and T gates are high-fidelity.

The second is a protocol for measuring $p_{\rm D}$ and $p_{\rm ND}$ for the CX gate $CX_{1,2}$ acting on two qubits, by interleaving with elements of \mathcal{Z}_2 and randomly swapping between the usual $CX_{1,2}$ gate and a $|0\rangle$ -controlled CX gate which flips the target qubit if the control qubit is in the $|0\rangle$ state. The $|0\rangle$ -controlled CX gate can be written as a CX gate conjugated by a Pauli X on the control qubit, $X_1 C X_{1,2} X_1$. This second protocol is specifically designed for biased-noise, stabilized-cat qubits, where the fidelity of diagonal Z gates is much higher than the fidelity of X, and where we can apply both a bias-preserving $CX_{1,2}$ and a bias-preserving $|0\rangle$ -controlled CX gate by simply changing the phase of a drive [9, 12, 13] so that these two operations have similar $p_{\rm D}$ and $p_{\rm ND}$. However, we expect this protocol to be generally applicable to other biasednoise architectures, since arbitrary biased-noise architectures will likely have much higher-fidelity diagonal gates than off-diagonal gates [9, 13], and $X_1CX_{1,2}X_1$ differs from $CX_{1,2}$ simply by swapping the roles of $|0\rangle$ and $|1\rangle$ in the control qubit.

A. CX-dihedral bias RB (BRB)

Standard Clifford RB measures the fidelity averaged over elements of the Clifford group; similarly, CXdihedral bias RB measures $p_{\rm D}$ and $p_{\rm ND}$ averaged over elements of the CX-dihedral group \mathcal{D}_N . Our procedure is essentially identical to the previous character RB procedure for estimating the fidelity of the CX-dihedral group [21, 23]. The only modification is a post-processing step to extract the dephasing and non-dephasing error probabilities instead of just the average fidelity.

For convenience, we make the standard RB assumption of gate-independent noise, so that the noisy implementation of any $U \in \mathcal{D}_N$ is $\Lambda \circ U$ for a noise channel Λ independent of U. However, like the usual character RB, our procedure works for gate-dependent noise, provided all $U \in \mathcal{D}_N$ are high-fidelity; see [23, 24] for proofs.

The CX-dihedral bias RB procedure is:

- 1. For each b in Table I:
 - (a) For arbitrary n, choose unitaries $U_0 \in \mathcal{P}_N$ and $U_1, ..., U_n \in \mathcal{D}_N$ at random. Set $U_{n+1} = U_1^{\dagger} \cdots U_n^{\dagger}$.

- (b) Prepare the initial state ρ_b listed in Table I.
- (c) Successively apply the gates (U_1U_0) , U_2 , U_3 , ..., U_{n+1} . Note that instead of applying U_0 and then U_1 , we compile the product U_1U_0 into a single element of \mathcal{D}_N .
- (d) Perform a measurement of the observable E_b in Table I. Weight the outcome by $\chi_b^*(U_0)$.
- (e) Repeat steps 1a-1d many times to estimate the character-weighted survival probability

$$S_b(n) := \underset{\substack{U_0 \in \mathcal{P}_N \\ U_1 \cdots U_n \in \mathcal{D}_N}}{\mathbb{E}} \left[\chi_b^*(U_0) P_{\{U_i\}} \right]$$
(15)

where $P_{\{U_i\}}$ denotes the expectation value of E_b after applying the gates in step 1c and $\mathbb{E}[\cdot]$ denotes the average over the choices of gates.

- (f) Repeat steps 1a-1e for many different values of n to estimate the whole $S_b(n)$ curve.
- (g) Fit the decay curve, $S_b(n)$, to the functional form listed in Table I to estimate λ_b .
- 2. Estimate the dephasing and non-dephasing error probabilities of the error channel Λ as

$$p_{\rm D} = \frac{2^N - 1}{4^N} \left[1 + \left(2^N - 1 \right) \lambda_1 - 2^N \lambda_2 \right]$$

$$p_{\rm ND} = \frac{2^N - 1}{2^N} \left[1 - \lambda_1 \right].$$
 (16)

Note that the weights and measurement outcomes satisfy $|\chi_b^*(U_0)P_{\{U_i\}}| \leq 1$, so that if we take \mathcal{N}_s samples to estimate $S_b(n)$ at a specific value of b and n, the statistical uncertainty in our estimate is roughly $1/\sqrt{\mathcal{N}_s}$, independent of n or N. To achieve a good relative uncertainty, then, we simply need the true value of $S_b(n)$ to be sufficiently large. For high-fidelity gates, we show in our derivation in Appendix A 4 that $A_b \approx 1$, $\lambda_b \approx 1$, so that we can reliably fit the decay curve.

We give an example of CX-dihedral BRB in Fig. 1. Here, we generate random error channels Λ by generating random sets of Kraus operators, and simulate a CXdihedral BRB experiment (see Appendix D for details on the random error channels). Fig. 1a illustrates the experiment for a single error channel, where we estimate the value of $S_b(n)$ for a few different values of n and fit the data to the functional forms given in Table I. In this simulation, each value of $S_b(n)$ is estimated using 5000 measurements. To demonstrate the effectiveness of our procedure, we repeat this experiment for many different error channels, using Eq. 16 to estimate $p_{\rm D}$ and $p_{\rm ND}$ and finally using these estimates to extract the bias. In Fig. 1b we plot our estimated probabilities versus the exact probabilities of the error channel, and in Fig. 1c we do the same for the bias. To estimate the error bars, we use bootstrap resampling. Visually, it is clear that we are accurately estimating $p_{\rm D}$, $p_{\rm ND}$, and η even for very high biases. To verify this, we compute the reduced- χ^2



FIG. 1. Simulated CX-dihedral bias RB experiments. (a) An example of estimating $S_b(n)$, with the $S_b(n)$ curves plotted in orange and our estimates of $S_b(n)$ given by blue dots. To estimate sensitivity to statistical errors, we generate multiple fits of $S_b(n)$ using bootstrap resampling of our data, which are plotted in grey and generally overlap the exact $S_b(n)$ curves. (b) The probabilities p_D and p_{ND} extracted from the fits for 50 randomly generated error channels. Error bars denote the standard deviation over 50 resamplings. (c) The estimated bias η for each of these error channels. We see that even at very high bias, we can accurately estimate the value of η .

statistic for our estimates of $p_{\rm D}$, $p_{\rm ND}$, and η , and find χ^2 between 1 and 1.3, indicating that our bootstrapping is accurately estimating the error bars to within a factor of between 1 and $\sqrt{1.3}$.

В. Interleaved bias RB (IBRB)

A common approach to estimate the fidelity of a single gate is interleaved RB, in which one interleaves the gate of interest with random elements from an interleaving group designed to simplify the error channel. Originally, the gate of interest was required to be from the interleaving group [20], but later work has relaxed this requirement [23, 28, 29, 31, 42]. Interleaved RB estimates the fidelity of the combined error channel $\Lambda_U \circ \Lambda_G$ of the gate U and interleaving group G; knowing the fidelity of $\Lambda_U \circ \Lambda_G$ and the fidelity of Λ_G allows one to bound the fidelity of Λ_U , with bounds that become tight as the fidelity of Λ_G approaches one [20, 33, 58]. Thus, we typically require the interleaving group to have high-fidelity. On the other hand, the advent of randomized compiling [59–61] implies that in some cases it is not necessary to separately estimate the fidelities of Λ_U and Λ_G , as $\Lambda_U \circ \Lambda_G$ is the relevant error channel for a circuit that has been randomly compiled with the group G. However, in the case of randomized compiling, it is still necessary for G to be high-fidelity, so that the randomized compilation does not add in significant additional noise.

We develop an interleaved bias RB procedure that directly estimates the bias of the CX gate in Kerr cat qubits [12, 46, 47] or other biased-noise platforms [9, 13, 14]. In biased-noise qubits, we generally expect gates that are diagonal in the Z-basis to have much higher fidelity than non-diagonal gates. It is thus desirable to have a protocol that uses \mathcal{Z}_2 as the interleaving group, as the fidelity of X and Y gates may be no better than the fidelity of the CX gate [62]. Restricting to an interleaving group such as \mathcal{Z}_2 that is diagonal in the Z-basis introduces considerable complications, as we will see below.

For convenience, we define $C := CX_{1,2}$ to be a CX gate on the two qubits. We also define the gate C' := $X_1CX_{1,2}X_1$, which is similar to C, except that it applies X_2 when qubit 1 is in $|0\rangle$ rather than $|1\rangle$. The Kerr cat system can implement bias-preserving versions of both of these gates using similar procedures, and we expect that both C and C' will have similar dephasing and non-dephasing probabilities. These features will likely be shared by any biased-noise system. Define the error channels Λ_C and $\Lambda_{C'}$ to be the error channels associated with C and C' respectively, so the noisy implementation of C is $C \circ \Lambda_C$ and similar for C'. In addition define Λ_G to be the error channel associated with gates $U \in \mathbb{Z}_2$, so the noisy implementation of U is $\Lambda_G \circ U$. Finally, define $\Lambda = \Lambda_C \circ \Lambda_G$ and $\Lambda' = \Lambda_{C'} \circ \Lambda_G$ to be the composed error channels. Our protocol will directly estimate the dephasing and non-dephasing probabilities of the average

5

b	$ ho_b^{(a)}$	$E_b^{(a)}$	$\chi_b \left(\mathbf{Z}_1^{\beta_1} \mathbf{Z}_2^{\beta_2} \right)$	$S_b(n)$
0+	$rac{1}{2} Z \otimes 0 angle \langle 0 $	$Z\otimes \mathbb{1}$	1	$A_{0+}\lambda_{0+}^n + B_{0+}$
0-	$rac{1}{2}\mathrm{Z}\otimes 0 angle\langle0 $	$\mathbf{Z}\otimes\mathbf{Z}$	1	$A_{0-}\lambda_{0-}^{n} + B_{0-}\kappa_{0-}^{n}$
1+	$rac{1}{2}\mathbbm{1}\otimes + angle\langle+ $	$\mathbb{1}\otimes X$	$(-1)^{\beta_2}$	$A_{1+}\lambda_{1+}^n + B_{1+}\kappa_{1+}^n$
	$rac{1}{2}Z\otimes + angle\langle+ $	$\mathbf{Z}\otimes\mathbf{X}$	(1)	
1-	$ +i+i\rangle\langle+i+i $	$\mathbb{1}\otimes Y$	$(-1)^{\beta_2}$	$A_{1-}\lambda_{1-}^{n} + B_{1-}\kappa_{1-}^{n}$
2+	$ +0 angle\langle+0 $	$X\otimes \mathbb{1}$	$(-1)^{\beta_1}$	$A_{2+}\lambda_{2+}^n + B_{2+}\kappa_{2+}^n$
	$ +i \ 0 angle\langle+i \ 0 $	$Y\otimes \mathbb{1}$	(-1)	
2-	$ +0 angle\langle+0 $	$\mathbf{X}\otimes\mathbf{Z}$	$(-1)^{\beta_1+\beta_2}$	$A_{2-}\lambda_{2-}^n + B_{2-}\kappa_{2-}^n$
	$ +i \ 0 angle\langle+i \ 0 $	$\mathbf{Y}\otimes\mathbf{Z}$	(-1)	

TABLE II. Initial states, measurements, weighting, and fitting functions for interleaved bias RB. In the fitting functions, λ_b and κ_b are two independent RB decay parameters. $|\pm i\rangle$ denotes eigenstates of the Y operator.

channel $(\Lambda + \Lambda')/2$. If the dephasing and non-dephasing probabilities of Λ_C and $\Lambda_{C'}$ are close and $\Lambda_G \approx 1$, this is identical to the dephasing and non-dephasing probabilities of Λ_C alone. On the other hand, even without these assumptions, we will see below that $(\Lambda + \Lambda')/2$ is the relevant error channel for a certain randomized compilation procedure.

The interleaved bias RB procedure is:

- 1. For $b = 0 \pm$ and $b = 1 \pm$ in Table II:
 - (a) For arbitrary n, choose unitaries $U_1, ..., U_{n+1} \in \mathcal{Z}_2$ uniformly at random. Also choose CX gates $C_1, ..., C_n \in \{C, C'\}$ uniformly at random.
 - (b) If there is more than one initial state ρ_b listed in Table II, randomly select one of the listed initial states and prepare it.
 - (c) Alternatively apply the gates from \mathcal{Z}_2 and the CX gates as $U_1, C_1, U_2, C_2, ..., U_n, C_n, U_{n+1}$.
 - (d) Perform a measurement of the observable E_b corresponding to the ρ_b selected in step 1b. Weight the outcome by $\chi_b^*(\{U_i\})\sigma_{\pm}(\{C_i\})$, where we define

$$\sigma_{\pm}(\{C_i\}) := (\pm 1)^{\delta_{C_1,C'} + \dots + \delta_{C_n,C'}}$$
(17)

$$\chi_b^*(\{U_i\}) := \prod_i \chi_b^*(U_i)$$
(18)

and $\chi_b(U_i)$ is given in Table II. If b = 0or b = 1 – the effect of σ is to multiply the measurement outcome by (-1) for each C' we apply, while otherwise σ is trivial.

(e) Repeat steps 1a-1d many times, to estimate the signed character-weighted survival probability

$$S_{b}(n) := \underset{\substack{\nu_{b} \\ U_{0}\cdots U_{n+1} \in \mathcal{Z}_{2} \\ C_{1},\dots,C_{n} \in \{C,C'\}}}{\mathbb{E}} \left[\chi_{b}^{*}(\{U_{i}\})\sigma_{\pm}(\{C_{i}\})P_{\{C_{i}\},\{U_{i}\},\rho_{b}} \right]$$
(19)

where $P_{\{C_i\},\{U_i\},\rho_b}$ denotes the expectation value of E_b after applying the gates in step 1c to ρ_b .

- (f) Repeat steps 1a-1e for many different values of n to estimate the whole $S_b(n)$ curve.
- (g) Fit $S_b(n)$ to the functional form listed in Table II to estimate λ_b and κ_b , where we take the convention $\Re(\lambda_b) > \Re(\kappa_b)$.
- 2. For $b = 2\pm$ in Table II:
 - (a) For arbitrary n, choose unitaries $U_1, ..., U_{2n+1} \in \mathbb{Z}_2$ uniformly at random. Also choose CX gates $C_1, ..., C_{2n} \in \{C, C'\}$ uniformly at random.
 - (b) If there is more than one initial state ρ_b listed in Table II, randomly select one of the listed initial states and prepare it.
 - (c) Alternatively apply the gates from \mathcal{Z}_2 and the CX gates as $U_1, C_1, U_2, C_2, ..., U_{2n}, C_{2n}, U_{2n+1}$.
 - (d) Perform a measurement of the observable E_b corresponding to the ρ_b selected in step 2b. Weight the outcome by $\chi_b^*(\{U_i\})\sigma_{\pm}(\{C_i\})$, where we define

$$\sigma_{\pm}(\{C_i\}) := (\pm 1)^{\delta_{C_1,C'} + \dots + \delta_{C_{2n},C'}}$$
(20)

$$\chi_b^*(\{U_i\}) := \prod_{i \text{ odd}} \chi_{2+}^*(U_i) \prod_{i \text{ even}} \chi_{2-}^*(U_i) \qquad (21)$$

where $\chi_{2\pm}(U_i)$ are given in Table II.

