



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Hardware-efficient random circuits to classify noise in a multiqubit system

Jin-Sung Kim, Lev S. Bishop, Antonio D. Córcoles, Seth Merkel, John A. Smolin, and Sarah Sheldon

Phys. Rev. A **104**, 022609 — Published 17 August 2021

DOI: [10.1103/PhysRevA.104.022609](https://doi.org/10.1103/PhysRevA.104.022609)

# Hardware-efficient random circuits to classify noise in a multi-qubit system

Jin-Sung Kim,\* Lev S. Bishop, Antonio D. Córcoles, Seth Merkel, John A. Smolin, and Sarah Sheldon

*IBM Quantum*

(Dated: July 26, 2021)

In this work we extend a multi-qubit benchmarking technique known as the Binned Output Generation (BOG) in order to discriminate between coherent and incoherent noise sources in the multi-qubit regime. While methods exist to discriminate coherent from incoherent noise at the single and few-qubit level, these methods scale poorly beyond a few qubits or must make assumptions about the form of the noise. On the other end of the spectrum, system-level benchmarking techniques exist, but fail to discriminate between coherent and incoherent noise sources. We experimentally verify the BOG against Randomized Benchmarking (RB) (the industry standard benchmarking technique) in the two-qubit regime, then apply this technique to a six qubit linear chain, a regime currently inaccessible to RB. In this experiment we inject an instantaneous coherent  $Z$ -type noise on each qubit and demonstrate that the measured coherent noise scales correctly with the magnitude of the injected noise, while the measured incoherent noise remains unchanged as expected. This demonstrates a robust technique to measure coherent errors in a variety of hardware.

A fundamental barrier in achieving quantum advantage in quantum computers is the presence of noise, which reduces the length and complexity of computations achievable by current quantum processors [1, 2]. Noise can be broadly categorized as incoherent noise, e.g. stochastic noise originating in relaxation ( $T_1$ ) and dephasing ( $T_2$ ) events [3, 4], or coherent (unitary) noise, e.g. repeatable over/under rotations caused by miscalibrations or cross-talk [5, 6]. In superconducting qubit systems, incoherent noise is often attributed to microscopic materials defects [7, 8], quasi-particles [3, 9, 10], and coupling to environmental noise [10, 11] whereas coherent noise is often due to unwanted qubit-qubit interactions and microwave cross-talk from control lines [6, 12]. As such, the level of incoherent noise present in a device can generally not be improved once the device is fabricated and cooled down but coherent noise can often be mitigated by control techniques [13, 14], echo pulses [15–17], and noise tailoring [18, 19].

Importantly, the resulting algorithmic error scales differently depending on the type of noise present in the system. Coherent noise results in a worst-case quadratic error scaling in the circuit size (both in the length of the circuit and number of qubits), whereas incoherent noise results in a linear error increase [5, 19, 20]. Therefore, for near-term applications, the presence of coherent noise within a given system is more detrimental to the algorithmic fidelity, at least before fault-tolerant quantum computing is achieved [20]. Once coherence times allowing for fault-tolerant quantum computing are achieved, though, it will be paramount to ensure coherent errors are minimal and uncorrelated enough to maintain gate error rates below the fault-tolerance threshold.

While a mature suite of hardware benchmarks exists to characterize qubits and gates, these methods tend to scale poorly beyond roughly three qubits [21], at least at current coherence and control limits. For example,

variants of tomography can be used to fully characterize the quantum process and state but are susceptible to state preparation and measurement errors (SPAM) and the number of experiments required to characterize the system scales exponentially with the number of qubits. Randomized benchmarking [22] is currently the industry standard for measuring single and two-qubit gate errors and is robust to SPAM errors.

While standard RB gives only a single noise parameter, it can be modified to detect coherent errors [5, 23], characterize the error of a specific gate [23], as well as measure leakage out of the computational space [24]. In addition, recent work on RB allows for characterization of gates outside of the Clifford group [25–28]. Multi-qubit RB is an active area of research with recent results demonstrating that rigorous confidence intervals may be realized with a significantly smaller number of random sequences required than previously believed [29]. In the same work, the bounds were found to be asymptotically independent of the system size and an explicit relationship relating the variance of RB to the unitarity of the noise was derived [29], though it remains to be seen whether this relationship is useful as a practical characterization technique to measure coherent noise.

However, in practice implementing RB beyond a few qubits is difficult because the depth of a  $n$ -qubit Clifford is a quadratic function of  $n$  which limits the application of multi-qubit Cliffords in current technologies, and with typical gate errors, to at most three or four qubits [21]. Characterizations on larger numbers of qubits are of immediate practical interest in order to gauge the efficacy of near-term quantum computations. Additional noise sources may also be revealed in multi-qubit scenarios that may go undetected simply looking at the isolated one- and two-qubit error rates present in the system [21, 30].

For system-level characterization, the Quantum Volume (QV) [1, 2] remains the gold standard for hardware-agnostic performance benchmarking and incrementing QV signifies meaningful system improvements. We can ask, however, within a single step of QV, can we im-

---

\* jin@ibm.com

plement less demanding random circuits to extract fine-grained information which discriminates between coherent and incoherent noise? Here we introduce “hardware-efficient random circuits,” a family of circuits that exploits the native interactions available in the hardware, and implement a binning method introduced by Bouland *et al.* called “Binned Output Generation” (BOG) [31] which allows us to characterize the output of these circuits. We extend this method and show that by binning the same experimental results in two separate ways we can discriminate between coherent noise and incoherent noise in a multi-qubit system, thus extending noise classification beyond just the few qubits practically available hitherto.