(e) Repeat steps 2a-2d many times, to estimate the signed character-weighted survival probability

$$S_{b}(n) := \underset{\substack{\rho_{b} \\ U_{0}\cdots U_{2n+1} \in \mathcal{Z}_{2} \\ C_{1},\dots,C_{2n} \in \{C,C'\}}}{\mathbb{E}} \left[\chi_{b}^{*}(\{U_{i}\})\sigma_{\pm}(\{C_{i}\})P_{\{C_{i}\},\{U_{i}\},\rho_{b}} \right]$$
(22)

where $P_{\{C_i\},\{U_i\},\rho_b}$ denotes the expectation value of E_b after applying the gates in step 2c to ρ_b .

- (f) Repeat steps 2a-2e for many different values of n to estimate the whole $S_b(n)$ curve.
- (g) Fit $S_b(n)$ to the functional form listed in Table II to estimate λ_b and κ_b , where we take the convention $\Re(\lambda_b) > \Re(\kappa_b)$.
- 3. Estimate the dephasing and non-dephasing error probabilities of the combined error channel Λ as

$$p_{\rm D} = \frac{1}{16} \begin{bmatrix} 3\lambda_{0+} + 3\lambda_{0-} - 3\kappa_{0-} - \lambda_{1+} \\ -\kappa_{1+} - \lambda_{1-} + \kappa_{1-} - \lambda_{2+} \\ -\kappa_{2+} - \lambda_{2-} - \kappa_{2-} - 1 \end{bmatrix}$$
(23)
$$p_{\rm ND} = 1 - \frac{1}{4} \begin{bmatrix} 1 + \lambda_{0+} + \lambda_{0-} - \kappa_{0-} \end{bmatrix}.$$

To realize the mixed, non-positive state $\rho = \frac{1}{2}Z \otimes |0\rangle\langle 0|$ we simply prepare either $|00\rangle\langle 00|$ or $|10\rangle\langle 10|$ with equal probability, and weight the resulting measurement by (-1) if we prepare $|10\rangle\langle 10|$. We realize the states $\rho = \frac{1}{2}\mathbb{1} \otimes |+\rangle\langle +|$ and $\rho = \frac{1}{2}Z \otimes |+\rangle\langle +|$ similarly.

In this procedure, for b = 0+ we need to accurately estimate λ_{0+} and for $b \neq 0+$ we need to accurately estimate both λ_b and κ_b . To accurately fit these decay parameters, we require the prefactors A_{0+} and A_b , B_b for $b \neq 0+$ to be large. As we will demonstrate in our derivation in Appendix A 5, for high-fidelity gates we expect $A_{0+} \approx 1$, as well as $A_b \approx 1/2$ and $B_b \approx 1/2$ for $b \neq 0+$, so that we may accurately fit the decay parameters.

In the case of b = 0- and b = 1-, we also need to distinguish between λ_b and κ_b , since they enter into Eq. 23 with different signs. As we demonstrate in Appendix A 5, for high-fidelity gates we expect $\lambda_b \approx 1$ and $\kappa_b \approx -1$, so we can define λ_b to be the decay constant with the largest real part. In the case of b = 0+, b = 1+, and $b = 2\pm$ we cannot differentiate between λ_b and κ_b , since they are both ≈ 1 for high-fidelity gates, but they enter Eq. 23 with the same sign and therefore do not need to be distinguished.

While there are nine initial states and measurements listed in Table II, one can use the same experimental data for b = 0+ and b = 0-, the two rows of b = 1+, for the first row of b = 2+ and the first row of b = 2-, and for the second row of b = 2+ and the second row of b = 2-, reducing the cost to five distinct pairs of initial states and measurements.

We give an example of IBRB in Fig. 2. Here, we generate random error channels Λ_G , Λ_C , and $\Lambda_{C'}$ by generating random sets of Kraus operators, and simulate an IBRB experiment (see Appendix D for details on the random error channels). Fig. 2a illustrates the experiment for a single error channel, where we estimate the value of $S_{h}(n)$ for a few different values of n and fit the data to the functional forms given in Table II. Again, each value of $S_b(n)$ is estimated using 5000 measurements. Note that for b = 1, 2- the function $S_b(n) = A_b \lambda_b^n + B_b \kappa_b^n$ oscillates with period 2, because $\lambda_b \approx 1$ and $\kappa_b \approx -1$. However, we can still accurately estimate the parameters $A_b, B_b, \lambda_b, \kappa_b$ by taking only a few widely spaced data points, as shown in Fig. 2a, provided we take data at both even and odd sequence lengths n. This is because the rapid oscillations are constrained to have period 2 by the form of $S_b(n)$, so it is not necessary to take finegrained data at nearby values of n to fit this rapidly oscillating function. To demonstrate the effectiveness of our procedure, we repeat this experiment for many different error channels, using Eq. 23 to estimate $p_{\rm D}$ and $p_{\rm ND}$ and finally using these estimates to extract the bias. In Fig. 2b we plot our estimated probabilities versus the exact probabilities of the error channel, and in Fig. 2c we do the same for the bias. To estimate the error bars, we again use bootstrap resampling. Visually, it is clear that we are accurately estimating $p_{\rm D}$, $p_{\rm ND}$, and η even for very high biases. To verify this, we again compute



FIG. 2. Simulated interleaved bias RB experiments. (a) An example of estimating $S_b(n)$, with the $S_b(n)$ curves plotted in orange and our estimates of $S_b(n)$ given by blue dots. Note that $S_b(n)$ is oscillatory for b = 0, 1-, since the decay constants λ_b, κ_b have opposite sign. To robustly fit $S_b(n)$ in the presence of these oscillations does not require densely sampling data points, since we know the oscillatory period is always 2; however, it does require taking data at both even and odd sequence lengths, as shown in the figure. To estimate sensitivity to statistical errors, we generate multiple fits of $S_b(n)$ using bootstrap resampling of our data, which are plotted in grey and generally overlap the exact $S_b(n)$ curves. (b) The probabilities p_D and p_{ND} extracted from the fits for 50 randomly generated error channels. Error bars denote the standard deviation over 50 resamplings. (c) The estimated bias η for each of these error channels. We see that even at very high bias, we can accurately estimate the value of η .

the reduced- χ^2 statistic for our estimates, and find χ^2 between 2 and 3, indicating that our bootstrapping is accurately estimating the error bars to within a factor of between $\sqrt{2}$ and $\sqrt{3}$.

C. Randomized compiling for interleaved bias RB

Our IBRB measures $p_{\rm D}$ and $p_{\rm ND}$ for the averaged, composite channel $(\Lambda + \Lambda')/2$, with $\Lambda = \Lambda_C \circ \Lambda_G$ and $\Lambda' = \Lambda_{C'} \circ \Lambda_G$. Provided that $p_{\rm D}$ and $p_{\rm ND}$ for Λ_C and $\Lambda_{C'}$ are equal and $\Lambda_G = 1$, these are also the dephasing and non-dephasing probabilities for Λ_C alone. However, in general we would like to avoid assuming the error channels Λ_C and $\Lambda_{C'}$ are identical, and we would like to allow for the possibility of $\Lambda_G \neq 1$.

Previous interleaved RB procedures for determining the fidelity of a gate C have a similar problem; in these procedures one separately estimates the fidelity of $\Lambda_C \circ \Lambda_G$ and Λ_G , and uses this information to provide bounds on the fidelity of Λ_C alone [20]. These bounds depend on the fidelity of Λ_G , with lower fidelity Λ_G resulting in looser bounds for the fidelity of Λ_C . From this point of view, it is important for the interleaving group to be high-fidelity, in order to be able to tightly bound the fidelity of Λ_C . It is also possible to use this method to bound p_D and p_{ND} of $(\Lambda_C + \Lambda_{C'})/2$, provided we know the corresponding probabilities for Λ_G and Λ (see Appendix B 3). However, benchmarking p_D and p_{ND} of Λ_G is challenging for $G = Z_2$ as averaging over sequences of Z_2 gates does not randomize error channels enough to guarantee a simple form of the survival probability.

On the other hand, the technique of randomized compiling [59–61] provides an alternative interpretation of interleaved RB results. In randomized compiling, one intentionally inserts random elements of a high-fidelity interleaving group between the lower-fidelity gates C in order to eliminate the coherence of the noise. In a circuit that has been randomly compiled, the error channel associated with a gate C is the combined error channel $\Lambda_C \circ \Lambda_G$. From this point of view, we require the interleaving group to be high-fidelity to ensure that randomly inserting elements of the interleaving group into a circuit does not notably increase the errors in the circuit.

We can do a modified version of randomized compiling for bias-preserving CX gates. Each time a CX gate appears in a circuit, we randomly insert an element of Z_2 before it, and with 50% probability replace it with $|0\rangle$ -controlled CX, $X_c C X_{c,t} X_c$. These extra Pauli operators can be commuted through the rest of the circuit and their effect can be tracked in software. This is illustrated in Fig. 3 for the example of a circuit that measures the stabilizer of the XZZX surface code [5].

The error channel for the resulting randomized gate



FIG. 3. Randomizing a single $CX_{c,t}$ gate in a circuit that measures the stabilizer of the XZZX code, where c denotes the control qubit and t the target. Rather than apply the bare $CX_{c,t}$ gate, we randomly insert an element of \mathcal{Z}_2 before the gate, and randomly apply either $CX_{c,t}$ or $X_cCX_{c,t}X_c$. This modification of the circuit results in an overall Pauli operator, which can be tracked in software.

then has the $p_{\rm D}$ and $p_{\rm ND}$ that we measure in our experiment. Therefore, we can always achieve the error rates $p_{\rm D}$ and $p_{\rm ND}$ by this randomized compiling procedure.

Like the usual randomized compiling, our biased-noise randomized compilation procedure has the additional benefit of limiting how errors can build up when composing noisy gates. We show in Appendix C that if two error channels Λ_A and Λ_B are randomly compiled as above, their composition $(\Lambda_A \circ \Lambda_B)$ will have a non-dephasing error probability

$$p_{\rm ND} \approx p_{\rm ND}^A + p_{\rm ND}^B,$$
 (24)

which, in contrast to to Eq. 9, says that $p_{\rm ND}$ grows linearly rather than quadratically in the number of composed error channels when the circuit is randomly compiled by Z_N . This is similar to the behavior of the average infidelity for full randomized compilation, which also may grow quadratically under composition for generic noise channels but grows linearly for randomly compiled circuits [58].

V. CONCLUSION

Measuring the bias of a highly-biased gate is a delicate process, as the non-dephasing error probability must be precisely estimated. By using techniques from randomized benchmarking, we can precisely estimate these error probabilities. The essential ingredient in our method is defining efficiently measurable weighted survival probabilities whose decay rates depend only on $p_{\rm ND}$. Because we consider variable sequence lengths, our method estimates gate error rates independently from SPAM errors, even if the SPAM errors are much larger than the nondephasing gate errors. By measuring the weighted survival probabilities for long gate sequences, we can magnify the effect of small non-dephasing errors, allowing us to precisely measure arbitrarily small error probabilities by simply increasing our sequence length.

Our interleaved bias RB in particular is highly tailored to the experimental constraints of biased noise qubits. In general, interleaved RB works because the interleaved gates randomize the error channel while not adding significant additional errors. However, in the case of biased noise qubits, we can only interleave Z gates without introducing additional errors; X and Y gates are generally as error-prone as CX gates. As a result, we were motivated to add additional randomization by swapping between C and C', which allowed for sufficient randomization of the error channel. This is in contrast to standard techniques for estimating Pauli channels, which assume one can freely add Pauli operators to a circuit without adding significant errors [34, 35, 44, 61]. We expect our techniques to be highly relevant to near-term experiments, as the numerous proposals for bias-preserving CX gates [12–14] are realized experimentally.

ACKNOWLEDGMENTS

We thank Joel J. Wallman for helpful discussions about randomized compiling and its implications for our procedure and Steven T. Flammia for a critical reading of our manuscript. This work is supported by the ARO under grant number W911NF18-1-0212.

Appendix A: Derivation of the procedures

To derive these procedures requires a detour into some background mathematics. To begin, we review a few necessary aspects of representation theory, and define a natural representation for quantum groups, the Liouville representations of the Liouville representation of \mathcal{P}_N , \mathcal{Z}_N , and \mathcal{D}_N . Then we derive how the dephasing and nondephasing probabilities may be written as the trace over invariant subspaces of the Liouville representations. Finally, armed with this background, we derive each of the bias RB procedures.

1. Background: Representation theory

Given a finite group G, a **unitary representation** is a map assigning a unitary matrix to each group element such that group multiplication is preserved:

$$\phi: G \to U(m), \quad \phi(g_1g_2) = \phi(g_1)\phi(g_2). \tag{A1}$$

where U(m) is the group of $m \times m$ unitary matrices acting on \mathbb{C}^m . A representation is **irreducible** if the image of ϕ doesn't preserve any proper subspace of \mathbb{C}^m . Every finite-dimensional representation can be uniquely decomposed as the direct sum of irreducible representations (irreps):

$$\phi \simeq a_1 \phi_1 \oplus \dots \oplus a_I \phi_I \tag{A2}$$

$$\mathbb{C}^m \simeq V_1^{a_1} \oplus \dots \oplus V_I^{a_I} \tag{A3}$$

where $\phi_i : G \to V_i$ is irreducible and $a_i \phi_i$ is standard shorthand for the direct sum of a_i copies of ϕ_i . The number a_i is referred to as the **multiplicity** of the *i*th irrep. Finally, the **character** χ of a representation is defined by

$$\chi(g) = \operatorname{Tr}\left(\phi(g)\right). \tag{A4}$$

We will repeatedly use two elementary facts about representations.

Fact 1 (Projection Formula). If $\phi \simeq a_1 \phi_1 \oplus \cdots \oplus a_I \phi_I$, then the projector $\hat{\Pi}_i$ onto $V_i^{a_i} \subseteq \mathbb{C}^m$ is given by

$$\hat{\Pi}_i = \frac{\dim(V_i)}{|G|} \sum_{g \in G} \chi_i^*(g) \phi(g) \tag{A5}$$

where χ_i is the character of ϕ_i .

Fact 2 (Schur's Lemma). If $\phi \simeq a_1 \phi_1 \oplus \cdots \oplus a_I \phi_I$, then for any matrix $Q \in GL(m)$ we have

$$\frac{1}{|G|} \sum_{g \in G} \phi(g^{-1}) Q \phi(g) \simeq (Q_1 \otimes \mathbb{1}_1) \oplus \dots \oplus (Q_I \otimes \mathbb{1}_I)$$
 (A6)

where Q_i is some $a_i \times a_i$ matrix, and $Q_i \otimes \mathbb{1}_i$ is a matrix that acts on $V_i^{a_i}$ by mixing the a_i copies of V_i but acting as the identity on the degrees of freedom within each copy of V_i . In particular, if $a_i = 1$ for all i, we have

$$\frac{1}{|G|} \sum_{g \in G} \phi(g^{-1}) Q \phi(g) = \sum_{i} \frac{\operatorname{Tr}(\hat{\Pi}_{i}Q)}{\operatorname{Tr}(\hat{\Pi}_{i})} \hat{\Pi}_{i}$$
(A7)

where $\hat{\Pi}_i$ is the projector onto the single copy of $V_i \subseteq \mathbb{C}^m$.

See [63] for proofs of these facts, as well as more details on the representation theory of finite groups.

In this paper, we will be interested in the case where $G \subset U(2^N)$ is a finite subgroup of the unitary group on N qubits. In this case, the standard action of $U \in G$ on a density matrix, $\rho \mapsto U\rho U^{\dagger}$, is a unitary representation of G on the vector space of density matrices. Choosing a basis for \mathcal{H} , we can more conveniently represent a density matrix $\rho = \sum_{i,j} \rho_{ij} |i\rangle \langle j|$ by a vector $|\rho\rangle\rangle := \sum_{i,j} \rho_{ij} |i\rangle |j\rangle$. In terms of this vectorized density matrix, it is simple to see that the action of a unitary U on $|\rho\rangle\rangle$ is given by

$$|U\rho U^{\dagger}\rangle\rangle = U \otimes U^{*}|\rho\rangle\rangle = \hat{U}|\rho\rangle\rangle \qquad (A8)$$

where we've defined $\hat{U} := U \otimes U^*$ to be the matrix representation of the unitary U acting on the space of vectorized density matrices. This representation $\phi : U \mapsto \hat{U}$

sending a $d \times d$ unitary to a $d^2 \times d^2$ unitary is known as the **Liouville representation**.

We can define the Liouville representation of a quantum channel $\Lambda : \rho \mapsto \sum_{a} K_{a} \rho K_{a}^{\dagger}$ by $\hat{\Lambda} := \sum_{a} K_{a} \otimes K_{a}^{*}$, in which case we have

$$|\Lambda(\rho)\rangle\rangle = \hat{\Lambda}|\rho\rangle\rangle.$$
 (A9)

Finally, we can write the expectation value of an observable E over a state ρ in terms of the Liouville representation as

$$\langle E \rangle_{\rho} = \operatorname{Tr}(E^{\dagger}\rho) = \langle \langle E | \rho \rangle \rangle$$
 (A10)

We refer to [48] for a more detailed treatment of both quantum channels and the Liouville representation.

2. Irreps of the quantum groups

For the Pauli, Z, and CX-dihedral groups, we will need to understand the decomposition of their Liouville representation into irreps, and the characters of those irreps.

In the case of the 1-qubit Pauli group, the Liouville representation decomposes into four non-isomorphic irreps, with projectors and characters given in Table III. The Liouville representation of the N-qubit Pauli group \mathcal{P}_N then decomposes into 4^N non-isomorphic irreps indexed by vectors $\vec{i} \in \mathbb{Z}_4^N$, with projectors and characters

$$\hat{\Pi}_{\vec{i}}^{\mathcal{P}} = \hat{\Pi}_{i_1}^{\mathcal{P}} \otimes \dots \otimes \hat{\Pi}_{i_N}^{\mathcal{P}}$$
(A11)
$$\chi_{\vec{i}} \left(X(\vec{\alpha}) Z(\vec{\beta}) \right) = \chi_{i_1} (X^{\alpha_1} Z^{\beta_1}) \cdots \chi_{i_N} (X^{\alpha_N} Z^{\beta_N}).$$

One can similarly determine a factorized form for the irreps of the Liouville representation of the Z group Z_N , but we will need only the case N = 2 here. The Liouville representation of Z_2 decomposes into 16 irreps with projectors and characters given Table IV. Note that in this case, each irrep has multiplicity 4.

Finally, the Liouville representation of the CX-dihedral group \mathcal{D}_N decomposes into the three irreps given in Table V. Note that the number of irreps is independent of N. Derivation of these irreps can be found in [22].

3. The dephasing/non-dephasing error probabilities in the Liouville representation

From the definition of the dephasing and nondephasing error probabilities, Eqs. 3 and 4, we want to express $p_{\rm D}$ and $p_{\rm ND}$ as the trace of $\hat{\Lambda}$ over subspaces \mathcal{Z}_N and $\mathcal{P}_N \setminus \mathcal{Z}_N$ of the Pauli group. Note that these subspaces are invariant under the action of any bias-

i	$\hat{\Pi}_i^{\mathcal{P}}$	$\chi_i(X^{\alpha}Z^{\beta})$
0	$\frac{1}{2} 1\rangle\rangle\langle\langle1 $	1
1	$\frac{1}{2} Z\rangle\rangle\langle\langle Z $	$(-1)^{\alpha}$
2	$\frac{1}{2} Y\rangle\rangle\langle\langle Y $	$(-1)^{\alpha+\beta}$
3	$\frac{1}{2} X\rangle\rangle\langle\langle X $	$(-1)^{\beta}$

TABLE III. Irreps, projectors, and characters of the Liouville representation of $\mathcal{P}_{1}.$

i	$\hat{\Pi}_i^{\mathcal{Z}}$	$\chi_i\left(Z_1^{\beta_1}Z_2^{\beta_2}\right)$
0	$\frac{1}{4} 11\rangle\rangle\langle\langle11 $ $\frac{1}{4} 12\rangle\rangle\langle\langle12 $ $\frac{1}{4} Z1\rangle\rangle\langle\langleZ1 $ $\frac{1}{4} Z2\rangle\rangle\langle\langleZ2 $	1
1	$\begin{array}{c} \frac{1}{4} \mathbbm{1}X\rangle\rangle\langle\langle\mathbbm{1}X \\ \frac{1}{4} \mathbbm{1}Y\rangle\rangle\langle\langle\mathbbm{1}Y \\ \frac{1}{4} ZX\rangle\rangle\langle\langle\mathbbm{2}X \\ \frac{1}{4} ZY\rangle\rangle\langle\langle\mathbbm{2}X \end{array}$	$(-1)^{\beta_2}$
2	$\begin{array}{c} \frac{1}{4} X\mathbbm{1}\rangle\rangle\langle\langle X\mathbbm{1} \\ \\ \frac{1}{4} XZ\rangle\rangle\langle\langle XZ \\ \\ \frac{1}{4} Y\mathbbm{1}\rangle\rangle\langle\langle Y\mathbbm{1} \\ \\ \\ \frac{1}{4} YZ\rangle\rangle\langle\langle YZ \end{array}$	$(-1)^{\beta_1}$
3	$\begin{split} & \frac{1}{4} \overline{X} \overline{X} \rangle \rangle \langle \langle X X \\ & \frac{1}{4} X Y \rangle \rangle \langle \langle X Y \\ & \frac{1}{4} Y X \rangle \rangle \langle \langle Y X \\ & \frac{1}{4} Y Y \rangle \rangle \langle \langle Y Y \end{split}$	$(-1)^{\beta_1+\beta_2}$

TABLE IV. Irreps, projectors, and characters of the Liouville representation of \mathcal{Z}_2 .

i	$\hat{\Pi}_i^{\mathcal{D}}$
0	$\frac{1}{2^N} \mathbb{1} \cdots \mathbb{1} \rangle \rangle \langle \langle \mathbb{1} \cdots \mathbb{1} $
1	$\frac{1}{2^N}\sum_{P\in\mathcal{Z}_N\backslash\{\mathbbm{1}\cdots\mathbbm{1}\}} P\rangle\rangle\langle\langle P $
2	$\frac{1}{2^N}\sum_{P\in\mathcal{P}_N\backslash\mathcal{Z}_N} P\rangle\rangle\langle\langle P $

TABLE V. Irreps and projectors of the Liouville representation of \mathcal{D}_N . For this work, we will not need the characters of \mathcal{D}_N .

$$\operatorname{Tr}_{\mathcal{Z}_{N}}(\hat{\Lambda}) = \frac{1}{2^{N}} \sum_{P \in \mathcal{Z}_{N}} \langle \langle P | \hat{\Lambda} | P \rangle \rangle \tag{A12}$$

$$=2^{N}(1-p_{\rm ND})$$
 (A13)

$$\operatorname{Tr}_{\mathcal{P}_N \setminus \mathcal{Z}_N}(\hat{\Lambda}) = \frac{1}{2^N} \sum_{P \in \mathcal{P}_N \setminus \mathcal{Z}_N} \langle \langle P | \hat{\Lambda} | P \rangle \rangle \tag{A14}$$

$$= (4^N - 2^N)(1 - p_{\rm ND}) - 4^N p_{\rm D} \qquad (A15)$$

Rearranging this gives formulas for the dephasing and non-dephasing probabilities

$$p_{\rm D} = \frac{1}{4^N} \left[\left(2^N - 1 \right) \operatorname{Tr}_{\mathcal{Z}_N}(\hat{\Lambda}) - \operatorname{Tr}_{\mathcal{P}_N \setminus \mathcal{Z}_N}(\hat{\Lambda}) \right]$$

$$p_{\rm ND} = 1 - \frac{1}{2^N} \operatorname{Tr}_{\mathcal{Z}_N}(\hat{\Lambda}).$$
 (A16)

Our goal in these RB procedures is to determine the two traces $\operatorname{Tr}_{\mathcal{Z}_N}(\hat{\Lambda})$ and $\operatorname{Tr}_{\mathcal{P}_N \setminus \mathcal{Z}_N}(\hat{\Lambda})$.

4. Deriving CX-dihedral bias RB

The character-weighted survival probability \mathcal{S}_b can be written as

$$S_{b}(n) = \mathop{\mathbb{E}}_{\substack{U_{0} \in \mathcal{P}_{N} \\ U_{1} \cdots U_{n} \in \mathcal{D}_{N}}} \left[\langle \langle E_{b} | \hat{\Lambda}_{M} \hat{\Lambda} \hat{U}_{n+1} \hat{\Lambda} \hat{U}_{N} \cdots \hat{U}_{2} \hat{\Lambda} \hat{U}_{1} \hat{U}_{0} \hat{\Lambda}_{P} | \rho_{b} \rangle \rangle \chi_{b}^{*}(U_{0}) \right]$$
(A17)

where Λ is the error channel associated with elements of \mathcal{D}_N , and we have included unknown preparation and measurement errors Λ_P and Λ_M . We make the standard RB change-of-variables, defining $V_1 := U_1$ and inductively defining $V_m := U_m V_{m-1}$. Note that the expectation value over $\{V_m\}$ is equivalent to the expectation value over $\{U_m\}$.

$$S_{b}(n) = \underset{\substack{U_{0} \in \mathcal{P}_{N} \\ V_{1} \cdots V_{n} \in \mathcal{D}_{N}}}{\mathbb{E}} \left[\langle \langle E_{b} | \hat{\Lambda}_{M} \hat{\Lambda} \hat{V}_{n}^{\dagger} \hat{\Lambda} \hat{V}_{n} \hat{V}_{n-1} \cdots \hat{V}_{2} V_{1}^{\dagger} \hat{\Lambda} \hat{V}_{1} \hat{U}_{0} \hat{\Lambda}_{P} | \rho_{b} \rangle \rangle \chi_{b}^{*}(U_{0}) \right]$$
(A18)

$$= \langle \langle E_b | \hat{\Lambda}_M \hat{\Lambda} \left[\mathop{\mathbb{E}}_{V \in \mathcal{D}_N} \left[\hat{V}^{\dagger} \hat{\Lambda} V \right] \right]^n \mathop{\mathbb{E}}_{U_0 \in \mathcal{P}_N} \left[\hat{U}_0 \chi_b^*(U_0) \right] \hat{\Lambda}_P | \rho_b \rangle \rangle.$$
(A19)

The two expectation values can be evaluated using Facts 1 and 2. First, we note from Eq. A11 that χ_1 is the character function of the irrep of the Liouville representation indexed by $\vec{i}_1 = (1, 1, ..., 1)$, and χ_2 is the character function of the irrep indexed by $\vec{i}_2 = (3, 3, ..., 3)$. Then Fact 1 says that

$$\mathop{\mathbb{E}}_{U_0 \in \mathcal{P}_N} \left[\hat{U}_0 \chi_b^*(U_0) \right] = \hat{\Pi}_{\vec{i}_b}^{\mathcal{P}}$$
(A20)

where the projectors have the explicit form (see Eq. A11)

$$\hat{\Pi}_{\vec{i}b}^{\mathcal{P}} = \begin{cases} \frac{1}{2^{N}} |Z \cdots Z\rangle \rangle \langle \langle Z \cdots Z|, & b = 1\\ \frac{1}{2^{N}} |X \cdots X\rangle \rangle \langle \langle X \cdots X|, & b = 2 \end{cases}$$
(A21)

Second, we note that, since the Liouville representation of the dihedral group is multiplicity-free (all $a_i = 1$), Fact 2 gives

$$\mathbb{E}_{V \in \mathcal{D}_N} \left[\hat{V}^{\dagger} \hat{\Lambda} V \right] = \sum_i \frac{\operatorname{Tr}(\hat{\Pi}_i^{\mathcal{D}} \hat{\Lambda})}{\operatorname{Tr}(\hat{\Pi}_i^{\mathcal{D}})} \hat{\Pi}_i^{\mathcal{D}}.$$
 (A22)

Combining these facts allows us to simplify the survival probability as

$$S_b(n) = \sum_i \langle \langle E_b | \hat{\Lambda}_M \hat{\Lambda} \left[\frac{\operatorname{Tr}(\hat{\Pi}_i^{\mathcal{D}} \hat{\Lambda})}{\operatorname{Tr}(\hat{\Pi}_i^{\mathcal{D}})} \right]^n \hat{\Pi}_i^{\mathcal{D}} \hat{\Pi}_{\vec{i}_b}^{\mathcal{P}} \hat{\Lambda}_P | \rho_b \rangle \rangle.$$
(A23)

We have carefully chosen χ_b to give us a projector $\hat{\Pi}_{\vec{i}_b}^{\mathcal{P}}$ such that $\hat{\Pi}_i^{\mathcal{D}} \hat{\Pi}_{\vec{i}_b}^{\mathcal{P}} = \delta_{i,b} \hat{\Pi}_{\vec{i}_b}$, as can be checked from the formulas for $\hat{\Pi}_i^{\mathcal{D}}$ in Table V. Therefore, all terms in the sum vanish except for i = b, and we can write the final form of $S_b(n)$ as

$$S_{b}(n) = \underbrace{\langle \langle E_{b} | \hat{\Lambda}_{M} \hat{\Lambda} \hat{\Pi}_{\vec{i}_{b}}^{\mathcal{P}} \hat{\Lambda}_{P} | \rho_{b} \rangle \rangle}_{A_{b}} \left[\underbrace{\frac{\operatorname{Tr}(\hat{\Pi}_{b}^{\mathcal{P}} \hat{\Lambda})}{\operatorname{Tr}(\hat{\Pi}_{b}^{\mathcal{P}})}}_{\lambda_{b}} \right]^{n} \quad (A24)$$

where we have defined A_b , λ_b to match the form given in Table I. Note that λ_b only depends on Λ , and all effects of the SPAM errors Λ_P and Λ_M are absorbed in A_b . Note also that provided we have reasonably high-fidelity preparation, gates, and measurements, we can approximate A_b as

$$A_b \approx \langle \langle E_b | \hat{\Pi}_{\vec{i}}^{\mathcal{P}} | \rho_b \rangle \rangle = 1 \tag{A25}$$

where the last equality is found by plugging in the explicit formulas for E_b , $\hat{\Pi}_{\vec{i}_b}^{\mathcal{P}}$, and ρ_b given in Eq. A21 and Table I. Therefore, the prefactor in front of λ_b^n is large, and we can accurately fit λ_b .