In the following, we describe the hardware-efficient random circuits we implement as well as the binning process. We experimentally verify this technique in the two-qubit regime and compare the CNOT rate extracted from the BOG to the CNOT rate from simultaneous two-qubit RB. We then perform a six-qubit BOG experiment and extract an average CNOT error rate which is slightly higher than the average CNOT error rate extracted from the two-qubit experiment, indicating additional sources of error not detected by the two-qubit measurement. In this regime we re-bin the experimental results in order to extract only the incoherent error. The average incoherent error per CNOT agrees well with measurements of the purity error measured by purity RB [32]. Purity RB is a method of measuring how coherent the state is by applying state tomography to the end of a RB sequence and measuring the magnitude of the Bloch vector. Finally, we inject a purely coherent  $Z$  noise source [33] into the six-qubit circuits and extract average CNOT error rates consistent with the injected  $Z$  noise with an additional static  $ZZ$  component of the same order as the independently-measured  $ZZ$  interaction. Importantly, the measured incoherent noise remains constant during this experiment, demonstrating that our re-binning technique only detects the incoherent error and is robust to changes in coherent error.

Hardware-efficient random circuits are a family of circuits which respect the native connectivity and native entangling gate of the hardware. They consist of alternating layers of Haar-random single qubit rotations and two-qubit entangling gates between coupled qubits. For our architecture, the CNOT is the native entangling gate. Entangling gates are applied to alternating adjacent pairs of qubits, i.e. on one cycle qubit  $i$  would be entangled with its neighbor  $i + 1$ , and on the next cycle with  $i - 1$ . In this way, these circuits are agnostic to the details of the hardware architecture and thus can be applied to a variety of different systems.

In a perfect, noiseless system, with sufficiently many cycles of the aforementioned operations, the resulting quantum state will land at a random point within the Hilbert space. Projecting this state onto a measurement basis results in series of bitstrings whose probability distribution will tend towards a Porter-Thomas distribution

[34, 35]. For Hilbert spaces with dimension much larger than order unity, the Porter-Thomas distribution is well approximated by an exponential decay function.

The premise of the BOG, and other similar metrics like the cross-entropy [35] and the heavy output generation (HOG) [1, 2], is to pre-compute (with a classical computer) the ideal (i.e. noiseless) probability distribution that results from executing the circuit and compare this ideal distribution to the noisy distribution produced by the quantum hardware. The main benefit these methods provide is that these quantities can be estimated efficiently, in terms of quantum resources, with a non-exponential number of experiments [31, 35], though the classical pre-computation still requires exponential resources.

We compute a fidelity score for the BOG based on how far the noisy distribution is from the ideal (noiseless) distribution and normalize said fidelity taking into account the other limiting case of an incoherent mixture. This is fidelity is computed in the following way. Given a target circuit described by the unitary matrix  $\hat{U}$ , we define the outcomes for the ideal probability distribution as

$$p_i = |\langle x_i | \hat{U} | 0 \rangle|^2, \quad (1)$$

where  $x_i \in \mathbb{Z}^n$  denote the possible output bitstrings from the circuit for  $n$  qubits. In the presence of noise, the experimentally measured probabilities  $q_i$  will deviate from the ideal probabilities.

To estimate the noisy probability distribution using the BOG, the probability space spanned by  $[0, 1]$  is divided into  $\text{poly}(n)$  number of bins [31]. The distribution is reconstructed by adding each experimentally measured probability  $q_i$  to the bin which contains the bitstring  $x_i$ 's ideal probability  $p_i$  (algorithm 1). For each bin  $[a/N, b/N]$ , where  $N = 2^n$ , the bin edges are constructed such that  $\int_a^b q e^{-Nq} = \Theta(1/\text{poly}(n))$  [31]. This ensures that for an ideal Porter-Thomas distribution, the weights in each bin are equal. It is worth noting that the cross-entropy and HOG can be viewed as two limiting cases of the BOG [31]. For the HOG, outcomes are sorted into just two bins (heavy or not heavy) and for the cross-entropy, each outcome is placed into its own bin. The BOG maximizes the amount of information obtained from the pre-computed distribution and can rule out “imposter” distributions [31].

To give an example of the binning procedure, suppose the all 0 bitstring's ( $x_0$ ) ideal probability  $p_0=0.05$  and the first two bins ( $\text{bin}_0$  and  $\text{bin}_1$ ) span  $[0,0.1]$  and  $[0.1,0.2]$ , respectively. Since the ideal probability of  $x_0$  is contained in  $\text{bin}_0$ , the experimentally measured probability  $q_0$  is added to  $\text{bin}_0$ . In the noiseless case, the measured experimental probability will equal the ideal probability ( $q_0=0.05$ ). Suppose now in the presence of noise, the experimental probability  $q_0=0.15$ . In this case, the value of 0.15 is added to  $\text{bin}_0$ . In this way the weights of the bins now deviate away from the ideal case and the noise

(both incoherent and coherent) is detected. This process is repeated for each bitstring.