To finish, we note that plugging in the explicit formulas from Table V for $\hat{\Pi}^{\mathcal{D}}_b$ gives

$$\lambda_1 = \frac{\operatorname{Tr}_{\mathcal{Z}}(\hat{\Lambda}) - 1}{2^N - 1} \tag{A26}$$

$$\lambda_2 = \frac{\operatorname{Tr}_{\mathcal{P}\backslash\mathcal{Z}}(\Lambda)}{4^N - 2^N}.$$
 (A27)

Plugging this into Eq. A16 for $p_{\rm D}$ and $p_{\rm ND}$ gives the estimates in Eq. 16, as desired.

5. Deriving interleaved bias RB

a. Deriving the survival probabilities for $b = 0 \pm$ and $b = 1 \pm$

We begin by considering the survival probability S_{0+} . For convenience, we define the operator \hat{M}_{0+} by

$$\hat{M}_{0+} := \frac{1}{2} \hat{\Pi}_0^{\mathcal{Z}} (\hat{C}\hat{\Lambda} + \hat{C}'\hat{\Lambda}') \hat{\Pi}_0^{\mathcal{Z}}.$$
 (A28)

Note that $\hat{\Pi}_0^{\mathbb{Z}}$ is a rank-4 projector, so \hat{M}_{0+} is a rank-4 operator in the general case.

We now evaluate the survival probability $S_{0+}(n)$ in terms of M_{0+} . We note that Fact 1 implies that $\mathbb{E}_{U \in \mathbb{Z}_2}[\hat{U}] = \hat{\Pi}_0^{\mathbb{Z}}$. We then have

$$S_{0+}(n) = \mathbb{E}_{\substack{U_1 \cdots U_{n+1} \in \mathcal{Z}_2 \\ C_1, \dots, C_n \in \{C, C'\}}} \left[\langle \langle E_{0+} | \hat{\Lambda}_M \hat{\Lambda}_G \hat{U}_{n+1} \hat{C}_n \hat{\Lambda}_n \hat{U}_n \hat{C}_{n-1} \hat{\Lambda}_{n-1} \hat{U}_{n-1} \cdots \hat{U}_2 \hat{C}_1 \hat{\Lambda}_1 \hat{U}_1 \hat{\Lambda}_P | \rho_{0+} \rangle \rangle \right]$$
(A29)

$$= \underset{C_1,\dots,C_n \in \{C,C'\}}{\mathbb{E}} \left[\langle \langle E_{0+} | \hat{\Lambda}_M \hat{\Lambda}_G \underbrace{\hat{\Pi}_0^{\mathcal{Z}} \hat{C}_n \hat{\Lambda}_n \hat{\Pi}_0^{\mathcal{Z}}}_{\hat{M}_{0+}} \underbrace{\hat{\Pi}_0^{\mathcal{Z}} \hat{C}_{n-1} \hat{\Lambda}_{n-1} \hat{\Pi}_0^{\mathcal{Z}}}_{\hat{M}_{0+}} \cdots \underbrace{\hat{\Pi}_0^{\mathcal{Z}} \hat{C}_1 \hat{\Lambda}_1 \hat{\Pi}_0^{\mathcal{Z}}}_{\hat{M}_{0+}} \hat{\Lambda}_P | \rho_{0+} \rangle \rangle \right]$$
(A30)

$$= \langle \langle E_{0+} | \hat{\Lambda}_M \hat{\Lambda}_G \hat{M}_{0+}^n \hat{\Lambda}_P | \rho_{0+} \rangle \rangle \tag{A31}$$

Given that \hat{M}_{0+} has rank 4, we can expand it in terms of its eigenvalues μ_j and corresponding left and right eigenvectors $|\phi_j\rangle\rangle$ and $|\psi_j\rangle\rangle$ as

$$\hat{M}_{0+} = \sum_{j=1}^{4} |\psi_j\rangle\rangle\mu_j\langle\langle\phi_j|$$
(A32)

where we have normalized our eigenvectors so that $\langle \langle \psi_j | \phi_k \rangle \rangle = \delta_{j,k}$. In this case, our survival probability becomes

$$S_{0+}(n) = \sum_{j=1}^{4} \langle \langle E_{0+} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{0+} \rangle \rangle \mu_j^n.$$
(A33)

Note that again the eigenvalues μ_j depend only on the gate errors Λ and Λ' , and not on the SPAM errors Λ_P and Λ_M .

In this form, the survival probability is the sum of four exponential decays, which is infeasible to fit to experimental data. However, in the case of high-fidelity gates, we can show that only two exponential decays are relevant using perturbation theory in $\delta \hat{\Lambda} := (\hat{\Lambda} - \mathbb{1})$ and $\delta \hat{\Lambda}' := (\hat{\Lambda}' - \mathbb{1})$. For perfect gates with $\delta \hat{\Lambda} = \delta \hat{\Lambda}' = 0$, \hat{M}_{0+} has eigenvalues and eigenvectors given by

$$|\psi_1^0\rangle\rangle = |\phi_1^0\rangle\rangle = \frac{1}{2}|\mathbb{1}\mathbb{1}\rangle\rangle \qquad \qquad \mu_1^0 = 1 \qquad (A34)$$

$$|\psi_2^0\rangle\rangle = |\phi_2^0\rangle\rangle = \frac{1}{2}|Z\mathbb{1}\rangle\rangle \qquad \qquad \mu_2^0 = 1 \qquad (A35)$$

$$|\psi_3^0\rangle\rangle = |\phi_3^0\rangle\rangle = \frac{1}{2}|\mathbb{1}Z\rangle\rangle \qquad \qquad \mu_3^0 = 0 \qquad (A36)$$

$$|\psi_4^0\rangle\rangle = |\phi_4^0\rangle\rangle = \frac{1}{2}|ZZ\rangle\rangle \qquad \qquad \mu_4^0 = 0. \tag{A37}$$

Then to first order in $\delta \hat{\Lambda}$ and $\delta \hat{\Lambda}'$, we have

$$\mu_1 = 1 \tag{A38}$$

$$\mu_2 = \frac{1}{4} \left\langle \left\langle Z \mathbb{1} \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| Z \mathbb{1} \right\rangle \right\rangle + O(\delta^2) \tag{A39}$$

$$\mu_3 = O(\delta) \tag{A40}$$

$$\mu_4 = O(\delta). \tag{A41}$$

We therefore see that one eigenvalue is always 1, and that we may neglect the eigenvalues μ_3 and μ_4 , since their contribution to the survival probability $S_{0+}(n)$ is $O(\delta^n)$. We can therefore fit $S_{0+}(n)$ to a single exponential decay plus a constant. In the notation of Table II, μ_2 corresponds to λ_{0+} .

From Eq. A33, the prefactor A_{0+} in front of the exponential decay is given by

$$A_{0+} = \langle \langle E_{0+} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_2 \rangle \rangle \langle \langle \phi_2 | \hat{\Lambda}_P | \rho_{0+} \rangle \rangle \tag{A42}$$

We can estimate the value of A_{0+} by assuming $\hat{\Lambda}_G, \hat{\Lambda}_P, \hat{\Lambda}_M \approx 1$ and evaluating Eq. A42 explicitly, provided we can estimate the eigenvectors $|\psi_2\rangle\rangle, |\phi_2\rangle\rangle$. Since the eigenvectors at $\delta = 0$ (Eqs. A34 and A35) are degenerate, we must use degenerate perturbation theory to find the eigenvectors at $\delta \neq 0$. This means that $|\psi_2\rangle\rangle$ and $|\phi_2\rangle\rangle$ will in general be (up to $O(\delta)$) some linear combination the $\delta = 0$ eigenvectors with eigenvalue 1. Specifically, we have

$$|\psi_2\rangle\rangle \approx \frac{1}{2}|Z\mathbb{1}\rangle\rangle$$
 (A43)

$$|\phi_2\rangle\rangle \approx \frac{1}{2}|Z\mathbb{1}\rangle\rangle + \frac{\gamma}{2}|\mathbb{1}\mathbb{1}\rangle\rangle$$
 (A44)

where γ is some constant determined by the specific perturbations $\delta\Lambda$ and $\delta\Lambda'$, and the overall form is restricted by the normalization condition $\langle\langle\phi_i|\psi_j\rangle\rangle = \delta_{i,j}$ and the fact that $|\phi_1\rangle\rangle = \frac{1}{2}|\mathbb{11}\rangle$ is a right-eigenvector with eigenvalue 1 for any trace-preserving map. Using these eigenvectors to evaluate Eq. A42 then gives $A_{0+} \approx 1$, so that we may accurately fit λ_{0+} .

The case of b = 0- and $b = 1\pm$ are similar. We define

$$\hat{M}_{0-} := \frac{1}{2} \hat{\Pi}_0 (\hat{C} \hat{\Lambda} - \hat{C}' \hat{\Lambda}') \hat{\Pi}_0$$
 (A45)

$$\hat{M}_{1+} := \frac{1}{2} \hat{\Pi}_1 (\hat{C}\hat{\Lambda} + \hat{C}'\hat{\Lambda}') \hat{\Pi}_1$$
 (A46)

$$\hat{M}_{1-} := \frac{1}{2} \hat{\Pi}_1 (\hat{C}\hat{\Lambda} - \hat{C}'\hat{\Lambda}') \hat{\Pi}_1$$
 (A47)

and repeat the above analysis to see that we have for b = 0- and b = 1-

$$S_b(n) = \sum_{j=1}^4 \langle \langle E_b | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_b \rangle \rangle \mu_j^n, \quad (A48)$$

or for b = 1 +

$$S_{1+}(n) = \frac{1}{2} \sum_{j=1}^{4} \left(\langle \langle E_{1+}^{(1)} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{1+}^{(1)} \rangle \right)$$

$$(A49)$$

$$+ \langle \langle E_{1+}^{(2)} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{1+}^{(2)} \rangle \rangle \right) \mu_j^n,$$

where μ_j , $|\psi_j\rangle\rangle$, and $|\phi_j\rangle\rangle$ are the eigenvalues and eigenvectors of the corresponding \hat{M}_b operator. Performing perturbation theory, we find that the unperturbed \hat{M}_{0-} and \hat{M}_{1-} have eigenvalues $\{\mu_j^0\} = \{1, -1, 0, 0\}$, while the unperturbed \hat{M}_{1+} has eigenvalues $\{\mu_j^0\} = \{1, 1, 0, 0\}$. We label the two largest-magnitude eigenvalues by λ_b and κ_b , and neglect the remaining eigenvalues that are $O(\delta)$. Working to first order in $\delta\Lambda$ and $\delta\Lambda'$, we find

$$\lambda_{0-} - \kappa_{0-} = \frac{1}{4} \left[\left\langle \left\langle \mathbb{1}Z \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| \mathbb{1}Z \right\rangle \right\rangle + \left\langle \left\langle ZZ \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| ZZ \right\rangle \right\rangle \right]$$
(A50)

$$\lambda_{1+} + \kappa_{1+} = \frac{1}{4} \left[\left\langle \left\langle \mathbb{1}X \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| \mathbb{1}X \right\rangle \right\rangle + \left\langle \left\langle ZX \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| ZX \right\rangle \right\rangle \right] \quad (A51)$$

$$\lambda_{1-} - \kappa_{1-} = \frac{1}{4} \left[\left\langle \left\langle \mathbb{1}Y \left| \frac{\Lambda + \Lambda'}{2} \right| \mathbb{1}Y \right\rangle \right\rangle + \left\langle \left\langle ZY \left| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \right| ZY \right\rangle \right\rangle \right]. \quad (A52)$$

In total, we've demonstrated that for these values of b, we can fit $S_b(n)$ to the sum of two exponential decays, as given in Table 23. We can again estimate the prefactors A_b and B_b in front of the decays by assuming $\hat{\Lambda}_G$, $\hat{\Lambda}_P$, $\hat{\Lambda}_M \approx 1$ and evaluating Eqs. A48 and A49. In the case of b = 1+ we again need to use degenerate perturbation theory, while the cases of b = 0- and b = 1- are non-degenerate. We find that $A_b, B_b \approx 1/2$, so that we can reliably fit both decay curves. Note that the only reason for averaging over two initial states and final measurements for b_{1+} is to ensure that $A_{1+}, B_{1+} \approx 1/2$.

b. Deriving the survival probabilities for $b = 2\pm$

We begin by considering the survival probability S_{2+} . For convenience, we define the operator \hat{M}_{2+} by

$$\hat{M}_{2+} := \frac{1}{4} \hat{\Pi}_2^{\mathcal{Z}} (\hat{C}\hat{\Lambda} + \hat{C}'\hat{\Lambda}') \hat{\Pi}_3^{\mathcal{Z}} (\hat{C}\hat{\Lambda} + \hat{C}'\hat{\Lambda}') \hat{\Pi}_2^{\mathcal{Z}}.$$
 (A53)

Note that $\hat{\Pi}_2^{\mathbb{Z}}$ is a rank-4 projector, so \hat{M}_{2+} is a rank-4 operator in the general case.

We now evaluate the survival probability $S_{2+}(n)$ in terms of M_{2+} . We note that Fact 1 implies that $\mathbb{E}_{U \in \mathbb{Z}_2}[\hat{U}\chi_{2+}^*(U)] = \hat{\Pi}_2^{\mathbb{Z}}$ and $\mathbb{E}_{U \in \mathbb{Z}_2}[\hat{U}\chi_{2-}^*(U)] = \hat{\Pi}_3^{\mathbb{Z}}$ (re the expression for $\chi_{2\pm}$ in Table II to the characters of the Liouville representation of \mathbb{Z}_2 in Table IV). We then have

$$S_{2+}(n) = \underset{\substack{(a)=(1),(2)\\U_1\cdots U_{2n+1}\in G\\C_1\dots\dots C_{2n}\in\{C,C'\}}}{\mathbb{E}} \left[\langle \langle E_{2+}^{(a)} | \hat{\Lambda}_M \hat{\Lambda}_G \hat{U}_{2n+1} \hat{C}_{2n} \hat{\Lambda}_{2n} \hat{U}_{2n} \hat{C}_{2n-1} \hat{\Lambda}_{n-1} \hat{U}_{n-1} \cdots \hat{U}_3 \hat{C}_2 \hat{\Lambda}_2 \hat{U}_2 \hat{C}_1 \hat{\Lambda}_1 \hat{U}_1 \hat{\Lambda}_P | \rho_{2+}^{(a)} \rangle \chi_2^*(\{U_i\}) \right]$$

$$= \underbrace{\mathbb{E}}_{\substack{(a)=(1),(2)\\C_1,\dots,C_{2n}\in\{C,C'\}}} \left[\langle \langle E_{2+}^{(a)} | \hat{\Lambda}_M \hat{\Lambda}_G \underbrace{\hat{\Pi}_2^{\mathcal{Z}} \hat{C}_{2n} \hat{\Lambda}_{2n} \hat{\Pi}_3^{\mathcal{Z}} \hat{C}_{2n-1} \hat{\Lambda}_{n-1} \hat{\Pi}_2^{\mathcal{Z}}}_{\hat{M}_{2+}} \cdots \underbrace{\hat{\Pi}_2^{\mathcal{Z}} \hat{C}_2 \hat{\Lambda}_2 \hat{\Pi}_3^{\mathcal{Z}} \hat{C}_1 \hat{\Lambda}_1 \hat{\Pi}_2^{\mathcal{Z}}}_{\hat{M}_{2+}} \hat{\Lambda}_P | \rho_{2+}^{(a)} \rangle \rangle \right]$$
(A55)

$$= \frac{1}{2} \left(\langle \langle E_{2+}^{(1)} | \hat{\Lambda}_M \hat{\Lambda}_G \hat{M}_{2+}^n \hat{\Lambda}_P | \rho_{2+}^{(1)} \rangle \rangle + \langle \langle E_{2+}^{(2)} | \hat{\Lambda}_M \hat{M}_{2+}^n \hat{\Lambda}_P | \rho_{2+}^{(2)} \rangle \rangle \right)$$
(A56)

$$=\sum_{j=1}^{4} \frac{1}{2} \left(\langle \langle E_{2+}^{(1)} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{2+}^{(1)} \rangle \rangle + \langle \langle E_{2+}^{(2)} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{2+}^{(2)} \rangle \right) \mu_j^n$$
(A57)

where in the last line, μ_j , $|\psi_j\rangle\rangle$, and $|\phi_j\rangle\rangle$ denote the eigenvalues and eigenvectors of \hat{M}_{2+} .