---

**Algorithm 1** Binned Output Generation
 

---

```

1: for  $x_i$  in  $x_{all}$  do
2:    $p_i$  = ideal_probabilities [ $x_i$ ]
3:    $q_i$  = experimental_probabilities [ $x_i$ ]
4:   bin_index = assign_bin( $p_i$ )
5:   bins [bin_index] +=  $q_i$ 
6: end for
  
```

---



---

**Algorithm 2** Binned Output Generation for Incoherent Errors
 

---

```

1: for  $x_i$  in  $x_{all}$  do
2:    $q_i$  = experimental_probabilities [ $x_i$ ]
3:   bin_index = assign_bin( $q_i$ )
4:   bins [bin_index] +=  $q_i$ 
5: end for
  
```

---

To convert the measured frequencies to a fidelity we define the following expression

$$\text{Fidelity} = \left\langle 1 - \frac{|\mathbf{bins}(\psi_{\text{ideal}}) - \mathbf{bins}(\rho_{\text{exp}})|_1}{|\mathbf{bins}(\psi_{\text{ideal}}) - \mathbf{bins}(\mathbb{I}/2^n)|_1} \right\rangle_{\hat{U}}, \quad (2)$$

that is the 1-norm distance of the binning of the experimental data ( $\rho_{\text{exp}}$ ) from the binning of the ideal output, normalized by the difference of the ideal binning and the maximally mixed state. In other words, it defines a normalized distance between the ideal state's distribution and the maximally mixed distribution that the experimental state achieves. By normalizing in this manner, we enforce that a noiseless result achieves a score of 1 and a fully depolarized result decays to 0. The average is taken over typical circuits  $\hat{U}$ .  $\mathbf{bins}$  is the array of summed probabilities computed by algorithms 1 and 2. We compute the weights of  $\mathbf{bins}(\psi_{\text{ideal}})$  with the pre-computed ideal probabilities  $p_i$  and  $\mathbf{bins}(\mathbb{I}/2^n)$  with the maximally mixed state  $\mathbb{I}/2^n$ . This method is sensitive to both incoherent and coherent errors.

We now extend the binning procedure to only detect incoherent errors. Instead of using the pre-computed ideal outcome probability to assign a bin to each measured bitstring, we bin the experimental outputs according to the experimentally measured probabilities (algorithm 2). The same experimental data used for algorithm 1 is simply rebinned and a separate experiment need not be performed. Using the aforementioned example probabilities, supposing  $x_0$ 's ideal probability  $p_0=0.05$  but the experimental probability measured is  $q_0=0.15$ , the experimental probability measured would be added to  $bin_1$ , the bin containing the *experimental probability*  $q_0$ , instead of the  $bin_0$ , the bin containing the *ideal probability*  $p_0$  as in algorithm 1. The intuition behind this binning strategy is that since the Haar measure is invariant under unitary operations, the resulting measurement after a purely unitary error is still described by a Porter-Thomas distribution, just not the specific Porter-Thomas distribution

as derived from the pre-computed quantum state. In contrast, incoherent errors in conjunction with random single-qubit gates will drive each bitstring probability to  $1/2^n$ .

For this binning strategy, since the measured outcomes are not pre-assigned a bin, the weights of  $\mathbf{bins}(\psi_{\text{ideal}})$  are computed by integrating the ideal Porter-Thomas distribution and taking the integration bounds to be the edges of each bin, i.e.  $\int_a^b q e^{-Nq}$ . The weights of  $\mathbf{bins}(\mathbb{I}/2^n)$  are computed similarly with the exception of integrating over a normalized Gaussian centered about  $1/2^n$  whose width is proportional to  $1/\sqrt{2^n}$  shots, i.e.  $\int_a^b q \exp[-\frac{1}{2}(\frac{q-1/2^n}{\sigma})^2]$ , where  $\sigma = 1/\sqrt{2^n}$  shots. A Gaussian is used here to account for shot noise, which is ubiquitous given a finite number of shots.

A simple example to give intuition for this method is to apply a bit flip error to each qubit at the end of the quantum circuit. The overall shape of the distribution is still a Porter-Thomas, but the labels of the bitstrings have been permuted. Binning these measurements according to the pre-computed quantum state, therefore, would yield a low fidelity. Binning by the experimental probabilities however, still yields a high fidelity as this methodology is blind to coherent errors.

We execute all of our experiments on a six qubit linear chain on *ibmq.dublin* (Fig. 1 (a)), a 27 qubit heavy-hex superconducting qubit processor. We first compare the results of two-qubit pairwise BOG and compare the CNOT error rate extracted in this way to the CNOT error rates measured by simultaneous RB [36]. These qubits are fixed frequency transmon with fixed nearest neighbor coupling where the native entangling gate is a CNOT mediated by an echoed cross-resonance microwave (CR) interaction [6, 37]. We execute the circuit shown in Fig. 1 (b) among adjacent pairs in the chain, executing pair (Q2, Q3) and (Q5, Q8) simultaneously in one experiment and (Q1, Q2), (Q3, Q5), and (Q8, Q11) simultaneously in another experiment. By measuring adjacent qubit pairs simultaneously, cross-talk errors from applying simultaneous CR tones as well as spectator errors can be detected [16, 36].

The experiments were repeated with 90 different seeds to generate 90 distinct circuits with unique random single-qubit rotations and each circuit was executed with 1000 shots. The shots from each experimental outcome were binned into 10 bins and summed over all seeds. As a rule of thumb, the number of seeds times the number of basis states in the Hilbert space should be much larger than the number of bins to build good statistics. The depths of the circuits were varied up to 270 cycles and the resulting fidelity curves were each fit to a single exponential decay of the form  $Ae^{-\lambda x} + B$  in order to extract the error per cycle. We assume a two-qubit depolarizing noise channel such that the exponential decay rate is equal to the depolarizing parameter  $\lambda$  and extract an error per gate equal to  $\frac{3}{4}\lambda$ , where the prefactor is due to the dimensionality of a two-qubit depolarizing channel [22]. While a full proof that the resulting noise

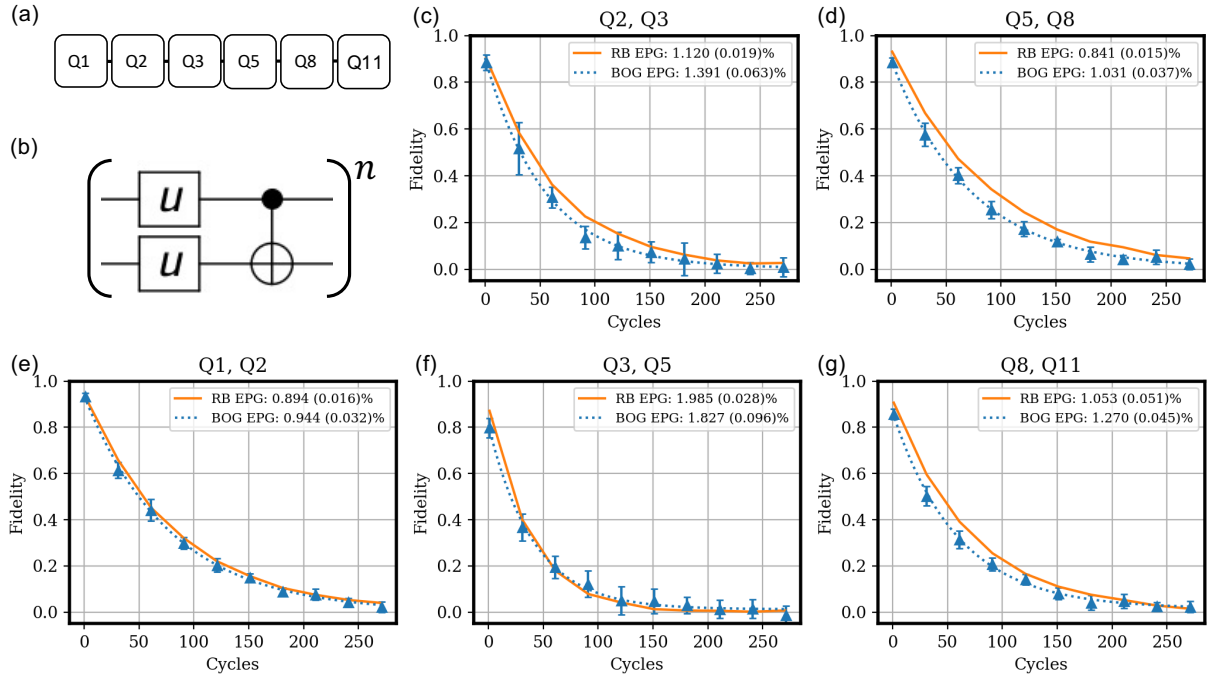


FIG. 1. (a) Connectivity diagram of the qubits used in this work. (b) One cycle of the two-qubit BOG circuit. Each  $u$  gate denotes a unique Haar-random single-qubit rotation, randomized for each gate for a given circuit seed. (c)-(g) Plots of the two-qubit BOG fidelity as a function of the circuit depth. The blue dotted lines denote exponential fits to the data. Error bars are computed by bootstrapping 10 groups of nine seeds together and computing the standard deviation of these values. Orange lines denote noisy simulations of the BOG circuits where the only noise parameters included are the readout error and the CNOT error measured by RB. Single-qubit errors are omitted from the simulation as the single-qubit error rate is typically one to two orders of magnitude lower than the CNOT error.

channel is a depolarizing channel is outside the scope of this work, it is likely a safe assumption given the Haar-random single-qubit gates and the averaging over different random seeds. In addition, by extracting an error per cycle in this way, the method is robust to SPAM errors, similar to RB.

Fig. 1 (c)-(g) illustrates the CNOT error rates extracted by the BOG compared to the CNOT error rates measured by simultaneous RB the same day. We measure an average CNOT error rate of 1.292 (0.054)% per gate from the BOG in comparison to 1.178 (0.025)% per gate from RB, demonstrating agreement within 10% between the two techniques.

Having established agreement between the RB and BOG in the two-qubit case, we now consider a six-qubit BOG experiment. In this regime, RB is difficult to perform because the decomposition of multi-qubit Clifford gates results in many more native entangling gates. Moreover, experiments beyond the two-qubit regime have been shown to capture error mechanisms absent from single and two-qubit measurements [21, 30].