theory. When $\delta \hat{\Lambda} = \delta \hat{\Lambda}' = 0$, \hat{M}_{2+} has eigenvalues and

We again simplify this expression through perturbation

eigenvectors given by

$$|\psi_1^0\rangle\rangle = |\phi_1^0\rangle\rangle = \frac{1}{2}|X\mathbb{1}\rangle\rangle \qquad \qquad \mu_1^0 = 1 \qquad (A58)$$

$$|\psi_2^0\rangle\rangle = |\phi_2^0\rangle\rangle = \frac{1}{2}|Y1\rangle\rangle \qquad \qquad \mu_2^0 = 1 \qquad (A59)$$

$$|\psi_3^0\rangle\rangle = |\phi_3^0\rangle\rangle = \frac{1}{2}|XZ\rangle\rangle \qquad \mu_3^0 = 0 \tag{A60}$$

$$|\psi_4^0\rangle\rangle = |\phi_4^0\rangle\rangle = \frac{1}{2}|YZ\rangle\rangle \qquad \qquad \mu_4^0 = 0. \tag{A61}$$

Then to first order in $\delta \hat{\Lambda}$ and $\delta \hat{\Lambda}'$, the eigenvalues satisfy

$$\mu_{1} + \mu_{2} = \frac{1}{4} \left[\left\langle \left\langle X \mathbb{1} \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| X \mathbb{1} \right\rangle \right\rangle + \left\langle \left\langle X X \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| X X \right\rangle \right\rangle + \left\langle \left\langle Y \mathbb{1} \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| Y \mathbb{1} \right\rangle \right\rangle$$
(A62)

$$+\left\langle\left\langle YX\Big|\frac{\hat{\Lambda}+\hat{\Lambda}'}{2}\Big|YX\right\rangle\right\rangle\right] - 2 + O(\delta^2)$$

$$\mu_3 = O(\delta) \tag{A63}$$

$$\mu_4 = O(\delta). \tag{A64}$$

We neglect μ_3 and μ_4 , and find S_{2+} is the sum of two exponential decays. In the notation of Table II, μ_1 corresponds to λ_{2+} and μ_2 corresponds to κ_{2+} .

Similarly, for $S_{2-}(b)$, we define

$$\hat{M}_{2-} := \frac{1}{4} \hat{\Pi}_2 (\hat{C}\hat{\Lambda} - \hat{C}'\hat{\Lambda}') \hat{\Pi}_3 (\hat{C}\hat{\Lambda} - \hat{C}'\hat{\Lambda}') \hat{\Pi}_2 \qquad (A65)$$

in terms of which we can write

$$S_{2-}(n) = \frac{1}{2} \sum_{j=1}^{4} \left(\langle \langle E_{2-}^{(1)} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{2-}^{(1)} \rangle \right)$$

$$(A66)$$

$$+ \langle \langle E_{2-}^{(2)} | \hat{\Lambda}_M \hat{\Lambda}_G | \psi_j \rangle \rangle \langle \langle \phi_j | \hat{\Lambda}_P | \rho_{2-}^{(2)} \rangle \right) \mu_j^n,$$

where now μ_j , $|\psi_j\rangle\rangle$, and $|\phi_j\rangle\rangle$ denote the eigenvalues and eigenvectors of \hat{M}_{2-} .

We again use perturbation theory; the unperturbed eigenvalues and eigenvectors are

$$|\psi_1^0\rangle\rangle = |\phi_1^0\rangle\rangle = \frac{1}{2}|XZ\rangle\rangle \qquad \qquad \mu_1^0 = 1 \qquad (A67)$$

$$|\psi_2^0\rangle\rangle = |\phi_2^0\rangle\rangle = \frac{1}{2}|YZ\rangle\rangle \qquad \mu_2^0 = 1 \qquad (A68)$$

$$|\psi_3^0\rangle\rangle = |\phi_3^0\rangle\rangle = \frac{1}{2}|X\mathbb{1}\rangle\rangle \qquad \qquad \mu_3^0 = 0 \qquad (A69)$$

$$|\psi_4^0\rangle\rangle = |\phi_4^0\rangle\rangle = \frac{1}{2}|Y\mathbb{1}\rangle\rangle \qquad \qquad \mu_4^0 = 0 \qquad (A70)$$

while to first order in $\delta \hat{\Lambda}$ and $\delta \hat{\Lambda}'$, the eigenvectors satisfy

$$\mu_{1} + \mu_{2} = \frac{1}{4} \left[\left\langle \left\langle XZ \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| XZ \right\rangle \right\rangle + \left\langle \left\langle YY \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| YY \right\rangle \right\rangle$$
(A71)
+ $\left\langle \left\langle YZ \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| YZ \right\rangle \right\rangle$
+ $\left\langle \left\langle XY \middle| \frac{\hat{\Lambda} + \hat{\Lambda}'}{2} \middle| XY \right\rangle \right\rangle \right] - 2 + O(\delta^{2})$
$$\mu_{3} = O(\delta)$$
(A72)
$$\mu_{4} = O(\delta).$$
(A73)

We again neglect μ_3 and μ_4 , and find S_{2-} is also the sum of two exponential decays. In the notation of Table II, μ_1 corresponds to λ_{2-} and μ_2 corresponds to κ_{2-} .

Finally, we can estimate the prefactors $A_{2\pm}$ and $B_{2\pm}$ by assuming $\hat{\Lambda}_G$, $\hat{\Lambda}_P$, $\hat{\Lambda}_M \approx 1$ and evaluating Eqs. A57 and A66. In both cases, we must use degenerate perturbation theory to find the approximate eigenvectors. We again find A_b , $B_b \approx 1/2$.

c. Finding the dephasing and non-dephasing probabilities

We can combine Eqs. A39 and A50 to evaluate the trace over \mathcal{Z}_2 as

$$\operatorname{Tr}_{\mathcal{Z}_2}\left(\frac{\hat{\Lambda}+\hat{\Lambda}'}{2}\right) = 1 + \lambda_{0+} + \lambda_{0-} - \kappa_{0-} \qquad (A74)$$

and we can combine Eqs. A51, A52, A62, and A71 to evaluate the trace over $\mathcal{P}_2 \setminus \mathcal{Z}_2$ as

$$\operatorname{Tr}_{\mathcal{P}_{2}\backslash\mathcal{Z}_{2}}\left(\frac{\hat{\Lambda}+\hat{\Lambda}'}{2}\right) = \lambda_{1+} + \kappa_{1+} + \lambda_{1-} - \kappa_{1-} + \lambda_{2+} + \kappa_{2+} + \lambda_{2-} + \kappa_{2-} + 4.$$
(A75)

Plugging Eqs. A74 and A75 into Eq. A16 for $p_{\rm D}$ and $p_{\rm ND}$ gives Eq. 23 as desired.

Appendix B: Dephasing/non-dephasing probabilities of a composite channel

Given two quantum channels

$$\Lambda_{A}(\rho) = \sum_{\substack{\vec{\alpha}_{1},\vec{\beta}_{1}\\\vec{\alpha}_{2},\vec{\beta}_{2}}} \chi^{A}_{\vec{\alpha}_{1}\vec{\beta}_{1},\vec{\alpha}_{2}\vec{\beta}_{2}} \mathbf{X}(\vec{\alpha}_{1})\mathbf{Z}(\vec{\beta}_{1})\rho\mathbf{Z}(\vec{\beta}_{2})\mathbf{X}(\vec{\alpha}_{2}) \quad (B1)$$

$$\Lambda_{B}(\rho) = \sum_{\substack{\vec{\alpha}_{1},\vec{\beta}_{1}\\\vec{\alpha}_{2},\vec{\beta}_{2}}} \chi^{B}_{\vec{\alpha}_{1}\vec{\beta}_{1},\vec{\alpha}_{2}\vec{\beta}_{2}} \mathbf{X}(\vec{\alpha}_{1})\mathbf{Z}(\vec{\beta}_{1})\rho\mathbf{Z}(\vec{\beta}_{2})\mathbf{X}(\vec{\alpha}_{2}) \quad (B2)$$

with dephasing and non-dephasing probabilities $p_{\rm D}^A$, $p_{\rm ND}^A$, $p_{\rm D}^B$, and $p_{\rm ND}^B$, we want to find bounds on the dephasing/non-dephasing probabilities of the combined channel ($\Lambda_A \circ \Lambda_B$). We will denote the combined error probabilities by simply $p_{\rm D}$ and $p_{\rm ND}$.

1. Finding the non-dephasing error probability

From the definition of $p_{\rm ND}$, Eq. 4, we have

$$p_{\rm ND} = \sum_{\substack{\vec{\alpha}_1 + \vec{\gamma}_1 = \vec{\alpha}_2 + \vec{\gamma}_2 \neq \vec{0} \\ \vec{\beta}_1 + \vec{\delta}_1 = \vec{\beta}_2 + \vec{\delta}_2}} (-1)^{\vec{\beta}_1 \cdot \vec{\gamma}_1 + \vec{\beta}_2 \cdot \vec{\gamma}_2} \chi^A_{\vec{\alpha}_1 \vec{\beta}_1, \vec{\alpha}_2 \vec{\beta}_2} \chi^B_{\vec{\gamma}_1 \vec{\delta}_1, \vec{\gamma}_2 \vec{\delta}_2}.$$
(B3)

For legibility in what follows, we will denote the diagonal elements $\chi_{\vec{\alpha}\vec{\beta},\vec{\alpha}\vec{\beta}}$ of the χ -matrices by simply $\chi_{\vec{\alpha}\vec{\beta}}$. Note that the complete-positivity of Λ_A and Λ_B requires that the χ -matrices are positive semidefinite, which in turn implies $|\chi_{\vec{\alpha}_1\vec{\beta}_1,\vec{\alpha}_2\vec{\beta}_2}| \leq \sqrt{\chi_{\vec{\alpha}_1\vec{\beta}_1}\chi_{\vec{\alpha}_2\vec{\beta}_2}}$ [33], a fact we will use repeatedly. We will also repeatedly use two elementary inequalities, both of which are versions of the Cauchy-Schwarz inequality:

$$\sum_{i} \sqrt{a_i b_i} \le \sqrt{\sum_{i} a_i} \sqrt{\sum_{i} b_i} \tag{B4}$$

$$\sum_{i=1}^{I} \sqrt{a_i} \le \sqrt{I} \sqrt{\sum_i a_i}.$$
 (B5)

The condition $\vec{\alpha}_1 + \vec{\gamma}_1 = \vec{\alpha}_2 + \vec{\gamma}_2 \neq \vec{0}$ implies at most two of $\{\vec{\alpha}_1, \vec{\alpha}_2, \vec{\gamma}_1, \vec{\gamma}_2\}$ can be equal to $\vec{0}$. We thus divide the terms in Eq. B3 into several subsets.

- 1. Terms with $\vec{\alpha}_1 = \vec{\alpha}_2 = \vec{0}$ (and thus $\vec{\gamma}_1 = \vec{\gamma}_2 \neq \vec{0}$).
- 2. Terms with $\vec{\gamma}_1 = \vec{\gamma}_2 = \vec{0}$ (and thus $\vec{\alpha}_1 = \vec{\alpha}_2 \neq \vec{0}$).
- 3. Terms with $\vec{\alpha}_1 = \vec{\gamma}_2 = \vec{0}$ (and thus $\vec{\alpha}_2 = \vec{\gamma}_1 \neq \vec{0}$).
- 4. Terms with $\vec{\alpha}_2 = \vec{\gamma}_1 = \vec{0}$ (and thus $\vec{\alpha}_1 = \vec{\gamma}_2 \neq \vec{0}$).
- 5. Terms with $\vec{\alpha}_1 = \vec{0}$ and $\vec{\alpha}_2, \vec{\gamma}_1, \vec{\gamma}_2 \neq \vec{0}$ (and thus $\vec{\gamma}_2 = \vec{\alpha}_2 + \vec{\gamma}_1$).
- 6. Terms with $\vec{\alpha}_2 = \vec{0}$ and $\vec{\alpha}_1, \vec{\gamma}_1, \vec{\gamma}_2 \neq \vec{0}$ (and $\vec{\gamma}_2 = \vec{\alpha}_1 + \vec{\gamma}_1$).
- 7. Terms with $\vec{\gamma}_1 = \vec{0}$ and $\vec{\alpha}_1, \vec{\alpha}_2, \vec{\gamma}_2 \neq \vec{0}$ (and $\vec{\gamma}_2 = \vec{\alpha}_1 + \vec{\alpha}_2$).
- 8. Terms with $\vec{\gamma}_2 = \vec{0}$ and $\vec{\alpha}_1, \vec{\alpha}_2, \vec{\gamma}_1 \neq \vec{0}$ (and $\vec{\gamma}_1 = \vec{\alpha}_1 + \vec{\alpha}_2$).
- 9. Terms with $\vec{\alpha}_1, \vec{\alpha}_2, \vec{\gamma}_1, \vec{\gamma}_2 \neq \vec{0}$ (and $\vec{\gamma}_2 = \vec{\alpha}_1 + \vec{\alpha}_2 + \vec{\gamma}_1$).