The structure of the circuits executed are illustrated in Fig. 2 (a). Random single qubit rotations are applied to each qubit before applying CNOTs in parallel to adjacent qubit pairs, alternating each cycle which qubit pairs perform the CNOT, thus exploiting the native connectivity

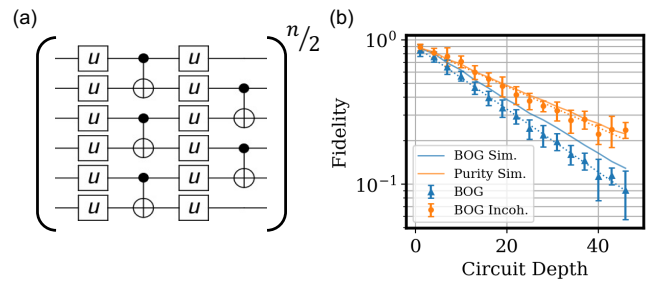


FIG. 2. (a) Six-qubit BOG circuits. Haar-random single-qubit rotations are applied to each qubit followed by CNOTs. Alternating layers of CNOTs are applied after single qubit rotations. (b) Data for the six-qubit BOG circuits. Data are binned according to quantum state (blue triangles) and are fit to an exponential to extract an average CNOT rate of 1.502 (0.022)%. The blue solid line is a noisy simulation of the circuit with the CNOT errors measured from the two-qubit BOG circuits. The orange circles are the data binned by the experimental probability which detects only incoherent errors. Fitting these data to an exponential yields an average incoherent error rate of 0.980 (0.038)%. The solid orange line is noisy simulation using error rates measured from purity RB.

	2Q RB	2Q BOG	2Q BOG (barriers)	6Q BOG
Avg. 2Q error rate	1.178 (0.025)%	1.292 (0.054)%	1.340 (0.054)%	1.502 (0.022)%

TABLE I. Average gate errors comparing between different benchmarking methods. Two-qubit BOG agrees within 10% of two-qubit RB. Adding barriers between cycles in the two-qubit bog circuits enforces the CNOTs between different qubits to start simultaneously each cycle, resulting in idling qubits since CNOTs between different qubits have variable lengths. The resulting idle time increases the average gate error. The six-qubit BOG detects a higher average effective error rate due to unavoidable idle times during the alternating cycles of CNOTs.

of our device via a hardware-efficient circuit structure. In general, any native entangling 2-qubit gate can be used in these circuits. These experiments are averaged over 40 random number seeds, repeated with 8000 shots each and the results were binned into 30 bins. Error bars are generated by calculating the standard deviation of eight groups of five seeds averaged together. Figure 2 (b) illustrates the fidelity of the six-qubit system as a function of the circuit depth, defined as one cycle of single qubit rotations and one cycle of parallel CNOTs. The blue points are the BOG fidelity binned by the pre-computed quantum state (algorithm 1), which captures both incoherent and coherent noise. By fitting this curve to a single exponential decay, we can relate the decay rate of the exponential,  $\lambda$ , to an average gate error ( $EPG$ ) by  $EPG = \frac{3}{4} \frac{2}{5} \lambda$ . Here again the prefactor of  $\frac{3}{4}$  is due to the dimensionality of a two-qubit depolarizing channel. The second prefactor  $\frac{2}{5}$  is to account for the five CNOT gates per two circuit layers. In this six-qubit circuit we extract an average CNOT error rate of 1.502 (0.022)%, notably higher than the average CNOT error rates measured in the two-qubit case, 1.292 (0.054)%. These error rates are summarized in table I.

This increase in average error rate is likely due to the additional qubit idling time when the six-qubit BOG circuits are executed in comparison to the six-qubit BOG circuits. Idling qubits are introduced in two separate instances in the six-qubit BOG circuits. First, because the CNOT pairs now alternate between adjacent qubit pairs (Fig. 2 (a)) and because the CNOT lengths are different, the qubit pairs with the fastest CNOTs must wait for the slower CNOTs to finish executing before the next layer of the circuit can be executed. This is in contrast to the two-qubit BOG circuits (Fig. 1 (b)) where the start of the execution of a CNOT between qubits is not constrained to the timing of the adjacent CNOTs. Indeed, repeating the two-qubit BOG experiments with barriers inserted between the CNOT layers, which enforces the simultaneous timing of adjacent CNOTs, supports this conclusion resulting in a higher average CNOT error rate of 1.340 (0.054)% (not shown).

A second qubit idle period occurs when Q1 and Q11

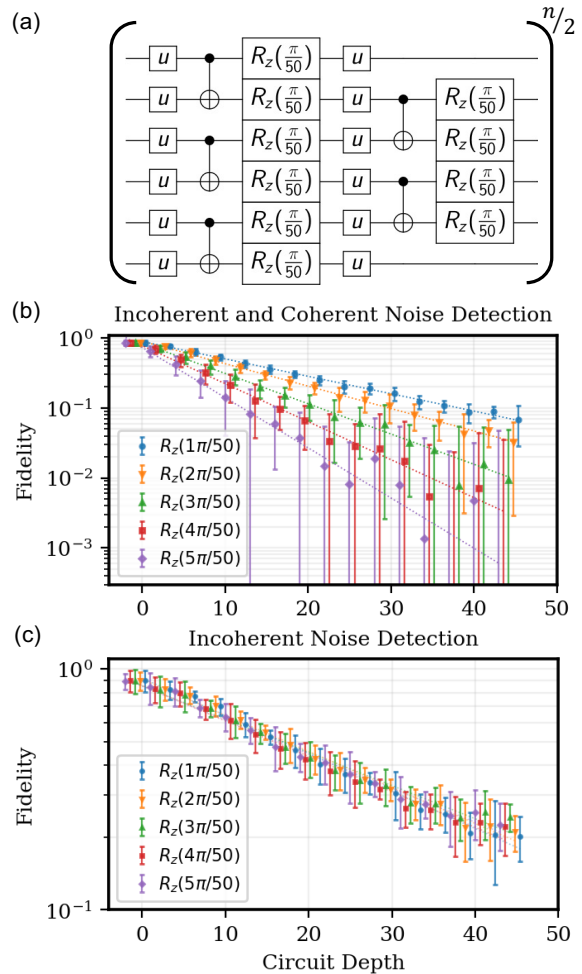


FIG. 3. (a) Six-qubit BOG circuits with injected  $Z$  noise following each CNOT. Virtual rotations of 1, 2, 3, 4, and 5% of  $2\pi$  radians are injected as a purely coherent noise source. (b) Fidelity decay rates of the  $Z$  noise injected circuits, binned by the precomputed quantum state (algorithm 1), demonstrating a spread in decay rates due to the increase in coherent noise. Markers are offset in the x direction for clarity. (c) Fidelity decay rates of the same coherent noise injected circuits, binned by the experimental probability (algorithm 2). Binning by this method only detects incoherent noise, and as such the decay rates of the five curves are identical. Markers have been offset in the x direction for clarity.

are left idle during the even cycles of the six-qubit BOG circuits. This another error which is not captured by two-qubit measurements. On these cycles, CNOTs are only applied to the inner two pairs of qubits, (Q2, Q3) and (Q5, Q8). This highlights the ability to detect errors that are likely to occur during the execution of a realistic algorithm, but are not evident by one and two-qubit benchmarking.

We apply our second binning method to this six-qubit experiment (Fig. 2 (b)) in order to extract the aver-

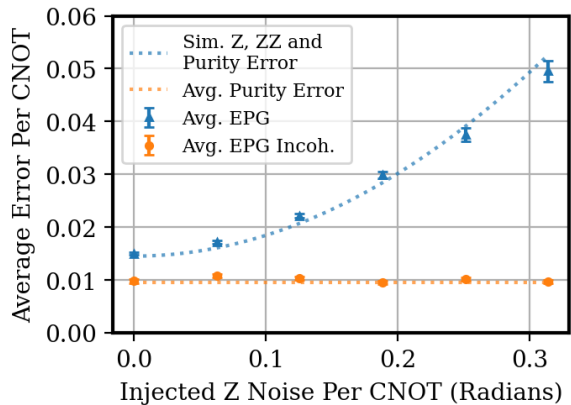


FIG. 4. Extracted error average CNOT error rates from the six-qubit BOG experiments with injected coherent  $Z$  noise. As the magnitude of the injected  $Z$  noise increases, the average CNOT error increases quadratically, as the average incoherent CNOT error remains constant. Simulation of the gate (blue dotted line) with the injected  $Z$  noise with an additional best fit  $ZZ$  noise of 56.7 kHz is also shown. Additionally, the average purity error per gate is shown (orange dotted line). The difference in the curves indicates the amount of residual coherent noise.

age incoherent error per gate and compare this figure to the average two-qubit purity error as measured by purity RB [32]. We demonstrate good agreement between the average incoherent error per gate as measured by the six-qubit BOG experiment, 0.980 (0.038)%, compared with the average purity error per gate, 0.958 (0.023)%. We note to build good statistics for this method, the number of shots taken must be much larger than the number of basis states in the Hilbert space.

Finally, we repeat the six-qubit experiment while injecting a purely coherent  $Z$  noise to each qubit each time a CNOT is applied to demonstrate the BOG is able to effectively discriminate between coherent and incoherent noise. After each CNOT, a small  $Z$  rotation from 1% to 5% of a  $2\pi$  rotation is applied to the control and target qubits as a virtual rotation (Fig. 3 (a)). This injected noise takes 0 time as it is implemented in software as a reorientation of the  $x$ -axis of the Bloch sphere and can be executed with near perfect fidelity [33]. We note that the injected  $Z$  noise of the last cycle will be undetected in the measurement basis of our experiment. This small error should not affect the extracted error per gate since this error is equivalent to a SPAM error.

When binning by the pre-computed quantum state (Algorithm 1), we observe an obvious increase in the decay rate of the fidelity curve which increases with the magnitude of the injected coherent noise (Fig. 3 (b)), as this method is sensitive to both incoherent and coherent errors. With increasing levels of coherent noise we observe a wider spread in the error bars. This effect is qualitatively consistent with calculations of the variance of RB with coherent noise [29] and is due to the con-

structive and destructive interference of unitary noise. When binning by the experimentally measured probabilities (Algorithm 2), the resulting decay rates remain constant as the injected noise is purely coherent and binning by this method only detects incoherent errors (Fig. 3 (b)).