Let's take each of these in turn. We have

$$\begin{aligned} (1) &= \sum_{\substack{\vec{\gamma}_{1} \neq \vec{0} \\ \vec{\beta}_{1}, \vec{\beta}_{2}, \vec{\delta}_{1}}} (-1)^{(\vec{\beta}_{1} + \vec{\beta}_{2}) \cdot \vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{\beta}_{1}, \vec{0}\vec{\beta}_{2}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}, \vec{\gamma}_{1}(\vec{\beta}_{1} + \vec{\beta}_{2} + \vec{\delta}_{1})} \\ &= \sum_{\substack{\vec{\gamma}_{1} \neq \vec{0} \\ \vec{\delta}_{1}}} \chi^{A}_{\vec{0}\vec{0}, \vec{0}\vec{0}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}, \vec{\gamma}_{1}\vec{\delta}_{1}} \\ &+ \sum_{\vec{\beta}_{2}, \vec{\gamma}_{1} \neq \vec{0}} (-1)^{\vec{\beta}_{2} \cdot \vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{0}, \vec{0}\vec{\beta}_{2}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}, \vec{\gamma}_{1}(\vec{\beta}_{2} + \vec{\delta}_{1})} \\ &+ \sum_{\vec{\beta}_{1}, \vec{\gamma}_{1} \neq \vec{0}} (-1)^{\vec{\beta}_{1} \cdot \vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{\beta}_{1}, \vec{0}\vec{0}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}, \vec{\gamma}_{1}(\vec{\beta}_{1} + \vec{\delta}_{1})} \\ &+ \sum_{\vec{\beta}_{1}, \vec{\beta}_{2}, \vec{\gamma}_{1} \neq \vec{0}} (-1)^{(\vec{\beta}_{1} + \vec{\beta}_{2}) \cdot \vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{\beta}_{1}, \vec{0}\vec{\beta}_{2}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}, \vec{\gamma}_{1}(\vec{\beta}_{1} + \vec{\beta}_{2} + \vec{\delta}_{1})} \end{aligned} \tag{B7}$$

The first term is equal to p_{ND}^B , up to irrelevant higherorder terms. We can bound the magnitude of the remaining terms:

$$\begin{vmatrix} \sum_{\vec{\beta}_{2},\vec{\gamma}_{1}\neq\vec{0}} (-1)^{\vec{\beta}_{2}\cdot\vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{0},\vec{0}\vec{\beta}_{2}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1},\vec{\gamma}_{1}(\vec{\beta}_{2}+\vec{\delta}_{1})} \\ \leq \sum_{\vec{\beta}_{2}\neq\vec{0}} \sqrt{\chi^{A}_{\vec{0}\vec{0}} \chi^{A}_{\vec{0}\vec{\beta}_{2}}} \sum_{\substack{\gamma_{1}\neq\vec{0}\\\vec{\delta}_{1}}} \sqrt{\chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}} \chi^{B}_{\vec{\gamma}_{1}(\vec{\beta}_{2}+\vec{\delta}_{1})}} \quad (B8) \end{aligned}$$

$$\leq \sum_{\vec{\beta}_2 \neq \vec{0}} \sqrt{\chi^A_{\vec{0}\vec{\beta}_2}} \sum_{\substack{\gamma_1 \neq \vec{0} \\ \vec{\delta}_1}} \chi^B_{\vec{\gamma}_1 \vec{\delta}_1} \tag{B9}$$

$$\leq 2^{N/2} \sqrt{\sum_{\beta_1 \neq \vec{0}} \chi^A_{\vec{0}\vec{\beta}_2}} p^B_{\rm ND} \tag{B10}$$

$$=2^{N/2}\sqrt{p_{\rm D}^A}p_{\rm ND}^B \tag{B11}$$

$$\begin{vmatrix} \sum_{\vec{\beta}_{1},\vec{\gamma}_{1}\neq\vec{0}} (-1)^{\vec{\beta}_{1}\cdot\vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{\beta}_{1},\vec{0}\vec{0}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1},\vec{\gamma}_{1}(\vec{\beta}_{1}+\vec{\delta}_{1})} \\ \leq 2^{N/2} \sqrt{p_{\rm D}^{A}} p_{\rm ND}^{B}$$
(B12)

$$\sum_{\vec{\beta}_{1},\vec{\beta}_{2},\vec{\gamma}_{1}\neq\vec{0}} (-1)^{(\vec{\beta}_{1}+\vec{\beta}_{2})\cdot\vec{\gamma}_{1}} \chi^{A}_{\vec{0}\vec{\beta}_{1},\vec{0}\vec{\beta}_{2}} \chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1},\vec{\gamma}_{1}(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{1})} \bigg|$$

$$\leq \sum_{\vec{\beta}_{1},\vec{\beta}_{2}\neq\vec{0}} \sqrt{\chi^{A}_{\vec{0}\vec{\beta}_{1}}\chi^{A}_{\vec{0}\vec{\beta}_{2}}} \sum_{\substack{\vec{\gamma}_{1}\neq\vec{0}\\\vec{\delta}_{1}}} \sqrt{\chi^{B}_{\vec{\gamma}_{1}\vec{\delta}_{1}}\chi^{B}_{\vec{\gamma}_{1}(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{1})}} \tag{B13}$$

$$\leq \sum_{\vec{\beta}_1, \vec{\beta}_2 \neq \vec{0}} \sqrt{\chi^A_{\vec{0}\vec{\beta}_1} \chi^A_{\vec{0}\vec{\beta}_2}} \sum_{\substack{\vec{\gamma}_1 \neq \vec{0} \\ \vec{\delta}_1}} \chi^B_{\vec{\gamma}_1 \vec{\delta}_1}$$
(B14)

$$\leq 2^N \sum_{\vec{\beta}_1 \neq \vec{0}} \chi^A_{\vec{0}\vec{\beta}_1} p^B_{\rm ND} \tag{B15}$$

$$=2^{N}p_{\rm D}^{A}p_{\rm ND}^{B} \tag{B16}$$

Thus in total, we have

$$|(1) - p_{\rm ND}^B| \le 2^{N/2+1} \sqrt{p_{\rm D}^A} p_{\rm ND}^B + 2^N p_{\rm D}^A p_{\rm ND}^B$$
 (B17)

By symmetry, we similarly have

$$|(2) - p_{\rm ND}^A| \le 2^{N/2+1} p_{\rm ND}^A \sqrt{p_{\rm D}^B} + 2^N p_{\rm ND}^A p_{\rm D}^B$$
 (B18)

The remaining terms are only higher-order corrections, and so we bound their magnitudes:

$$\begin{aligned} |(3)| &\leq \sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{1},\vec{\beta}_{2},\vec{\delta}_{2}}} \sqrt{\chi_{\vec{0}\vec{\beta}_{1}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{\vec{\alpha}_{2}(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}^{B}\chi_{\vec{0}\vec{\delta}_{2}}^{B}} \quad (B19) \\ &= \sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{B}\chi_{\vec{0}\vec{0}}^{B}} \\ &+ \sum_{\substack{\vec{\alpha}_{2},\vec{\beta}_{1}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{\vec{\alpha}_{2}(\vec{\beta}_{1}+\vec{\beta}_{2})}^{B}\chi_{\vec{0}\vec{0}}^{B}} \\ &+ \sum_{\substack{\vec{\alpha}_{2},\vec{\delta}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{\vec{\alpha}_{2}(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}^{B}\chi_{\vec{0}\vec{\delta}_{2}}^{B}} \\ &+ \sum_{\substack{\vec{\alpha}_{2},\vec{\delta}_{1},\vec{\delta}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{0}\vec{\beta}_{1}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{\vec{\alpha}_{2}(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}^{B}\chi_{\vec{0}\vec{\delta}_{2}}^{B}} \quad (B20) \end{aligned}$$

We can bound each of these four terms as

$$\sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A} \chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A} \chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{B} \chi_{\vec{0}\vec{0}}^{B}}$$

$$\leq \sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A} \chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{B}}$$
(B21)

$$\leq \sqrt{\sum_{\substack{\vec{\alpha}_2 \neq \vec{0} \\ \vec{\beta}_2}} \chi^A_{\vec{\alpha}_2 \vec{\beta}_2}} \sqrt{\sum_{\substack{\vec{\alpha}_2 \neq \vec{0} \\ \vec{\beta}_2}} \chi^B_{\vec{\alpha}_2 \vec{\beta}_2}} \tag{B22}$$

$$=\sqrt{p_{\rm ND}^A p_{\rm ND}^B} \tag{B23}$$

 $\sum_{\vec{\alpha}_{2},\vec{\beta}_{1}\neq\vec{0}} \sqrt{\chi^{A}_{\vec{0}\vec{\beta}_{1}}\chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}}\chi^{B}_{\vec{\alpha}_{2}(\vec{\beta}_{1}+\vec{\beta}_{2})}\chi^{B}_{\vec{0}\vec{0}}}$

$$\leq \sum_{\vec{\beta}_{1}\neq\vec{0}} \sqrt{\chi^{A}_{\vec{0}\vec{\beta}_{1}}} \sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}}\chi^{B}_{\vec{\alpha}_{2}(\vec{\beta}_{1}+\vec{\beta}_{2})}} \tag{B24}$$

$$\leq \sum_{\vec{\beta}_{1}\neq\vec{0}} \sqrt{\chi^{A}_{\vec{0}\vec{\beta}_{1}}} \sqrt{\sum_{\vec{\alpha}_{2}\neq\vec{0}} \chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}}} \sqrt{\sum_{\vec{\alpha}_{2}\neq\vec{0}} \chi^{B}_{\vec{\alpha}_{2}\vec{\beta}_{2}}}$$
(B25)

$$\leq 2^{N/2} \sqrt{\sum_{\vec{\beta}_1 \neq \vec{0}} \chi^A_{\vec{0}\vec{\beta}_1}} \sqrt{p^A_{\rm ND} p^B_{\rm ND}} \tag{B26}$$

$$\leq 2^{N/2} \sqrt{p_{\rm D}^A p_{\rm ND}^A p_{\rm ND}^B} \tag{B27}$$

$$\sum_{\substack{\vec{\alpha}_{2},\vec{\delta}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{\vec{\alpha}_{2}(\vec{\beta}_{2}+\vec{\delta}_{2})}^{B}\chi_{\vec{0}\vec{\delta}_{2}}^{B}} \le 2^{N/2}\sqrt{p_{\mathrm{D}}^{B}p_{\mathrm{ND}}^{A}p_{\mathrm{ND}}^{B}}$$
(B28)

$$\leq 2^{N} \sqrt{\sum_{\vec{\beta}_{1} \neq \vec{0}} \chi^{A}_{\vec{0}\vec{\beta}_{1}}} \sqrt{\sum_{\vec{\delta}_{2} \neq \vec{0}} \chi^{B}_{\vec{0}\vec{\delta}_{2}}} \sqrt{p^{A}_{\text{ND}} p^{B}_{\text{ND}}}$$
(B31)

$$\leq 2^N \sqrt{p_{\rm D}^A p_{\rm D}^B p_{\rm ND}^A p_{\rm ND}^B}.$$
(B32)

Thus, in total we have

$$|(3)| \le \sqrt{p_{\rm ND}^A p_{\rm ND}^B} \left(1 + 2^{N/2} \sqrt{p_{\rm D}^A} \right) \left(1 + 2^{N/2} \sqrt{p_{\rm D}^B} \right).$$
(B33)

By symmetry, we also have

$$|(4)| \le \sqrt{p_{\rm ND}^A p_{\rm ND}^B} \left(1 + 2^{N/2} \sqrt{p_{\rm D}^A} \right) \left(1 + 2^{N/2} \sqrt{p_{\rm D}^B} \right).$$
(B34)

Continuing, we have

$$\begin{aligned} |(5)| &\leq \sum_{\substack{\vec{\alpha}_{2}, \vec{\gamma}_{2} \neq \vec{0} \\ \vec{\alpha}_{2} \neq \vec{\gamma}_{2} \\ \vec{\sigma}_{2} \neq \vec{\gamma}_{2} \\ \vec{\sigma}_{1}, \vec{\beta}_{2}, \vec{\delta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A} \chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A} \chi_{(\vec{\alpha}_{2} + \vec{\gamma}_{2})(\vec{\beta}_{1} + \vec{\beta}_{2} + \vec{\delta}_{2})}^{B} \chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}} \\ &= \sum_{\substack{\vec{\alpha}_{2}, \vec{\gamma}_{2} \neq \vec{0} \\ \vec{\alpha}_{2} \neq \vec{\gamma}_{2} \\ \vec{\beta}_{2}, \vec{\delta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A} \chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A} \chi_{(\vec{\alpha}_{2} + \vec{\gamma}_{2})(\vec{\beta}_{1} + \vec{\beta}_{2} + \vec{\delta}_{2})}^{B} \chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}} \\ &+ \sum_{\substack{\vec{\alpha}_{2}, \vec{\beta}_{1}, \vec{\gamma}_{2} \neq \vec{0} \\ \vec{\alpha}_{2} \neq \vec{\gamma}_{2} \\ \vec{\beta}_{2}, \vec{\delta}_{2}}} \sqrt{\chi_{\vec{0}\vec{\beta}_{1}}^{A} \chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A} \chi_{(\vec{\alpha}_{2} + \vec{\gamma}_{2})(\vec{\beta}_{1} + \vec{\beta}_{2} + \vec{\delta}_{2})}^{B} \chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}} \end{aligned} \tag{B36}$$

each of which can be bounded as

$$\sum_{\substack{\vec{\alpha}_{2},\vec{\gamma}_{2}\neq\vec{0}\\\vec{\alpha}_{2}\neq\vec{\gamma}_{2}\\\vec{\beta}_{2},\vec{\delta}_{2}}} \sqrt{\chi_{\vec{0}\vec{0}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{(\vec{\alpha}_{2}+\vec{\gamma}_{2})(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}^{B}\chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}} \\
\leq \sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}} \sum_{\substack{\vec{\gamma}_{2}\neq\vec{0},\vec{\alpha}_{2}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{B}} \sum_{\vec{\lambda}_{2}} \chi_{\vec{\alpha}_{2}\vec{\delta}_{2}}^{B}} (B37) \\
\leq \sum_{\substack{\vec{\lambda}_{2}\neq\vec{0}\\\vec{\beta}_{2}}} \sqrt{\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}} \sum_{\substack{\vec{\lambda}_{2}\neq\vec{0},\vec{\lambda}_{2}\\\vec{\delta}_{2}}} \chi_{\vec{\beta}_{2}\vec{\delta}_{2}}^{B}} (B38)$$

$$\leq \sum_{\substack{\vec{\alpha}_2 \neq \vec{0} \\ \vec{\beta}_2}} \sqrt{\chi^A_{\vec{\alpha}_2 \vec{\beta}_2}} \sum_{\substack{\vec{\gamma}_2 \neq \vec{0} \\ \vec{\delta}_2}} \chi^B_{\vec{\alpha}_2 \vec{\delta}_2} \tag{B38}$$

$$\leq 2^{N} \sqrt{\sum_{\substack{\vec{\alpha}_{2} \neq \vec{0} \\ \vec{\beta}_{2}}} \chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}} p^{B}_{\rm ND}} \tag{B39}$$

$$=2^N \sqrt{p_{\rm ND}^A} p_{\rm ND}^B \tag{B40}$$

$$\sum_{\substack{\vec{\alpha}_{2},\vec{\beta}_{1},\vec{\gamma}_{2}\neq\vec{0}\\\vec{\alpha}_{2}\neq\vec{\gamma}_{2}\\\vec{\beta}_{2},\vec{\delta}_{2}}} \sqrt{\chi_{\vec{0}\vec{\beta}_{1}}^{A}\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}\chi_{(\vec{\alpha}_{2}+\vec{\gamma}_{2})(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}^{B}\chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}} = \sum_{\vec{\beta}_{1}\neq\vec{0}} \sqrt{\chi_{\vec{0}\vec{\beta}_{1}}^{A}} \sum_{\substack{\vec{\alpha}_{2}\neq\vec{0}\\\vec{\beta}_{2}\neq\vec{0}}} \sqrt{\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}} \sum_{\vec{\gamma}_{2}\neq\vec{0},\vec{\alpha}_{2}} \sqrt{\chi_{(\vec{\alpha}_{2}+\vec{\gamma}_{2})(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}} \chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}}$$
(B41)

$$\leq \sum_{\vec{\beta}_1 \neq \vec{0}} \sqrt{\chi^A_{\vec{0}\vec{\beta}_1}} \sum_{\substack{\vec{\alpha}_2 \neq \vec{0} \\ \vec{\beta}_2}} \sqrt{\chi^A_{\vec{\alpha}_2\vec{\beta}_2}} \sum_{\substack{\vec{\gamma}_2 \neq \vec{0} \\ \vec{\delta}_2}} \chi^B_{\vec{\gamma}_2\vec{\delta}_2}$$
(B42)

$$\leq 2^{3N/2} \sqrt{\sum_{\vec{\beta}_1 \neq \vec{0}} \chi^A_{\vec{0}\vec{\beta}_1}} \sqrt{\sum_{\vec{\alpha}_2 \neq \vec{0}} \chi^A_{\vec{\alpha}_2 \vec{\beta}_2}} p^B_{\rm ND}$$
(B43)

$$\leq 2^{3N/2} \sqrt{p_{\rm ND}^A p_{\rm D}^A} p_{\rm ND}^B.$$
 (B44)