The extracted error rates from Fig. 3 are summarized in Fig. 4. In addition to the extracted average error per gate and incoherent error per gate, we plot the average purity error, measured from purity RB, and simulations of the average error per gate with coherent noise. As expected, the error per gate increases quadratically as the magnitude of the  $Z$  noise increases, a hallmark of coherent noise [5, 19]. The difference of these two curves indicates the relative amount of coherent noise in the gate error. Notably at 0 injected  $Z$  noise, there is a discrepancy between the average gate error and the incoherent gate error, indicating some amount of coherent noise that is unaccounted for. We model this residual noise as a static  $ZZ$  interaction between the control and target qubit, applied after each CNOT. Fitting this  $ZZ$  interaction with the average CNOT gate time of 443.73 ns, we extract an effective  $ZZ$  interaction strength of 56.7 kHz, the same order of magnitude as the average independently-measured static  $ZZ$  interaction of 24.4 kHz. With this simple model we capture the scaling behavior of the experimental data, though this effective  $ZZ$  strength is likely an overestimate since it only considers the  $ZZ$  interaction between the control and target qubits. The model omits the additional  $ZZ$  interactions from nearest-neighbor spectator qubits and other higher weight interactions [30], as well as the mitigating effects of echo pulses applied during the CNOT [16]. Therefore, the  $ZZ$  interaction we extract is likely some sum of all of these effects and our noise model is the simplest non-local model which reasonably reflects the experimental data.

In conclusion, we have demonstrated a hardware-efficient multi-qubit metric capable of discriminating between coherent and incoherent noise, a regime inaccessible to current benchmarking techniques. This technique can be used as a diagnostic tool for different hardware platforms as the BOG accommodates different native two-qubit entangling gates and requires no special qubit connectivity or layout. It can determine whether the algorithmic performance of a quantum computing system is limited by its coherence or by the quality of the control and calibrations. If the system fidelity is limited by unitary errors, it can determine how much of the error can potentially be corrected by improvements in control. In addition, it can detect error sources that are only evident in multi-qubit experiments (namely true quantum algorithms), errors that may not be evident from standard two-qubit benchmarks. As we move towards circuits for error correcting codes, the ability to detect coherent vs. incoherent errors is critical. There are also further open questions about directly connecting the metrics in this manuscript, or any metrics derived from random circuits, to the performance of large scale codes. A possible

future extension is to replace the random unitary operators in our circuit construction with random elements of the Clifford group, producing circuits that are analogous to those used in stabilizer measurements.

This work was sponsored by the Army Research Office and was accomplished under Grant Numbers W911NF-14-1-0124 and W911NF-21-1-0002. The views and con-

clusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein

- 
- [1] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Gnllk, T. Itoko, N. Kanazawa, A. Kandala, G. A. Keefe, K. Krsulich, W. Landers, E. P. Lewandowski, D. T. McClure, G. Nannicini, A. Narasgond, H. M. Nayfeh, E. Pritchett, M. B. Rothwell, S. Srinivasan, N. Sundaresan, C. Wang, K. X. Wei, C. J. Wood, J.-B. Yau, E. J. Zhang, O. E. Dial, J. M. Chow, and J. M. Gambetta, Demonstration of quantum volume 64 on a superconducting quantum computing system (2020), arXiv:2008.08571 [quant-ph].
- [2] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
- [3] G. Catelani, S. E. Nigg, S. M. Girvin, R. J. Schoelkopf, and L. I. Glazman, Decoherence of superconducting qubits caused by quasiparticle tunneling, *Phys. Rev. B* **86**, 184514 (2012).
- [4] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, *Applied Physics Reviews* **6**, 021318 (2019), <https://doi.org/10.1063/1.5089550>.
- [5] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Characterizing errors on qubit operations via iterative randomized benchmarking, *Phys. Rev. A* **93**, 012301 (2016).
- [6] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Procedure for systematically tuning up cross-talk in the cross-resonance gate, *Phys. Rev. A* **93**, 060302 (2016).
- [7] J. Lisenfeld, A. Bilmes, S. Matityahu, S. Zanker, M. Marthaler, M. Schechter, G. Schön, A. Shnirman, G. Weiss, and A. V. Ustinov, Decoherence spectroscopy with individual two-level tunneling defects, *Scientific Reports* **6**, 23786 (2016).
- [8] A. P. M. Place, L. V. H. Rodgers, P. Mundada, B. M. Smitham, M. Fitzpatrick, Z. Leng, A. Premkumar, J. Bryon, S. Sussman, G. Cheng, T. Madhavan, H. K. Babla, B. Jaeck, A. Gyenis, N. Yao, R. J. Cava, N. P. de Leon, and A. A. Houck, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds (2020), arXiv:2003.00024 [quant-ph].
- [9] S. E. de Graaf, L. Faoro, L. B. Ioffe, S. Mahashabde, J. J. Burnett, T. Lindström, S. E. Kubatkin, A. V. Danilov, and A. Y. Tzalenchuk, Two-level systems in superconducting quantum devices due to trapped quasiparticles, *Science Advances* **6**, 10.1126/sciadv.abc5055 (2020).
- [10] R. Barends, J. Wenner, M. Lenander, Y. Chen, R. C. Bialczak, J. Kelly, E. Lucero, P. O'Malley, M. Mariantoni, D. Sank, H. Wang, T. C. White, Y. Yin, J. Zhao, A. N. Cleland, J. M. Martinis, and J. J. A. Baselmans, Minimizing quasiparticle generation from stray infrared light in superconducting quantum circuits, *Applied Physics Letters* **99**, 113507 (2011), <https://doi.org/10.1063/1.3638063>.
- [11] C. Rigetti, J. M. Gambetta, S. Poletto, B. L. T. Plourde, J. M. Chow, A. D. Córcoles, J. A. Smolin, S. T. Merkel, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, Superconducting qubit in a waveguide cavity with a coherence time approaching 0.1 ms, *Phys. Rev. B* **86**, 100506 (2012).
- [12] M. Malekakhlagh, E. Magesan, and D. C. McKay, First-principles analysis of cross-resonance gate operation, *Phys. Rev. A* **102**, 042605 (2020).
- [13] H. Ball, M. J. Biercuk, A. Carvalho, J. Chen, M. Hush, L. A. D. Castro, L. Li, P. J. Liebermann, H. J. Slatyer, C. Edmunds, V. Frey, C. Hempel, and A. Milne, Software tools for quantum control: Improving quantum computer performance through noise and error suppression (2020), arXiv:2001.04060 [quant-ph].
- [14] M. H. Goerz, F. Motzoi, K. B. Whaley, and C. P. Koch, Charting the circuit qed design landscape using optimal control theory, *npj Quantum Information* **3**, 37 (2017).
- [15] K. X. Wei, I. Lauer, S. Srinivasan, N. Sundaresan, D. T. McClure, D. Toyli, D. C. McKay, J. M. Gambetta, and S. Sheldon, Verifying multipartite entangled greenberger-horne-zeilinger states via multiple quantum coherences, *Phys. Rev. A* **101**, 032343 (2020).
- [16] N. Sundaresan, I. Lauer, E. Pritchett, E. Magesan, P. Jurcevic, and J. M. Gambetta, Reducing unitary and spectator errors in cross resonance with optimized rotary echoes, *PRX Quantum* **1**, 020318 (2020).
- [17] A. Kandala, K. X. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. A. Keefe, D. Klaus, O. Dial, and D. C. McKay, Demonstration of a high-fidelity cnot for fixed-frequency transmons with engineered zz suppression (2020), arXiv:2011.07050 [quant-ph].
- [18] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor (2020), arXiv:2010.00215 [quant-ph].
- [19] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, *Phys. Rev. A* **94**, 052325 (2016).
- [20] J. K. Iverson and J. Preskill, Coherence in logical quantum channels, *New Journal of Physics* **22**, 073066 (2020).
- [21] D. C. McKay, S. Sheldon, J. A. Smolin, J. M. Chow, and J. M. Gambetta, Three-qubit randomized benchmarking, *Phys. Rev. Lett.* **122**, 200502 (2019).
- [22] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum pro-