Thus in total,

$$|(5)| \le 2^N p_{\rm ND}^B \sqrt{p_{\rm ND}^A} \left(1 + 2^{N/2} \sqrt{p_{\rm D}^A}\right)$$
 (B45)

By symmetry, we also have

$$|(6)| \le 2^N p_{\rm ND}^B \sqrt{p_{\rm ND}^A} \left(1 + 2^{N/2} \sqrt{p_{\rm D}^A}\right)$$
 (B46)

$$|(7)| \le 2^N p_{\rm ND}^A \sqrt{p_{\rm ND}^B} \left(1 + 2^{N/2} \sqrt{p_{\rm D}^B}\right)$$
 (B47)

$$|(8)| \le 2^N p_{\rm ND}^A \sqrt{p_{\rm ND}^B} \left(1 + 2^{N/2} \sqrt{p_{\rm D}^B}\right).$$
 (B48)

Finally,

$$\begin{split} |(9)| &\leq \sum_{\substack{\vec{\alpha}_{1},\vec{\alpha}_{2},\vec{\gamma}_{2}\neq\vec{0}\\\vec{\alpha}_{1}+\vec{\alpha}_{2}\neq\vec{\gamma}_{2}\\\vec{\beta}_{1},\vec{\beta}_{2},\vec{\delta}_{2}}} \sqrt{\chi^{A}_{\vec{\alpha}_{1}\vec{\beta}_{1}}\chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}}\chi^{B}_{(\vec{\alpha}_{1}+\vec{\alpha}_{2}+\vec{\gamma}_{2})(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})}\chi^{B}_{\vec{\gamma}_{2}\vec{\delta}_{2}}} \end{split}$$
(B49)

$$= \sum_{\substack{\vec{\alpha}_{1} \neq \vec{0} \\ \vec{\beta}_{1}}} \sqrt{\chi_{\vec{\alpha}_{1}\vec{\beta}_{1}}^{A}} \sum_{\substack{\vec{\alpha}_{2} \neq \vec{0} \\ \vec{\beta}_{2}}} \sqrt{\chi_{\vec{\alpha}_{2}\vec{\beta}_{2}}^{A}} \times \sum_{\substack{\vec{\gamma}_{2} \neq \vec{0}, (\vec{\alpha}_{1} + \vec{\alpha}_{2}) \\ \vec{\gamma}_{2} \neq \vec{0}, (\vec{\alpha}_{1} + \vec{\alpha}_{2})}} \sqrt{\chi_{(\vec{\alpha}_{1} + \vec{\alpha}_{2} + \vec{\gamma}_{2})(\vec{\beta}_{1} + \vec{\beta}_{2} + \vec{\delta}_{2})} \chi_{\vec{\gamma}_{2}\vec{\delta}_{2}}^{B}}$$
(B50)

$$\leq \sum_{\substack{\vec{\alpha}_{1} \neq \vec{0} \\ \vec{\beta}_{1}}} \sqrt{\chi^{A}_{\vec{\alpha}_{1}\vec{\beta}_{1}}} \sum_{\substack{\vec{\alpha}_{2} \neq \vec{0} \\ \vec{\beta}_{2}}} \sqrt{\chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}}} \sum_{\substack{\vec{\gamma}_{2} \neq \vec{0} \\ \vec{\delta}_{2}}} \chi^{B}_{\vec{\gamma}_{2}\vec{\delta}_{2}}$$
(B51)

$$\leq 2^{2N} \sum_{\substack{\vec{\alpha}_{1} \neq \vec{0} \\ \vec{\beta}_{1}}} \chi^{A}_{\vec{\alpha}_{1}\vec{\beta}_{1}} \chi^{A}_{\vec{\alpha}_{2}\vec{\beta}_{2}} p^{B}_{\rm ND}$$
(B52)

$$\leq 2^{2N} p_{\rm ND}^A p_{\rm ND}^B \tag{B53}$$

Combining the bounds given in Eqs. B17, B18, B33, B34, B45, B46, B47, B48, and B53, we have that

$$p_{\rm ND} = p_{\rm ND}^A + p_{\rm ND}^B + \epsilon \tag{B54}$$

where the error term ϵ is bounded by

$$\begin{split} |\epsilon| <& 2\sqrt{p_{\rm ND}^A p_{\rm ND}^B} \\ &+ 2^{N/2+1} \left[\sqrt{p_{\rm D}^A p_{\rm ND}^B} + p_{\rm ND}^A \sqrt{p_{\rm D}^B} \\ &+ \sqrt{p_{\rm D}^A p_{\rm ND}^A p_{\rm ND}^B} + \sqrt{p_{\rm ND}^A p_{\rm D}^B p_{\rm ND}^B} \right] \\ &+ 2^N \left[p_{\rm D}^A p_{\rm ND}^B + p_{\rm ND}^A p_{\rm D}^B + 2\sqrt{p_{\rm D}^A p_{\rm ND}^A p_{\rm D}^B p_{\rm ND}^B} \\ &+ 2p_{\rm ND}^A \sqrt{p_{\rm D}^B} + 2\sqrt{p_{\rm D}^A p_{\rm ND}^B} \right] \\ &+ 2^{3N/2+1} \left[p_{\rm ND}^A \sqrt{p_{\rm D}^B p_{\rm ND}^B} + \sqrt{p_{\rm D}^A p_{\rm ND}^A p_{\rm ND}^B} \right] \\ &+ 2^{2N} p_{\rm ND}^A p_{\rm ND}^B. \end{split}$$
(B55)

While the bound on $|\epsilon|$ involves many terms, in the case where $p_{\rm D}^{A,B} \ll 2^{-N}$ and $p_{\rm ND}^{A,B}/p_{\rm D}^{A,B} \ll 2^{-N}$ the bound is essentially just $2\sqrt{p_{\rm ND}^A p_{\rm ND}^B}$, which gives Eq. 9 of the main text. This is the relevant regime for high fidelity, highlybiased channels on a small number of qubits. It is also the relevant regime for error channels in which off-diagonal elements of the χ -matrix are exponentially suppressed in the weight of the Pauli errors, which is likely the case for error-correction circuits as they tend to decohere noise [52–55].

2. Finding the dephasing error probability

From the definition of $p_{\rm D}$, Eq. 3, we have

$$p_{\mathrm{D}} = \sum_{\substack{\vec{\alpha}_1 + \vec{\gamma}_1 = \vec{\alpha}_2 + \vec{\gamma}_2 = \vec{0} \\ \vec{\beta}_1 + \vec{\delta}_1 = \vec{\beta}_2 + \vec{\delta}_2}} (-1)^{\vec{\beta}_1 \cdot \vec{\gamma}_1 + \vec{\beta}_2 \cdot \vec{\gamma}_2} \chi^A_{\vec{\alpha}_1 \vec{\beta}_1, \vec{\alpha}_2 \vec{\beta}_2} \chi^B_{\vec{\gamma}_1 \vec{\delta}_1, \vec{\gamma}_2 \vec{\delta}_2}.$$
(B56)

We could similarly divide the terms in this sum into categories and bound them, as we did for $p_{\rm ND}$ above. However, if we assume $p_{\rm D}^{A,B} \gg p_{\rm ND}^{A,B}$, we can instead approximate $p_{\rm D}$ by $p = p_{\rm D} + p_{\rm ND}$, the total error probability of the channel. We then use the known result [58, Theorem 1]

$$p = p^A + p^B + \epsilon, \qquad |\epsilon| \le 2\sqrt{p^A p^B}.$$
 (B57)

3. Extracting the dephasing and non-dephasing probabilities in IBRB

Given the estimates of $p_{\rm D}$ and $p_{\rm ND}$ given in Eqs. B57 and B54, respectively, we can reverse these equations to try and estimate $p_{\rm D}^A$ and $p_{\rm ND}^A$ from $p_{\rm D}^B$, $p_{\rm ND}^B$, $p_{\rm D}$, and $p_{\rm ND}$. This is relevant if Λ_A is the error channel associated to our gate of interest and Λ_B is the error channel of our interleaving group. For convenience, we'll assume we're in the regime where we can neglect the error terms proportional to $2^{N/2}$ and higher powers. Rearranging Eqs. B57 and B54 gives

$$p_{\rm D}^A = p_{\rm D} + p_{\rm D}^B + \epsilon, \qquad |\epsilon| < 2\sqrt{p_{\rm D}p_{\rm D}^B} \quad (B58)$$

$$p_{\rm ND}^A = p_{\rm ND} + p_{\rm ND}^B + \epsilon, \qquad |\epsilon| < 2\sqrt{p_{\rm ND}p_{\rm ND}^B}.$$
 (B59)

We note that in our particular case, it seems difficult to extract $p_{\rm D}^B$ and $p_{\rm ND}^B$ for the group \mathcal{Z}_2 , as the Liouville representation of this group has high multiplicity. This is why we prefer the randomized compiling interpretation of $p_{\rm D}$ and $p_{\rm ND}$ given in the main text.

Appendix C: Randomized compiling with Z randomization

In this section, we explain the randomized compiling by \mathcal{Z}_N in more detail, and prove that it ensures nondephasing errors increase linearly under composition (Eq. 24 of the main text). Given an N-qubit bias-preserving Clifford circuit composed of gates C_1, C_2, \ldots, C_n , a noisy implementation of the circuit is given by

$$\hat{C}_n \hat{\Lambda}_n \cdots \hat{C}_2 \hat{\Lambda}_2 \hat{C}_1 \hat{\Lambda}_1 \tag{C1}$$

where $\hat{\Lambda}_1, \hat{\Lambda}_2, \ldots, \hat{\Lambda}_n$ are the associated error channels. Let's assume that we can implement arbitrary elements of \mathcal{Z}_N with a gate-independent error channel $\hat{\Lambda}_G$, where $\hat{\Lambda}_G$ is negligible red to the Clifford errors. We take the convention that the In this case, we can interleave randomly chosen gates $U \in \mathcal{Z}_N$ between the Clifford elements without increasing the error. Note that because the circuit elements are Clifford, the effect of interleaving U can be corrected by an efficiently computable Pauli correction operator. We'll denote the correction operator by U_{n+1} . Note that because the Cliffords preserve the bias, we have $U_{n+1} \in \mathcal{Z}_N$.

The resulting noisy circuit is then

$$\hat{\Lambda}_G \hat{U}_{n+1} \hat{C}_n \hat{\Lambda}_n \hat{\Lambda}_G \hat{U}_n \cdots \hat{C}_2 \hat{\Lambda}_2 \hat{\Lambda}_G \hat{U}_2 \hat{C}_1 \hat{\Lambda}_1 \hat{\Lambda}_G \hat{U}_1.$$
(C2)

Because the Cliffords preserve the bias, commuting any element of Z_N past a Clifford results in another element of Z_N . We can thus rewrite the circuit as

$$\hat{\Lambda}_G \hat{C}_n \hat{V}_n^{\dagger} \hat{\Lambda}_n \hat{\Lambda}_G \hat{V}_n \cdots \hat{C}_2 \hat{V}_2^{\dagger} \hat{\Lambda}_2 \hat{\Lambda}_G \hat{V}_2 \hat{C}_1 \hat{V}_1^{\dagger} \hat{\Lambda}_1 \hat{\Lambda}_G \hat{V}_1.$$
(C3)

for some $V_i \in \mathcal{Z}_N$ that are also distributed uniformly over \mathcal{Z}_N . Taking the expectation value over $\hat{V}_1, \ldots, \hat{V}_n$ results in an effective circuit of the form

$$\hat{\Lambda}_G \hat{C}_n \hat{\Lambda}_n^T \cdots \hat{C}_2 \hat{\Lambda}_2^T \hat{C}_1 \hat{\Lambda}_1^T.$$
(C4)

where the twirled error channel Λ_i^T is given by

$$\hat{\Lambda}_{i}^{T} = \mathop{\mathbb{E}}_{V \in \mathcal{Z}_{N}} \left[\hat{V}^{\dagger} \hat{\Lambda}_{i} \hat{\Lambda}_{G} \hat{V} \right].$$
(C5)

The twirled version of the error channels is highly simplified compared to the original error channel. In terms of the χ -matrix, if the original error channel is given by

$$(\Lambda_{i} \circ \Lambda_{G})(\rho) = \sum_{\substack{\vec{\alpha}_{1}, \vec{\beta}_{1} \\ \vec{\alpha}_{2}, \vec{\beta}_{2}}} \chi_{\vec{\alpha}_{1}\vec{\beta}_{1}, \vec{\alpha}_{2}\vec{\beta}_{2}} \mathbf{X}(\vec{\alpha}_{1}) \mathbf{Z}(\vec{\beta}_{1})\rho \mathbf{Z}(\vec{\beta}_{2}) \mathbf{X}(\vec{\alpha}_{2})$$
(C6)

then the twirled error channel is given by

$$\Lambda_i^T(\rho) = \sum_{\substack{\vec{\alpha} \\ \vec{\beta}_1, \vec{\beta}_2}} \chi_{\vec{\alpha}\vec{\beta}_1, \vec{\alpha}\vec{\beta}_2} \mathbf{X}(\vec{\alpha}) \mathbf{Z}(\vec{\beta}_1) \rho \mathbf{Z}(\vec{\beta}_2) \mathbf{X}(\vec{\alpha})$$
(C7)

where the off-diagonal elements of the χ -matrix with $\vec{\alpha}_1 \neq \vec{\alpha}_2$ are set to zero.

If we consider the composition of two twirled error channels $(\Lambda_A^T \circ \Lambda_B^T)$, we can again estimate the nondephasing probability of the composition in terms of the dephasing and non-dephasing probabilities of Λ_A and Λ_B . In evaluating the sum in Section B 1, the fact that the error channels are twirled means the only nonzero terms in the sum have $\vec{\alpha}_1 = \vec{\alpha}_2$ and $\vec{\gamma}_1 = \vec{\gamma}_2$. These leaves the sum over subsets (1) and (2) unchanged, makes the sum over subsets (3)-(8) zero, and lets us reevaluate the sum over subset (9) as

$$|(9)| \leq \sum_{\vec{\alpha},\vec{\gamma}\neq\vec{0}} \sqrt{\chi^A_{\vec{\alpha}\vec{\beta}_1} \chi^A_{\vec{\alpha}\vec{\beta}_2} \chi^B_{\vec{\gamma}(\vec{\beta}_1+\vec{\beta}_2+\vec{\delta}_2)} \chi^B_{\vec{\gamma}\vec{\delta}_2}}$$
(C8)

$$=\sum_{\substack{\vec{\alpha}\neq\vec{0}\\\vec{\beta}_{1}}} \sqrt{\chi^{A}_{\vec{\alpha}\vec{\beta}_{1}}} \sum_{\vec{\beta}_{2}} \sqrt{\chi^{A}_{\vec{\alpha}\vec{\beta}_{2}}} \sum_{\substack{\vec{\gamma}\neq\vec{0}\\\vec{\delta}_{2}}} \sqrt{\chi^{B}_{\vec{\gamma}(\vec{\beta}_{1}+\vec{\beta}_{2}+\vec{\delta}_{2})} \chi^{B}_{\vec{\gamma}\vec{\delta}_{2}}}$$
(C9)

$$\leq \sum_{\substack{\vec{\alpha}\neq\vec{0}\\\vec{\beta}_1}} \sqrt{\chi^A_{\vec{\alpha}\vec{\beta}_1}} \sum_{\vec{\beta}_2} \sqrt{\chi^A_{\vec{\alpha}\vec{\beta}_2}} \sum_{\substack{\vec{\gamma}\neq\vec{0}\\\vec{\delta}_2}} \chi^B_{\vec{\gamma}\vec{\delta}_2} \tag{C10}$$

$$\leq 2^{N} \sum_{\vec{\alpha} \neq \vec{0}} \sqrt{\sum_{\vec{\beta}_{1}} \chi^{A}_{\vec{\alpha}\vec{\beta}_{1}}} \sqrt{\sum_{\vec{\beta}_{1}} \chi^{A}_{\vec{\alpha}\vec{\beta}_{2}}} p^{B}_{\rm ND} \tag{C11}$$

$$\leq 2^{N} \sum_{\substack{\vec{\alpha} \neq \vec{0} \\ \vec{\beta}_{1}}} \chi^{A}_{\vec{\alpha}\vec{\beta}_{1}} p^{B}_{\rm ND} \tag{C12}$$

$$\leq 2^N p_{\rm ND}^A p_{\rm ND}^B. \tag{C13}$$

Thus by combining Eqs. B17, B18, and C13, we have

$$p_{\rm ND} = p_{\rm ND}^{A} + p_{\rm ND}^{B} + \epsilon$$
(C14)
$$|\epsilon| \le 2^{N/2+1} \left[\sqrt{p_{\rm D}^{A}} p_{\rm ND}^{B} + p_{\rm ND}^{A} \sqrt{p_{\rm D}^{B}} \right]$$
$$+ 2^{N} \left[p_{\rm ND}^{A} p_{\rm D}^{B} + p_{\rm D}^{A} p_{\rm ND}^{B} + p_{\rm ND}^{A} p_{\rm ND}^{B} \right].$$
(C15)

In the regime of high-fidelity gates on a small number of qubits, $|\epsilon|$ is negligible, which gives Eq. 24 in the main text.

In contrast, randomized Z_2 compiling produces no notable improvement to the bounds on p_D for a highly biased noise channel, as in this case the dominant uncertainty in p_D comes from off-diagonal elements of the χ -

- A. M. Stephens, W. J. Munro, and K. Nemoto, Highthreshold topological quantum error correction against biased noise, Phys. Rev. A 88, 060301 (2013).
- [2] D. K. Tuckett, S. D. Bartlett, and S. T. Flammia, Ultrahigh error threshold for surface codes with biased noise, Phys. Rev. Lett. **120**, 050505 (2018).
- [3] D. K. Tuckett, A. S. Darmawan, C. T. Chubb, S. Bravyi, S. D. Bartlett, and S. T. Flammia, Tailoring surface codes for highly biased noise, Phys. Rev. X 9, 041031 (2019).
- [4] D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J.

matrices χ^A and χ^B with $\vec{\alpha}_1 = \vec{\alpha}_2 = \vec{\gamma}_1 = \vec{\gamma}_2 = \vec{0}$, which are not affected by twirling.

Appendix D: Generating random biased-noise error channels

Here, we give more details on our procedure to generate random biased-noise error channels to use in our simulation data. We make no claim that this procedure is optimal or generates all realistic error channels; our goal was simply to generate error channels that were both biased and not Pauli-diagonal to illustrate the power of our BRB methods.

We first randomly choose the number $d = 1, ..., 4^N$ of Kraus operators to include, as well as the approximate target probabilities p_D and p_{ND} for the channel. For i = 1, ..., (d-1), we choose each Kraus operator to be either dephasing or non-dephasing with probability 1/2, and set it to

$$K_{i} := \begin{cases} \sqrt{\frac{10p_{\mathrm{D}}}{d}} \sum_{\vec{\beta}} c_{\vec{\beta}} \mathbf{Z}(\vec{\beta}), & \text{dephasing} \\ \sqrt{\frac{10p_{\mathrm{ND}}}{d}} \sum_{\vec{\alpha} \neq \vec{0} \atop \vec{\beta}} c_{\vec{\alpha},\vec{\beta}} \mathbf{X}(\vec{\alpha}) \mathbf{Z}(\vec{\beta}), & \text{non-dephasing} \end{cases}$$
(D1)

where each c is generated by choosing a uniform $r \in [0,1]$, $\theta \in [0,2\pi]$, and setting $c = re^{i\theta}$. The factor of 10 was inserted "by hand" to make the resulting error channel approximately have the desired dephasing/non-dephasing probabilities.

Finally, having defined K_i for i = 1, ..., (d-1), we define K_d to be a matrix satisfying

$$K_d^{\dagger} K_d = \mathbb{1} - \sum_{i=1}^{d-1} K_i^{\dagger} K_i$$
 (D2)

to ensure the overall channel is trace-preserving. While this description of K_d is not unique, one matrix satisfying this equation is given by the Cholesky decomposition of $(\mathbb{1} - \sum_i K_i^{\dagger} K_i)$ [64], which is what we used in our simulations.

Brown, Fault-tolerant thresholds for the surface code in excess of 5% under biased noise, Phys. Rev. Lett. **124**, 130501 (2020).

- [5] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, The XZZX surface code, Nat. Commun. 12, 1 (2021).
- [6] A. Dua, A. Kubica, L. Jiang, S. T. Flammia, and M. J. Gullans, Clifford-deformed surface codes, arXiv preprint arXiv:2201.07802 (2022).
- [7] J. Claes, J. E. Bourassa, and S. Puri, Tailored clus-

- [8] A. S. Darmawan, B. J. Brown, A. L. Grimsmo, D. K. Tuckett, and S. Puri, Practical quantum error correction with the XZZX code and Kerr-cat qubits, PRX Quantum 2, 030345 (2021).
- [9] C. Chamberland, K. Noh, P. Arrangoiz-Arriola, E. T. Campbell, C. T. Hann, J. Iverson, H. Putterman, T. C. Bohdanowicz, S. T. Flammia, A. Keller, *et al.*, Building a fault-tolerant quantum computer using concatenated cat codes, PRX Quantum **3**, 010329 (2022).
- [10] C. Chamberland and E. T. Campbell, Universal quantum computing with twist-free and temporally encoded lattice surgery, PRX Quantum 3, 010331 (2022).
- [11] P. Aliferis and J. Preskill, Fault-tolerant quantum computation against biased noise, Phys. Rev. A 78, 052331 (2008).
- [12] S. Puri, L. St-Jean, J. A. Gross, A. Grimm, N. E. Frattini, P. S. Iyer, A. Krishna, S. Touzard, L. Jiang, A. Blais, *et al.*, Bias-preserving gates with stabilized cat qubits, Sci. Adv. 6, 5901 (2020).
- [13] J. Guillaud and M. Mirrahimi, Repetition cat qubits for fault-tolerant quantum computation, Phys. Rev. X 9, 041053 (2019).
- [14] I. Cong, S.-T. Wang, H. Levine, A. Keesling, and M. D. Lukin, Hardware-efficient, fault-tolerant quantum computation with Rydberg atoms, arXiv preprint arXiv:2105.13501 (2021).
- [15] Q. Xu, J. K. Iverson, F. G. Brandao, and L. Jiang, Engineering fast bias-preserving gates on stabilized cat qubits, arXiv preprint arXiv:2105.13908 (2021).
- [16] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. 106, 180504 (2011).
- [17] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, Phys. Rev. A 85, 042311 (2012).
- [18] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, Journal of Optics B: Quantum and Semiclassical Optics 7, S347 (2005).
- [19] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, Physical Review A 77, 012307 (2008).
- [20] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. Da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, *et al.*, Efficient measurement of quantum gate error by interleaved randomized benchmarking, Phys. Rev. Lett. **109**, 080505 (2012).
- [21] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Characterizing universal gate sets via dihedral benchmarking, Phys. Rev. A 92, 060302 (2015).
- [22] A. W. Cross, E. Magesan, L. S. Bishop, J. A. Smolin, and J. M. Gambetta, Scalable randomised benchmarking of non-Clifford gates, npj Quantum Inf. 2, 1 (2016).
- [23] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, A new class of efficient randomized benchmarking protocols, npj Quantum Inf. 5, 1 (2019).
- [24] J. Claes, E. Rieffel, and Z. Wang, Character randomized benchmarking for non-multiplicity-free groups with applications to subspace, leakage, and matchgate randomized benchmarking, PRX Quantum 2, 010351 (2021).

- [25] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, A general framework for randomized benchmarking, arXiv preprint arXiv:2010.07974 (2020).
- [26] W. G. Brown and B. Eastin, Randomized benchmarking with restricted gate sets, Phys. Rev. A 97, 062323 (2018).
- [27] D. S. França and A. Hashagen, Approximate randomized benchmarking for finite groups, J. Phys. A 51, 395302 (2018).
- [28] R. Harper and S. T. Flammia, Estimating the fidelity of T gates using standard interleaved randomized benchmarking, Quantum Sci. Technol. 2, 015008 (2017).
- [29] E. Onorati, A. Werner, and J. Eisert, Randomized benchmarking for individual quantum gates, Phys. Rev. Lett. 123, 060501 (2019).
- [30] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. 10, 1 (2019).
- [31] T. Chasseur, D. M. Reich, C. P. Koch, and F. K. Wilhelm, Hybrid benchmarking of arbitrary quantum gates, Phys. Rev. A 95, 062335 (2017).
- [32] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, New J. Phys 17, 113020 (2015).
- [33] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Robust extraction of tomographic information via randomized benchmarking, Phys. Rev. X 4, 011050 (2014).
- [34] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Trans. Quantum Comput. 1, 1 (2020).
- [35] R. Harper, S. T. Flammia, and J. J. Wallman, Efficient learning of quantum noise, Nat. Phys. 16, 1184 (2020).
- [36] C. Baldwin, B. Bjork, J. Gaebler, D. Hayes, and D. Stack, Subspace benchmarking high-fidelity entangling operations with trapped ions, Physical Review Research 2, 013317 (2020).
- [37] J. J. Wallman, Randomized benchmarking with gatedependent noise, Quantum 2, 47 (2018).
- [38] S. T. Merkel, E. J. Pritchett, and B. H. Fong, Randomized benchmarking as convolution: Fourier analysis of gate dependent errors, Quantum 5, 581 (2021).
- [39] J. J. Wallman and S. T. Flammia, Randomized benchmarking with confidence, New J. Phys 16, 103032 (2014).
- [40] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, Multiqubit randomized benchmarking using few samples, Phys. Rev. A 100, 032304 (2019).
- [41] A. Carignan-Dugas, K. Boone, J. J. Wallman, and J. Emerson, From randomized benchmarking experiments to gate-set circuit fidelity: how to interpret randomized benchmarking decay parameters, New J. Phys 20, 092001 (2018).
- [42] X. Xue, T. Watson, J. Helsen, D. R. Ward, D. E. Savage, M. G. Lagally, S. N. Coppersmith, M. Eriksson, S. Wehner, and L. Vandersypen, Benchmarking gate fidelities in a si/sige two-qubit device, Phys. Rev. X 9, 021011 (2019).
- [43] G. Feng, J. J. Wallman, B. Buonacorsi, F. H. Cho, D. K. Park, T. Xin, D. Lu, J. Baugh, and R. Laflamme, Estimating the coherence of noise in quantum control of a solid-state qubit, Phys. Rev. Lett. **117**, 260501 (2016).
- [44] S. T. Flammia, Averaged circuit eigenvalue sampling, arXiv preprint arXiv:2108.05803 (2021).

- [45] J. Helsen, S. Nezami, M. Reagor, and M. Walter, Matchgate benchmarking: Scalable benchmarking of a continuous family of many-qubit gates, Quantum 6, 657 (2022).
- [46] A. Grimm, N. E. Frattini, S. Puri, S. O. Mundhada, S. Touzard, M. Mirrahimi, S. M. Girvin, S. Shankar, and M. H. Devoret, Stabilization and operation of a Kerr-cat qubit, Nature 584, 205 (2020).
- [47] S. Puri, S. Boutin, and A. Blais, Engineering the quantum states of light in a Kerr-nonlinear resonator by twophoton driving, npj Quantum Inf. 3, 1 (2017).
- [48] C. J. Wood, J. D. Biamonte, and D. G. Cory, Tensor networks and graphical calculus for open quantum systems, arXiv preprint arXiv:1111.6950 (2011).
- [49] The usual definition of the χ -matrix differs from this one by factors of *i* off the diagonal, which are not relevant for defining the bias.
- [50] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, Phys. Lett. A 303, 249 (2002).
- [51] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A 60, 1888 (1999).
- [52] E. Huang, A. C. Doherty, and S. Flammia, Performance of quantum error correction with coherent errors, Phys. Rev. A 99, 022313 (2019).
- [53] S. J. Beale, J. J. Wallman, M. Gutiérrez, K. R. Brown, and R. Laflamme, Quantum error correction decoheres noise, Phys. Rev. Lett. **121**, 190501 (2018).
- [54] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, npj Quantum Inf. 4, 1 (2018).

- [55] M. R. Geller and Z. Zhou, Efficient error models for faulttolerant architectures and the Pauli twirling approximation, Phys. Rev. A 88, 012314 (2013).
- [56] S. Garion and A. W. Cross, Synthesis of CNOT-dihedral circuits with optimal number of two qubit gates, Quantum 4, 369 (2020).
- [57] M. Amy, J. Chen, and N. J. Ross, A finite presentation of CNOT-dihedral operators, arXiv preprint arXiv:1701.00140 (2016).
- [58] A. Carignan-Dugas, J. J. Wallman, and J. Emerson, Bounding the average gate fidelity of composite channels using the unitarity, New J. Phys 21, 053016 (2019).
- [59] E. Knill, Quantum computing with realistically noisy devices, Nature 434, 39 (2005).
- [60] L. Viola and E. Knill, Random decoupling schemes for quantum dynamical control and error suppression, Physical review letters 94, 060502 (2005).
- [61] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A 94, 052325 (2016).
- [62] We could also include single-qubit T gates or even twoqubit CZ gates in our group, as these gates are also highfidelity for Kerr cat qubits, but we did not find these additional gates helpful in designing our protocol.
- [63] W. Fulton and J. Harris, *Representation theory: a first course*, Vol. 129 (Springer Science & Business Media, 2013).
- [64] G. H. Golub and C. F. Van Loan, *Matrix computations* (JHU press, 2013).