- cesses, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [23] 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, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [24] R. W. Andrews, C. Jones, M. D. Reed, A. M. Jones, S. D. Ha, M. P. Jura, J. Kerckhoff, M. Levendorf, S. Meenehan, S. T. Merkel, A. Smith, B. Sun, A. J. Weinstein, M. T. Rakher, T. D. Ladd, and M. G. Borselli, Quantifying error and leakage in an encoded si/sige triple-dot qubit, *Nature Nanotechnology* **14**, 747 (2019).
- [25] D. S. França and A. K. Hashagen, Approximate randomized benchmarking for finite groups, *Journal of Physics A: Mathematical and Theoretical* **51**, 395302 (2018).
- [26] E. Derbyshire, R. Mezher, T. Kapourniotis, and E. Kashefi, Randomized benchmarking with stabilizer verification and gate synthesis (2021), arXiv:2102.13044 [quant-ph].
- [27] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, A new class of efficient randomized benchmarking protocols, *npj Quantum Information* **5**, 71 (2019).
- [28] S. Garion and A. W. Cross, Synthesis of CNOT-Dihedral circuits with optimal number of two qubit gates, *Quantum* **4**, 369 (2020).
- [29] J. Helsen, J. J. Wallman, S. T. Flammia, and S. Wehner, Multiqubit randomized benchmarking using few samples, *Phys. Rev. A* **100**, 032304 (2019).
- [30] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, Correlated randomized benchmarking (2020), arXiv:2003.02354 [quant-ph].
- [31] A. Bouland, B. Fefferman, C. Nirkhe, and U. Vazirani, On the complexity and verification of quantum random circuit sampling, *Nature Physics* **15**, 159 (2019).
- [32] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, *New Journal of Physics* **17**, 113020 (2015).
- [33] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient  $z$  gates for quantum computing, *Phys. Rev. A* **96**, 022330 (2017).
- [34] C. E. Porter and R. G. Thomas, Fluctuations of nuclear reaction widths, *Phys. Rev.* **104**, 483 (1956).
- [35] C. Neill, P. Roushan, K. Kechedzhi, S. Boixo, S. V. Isakov, V. Smelyanskiy, A. Megrant, B. Chiaro, A. Dunsworth, K. Arya, R. Barends, B. Burkett, Y. Chen, Z. Chen, A. Fowler, B. Foxen, M. Giustina, R. Graff, E. Jeffrey, T. Huang, J. Kelly, P. Klimov, E. Lucero, J. Mutus, M. Neeley, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, H. Neven, and J. M. Martinis, A blueprint for demonstrating quantum supremacy with superconducting qubits, *Science* **360**, 195 (2018).
- [36] J. M. Gambetta, A. D. Croles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, and et al., Characterization of addressability by simultaneous randomized benchmarking, *Physical Review Letters* **109**, 10.1103/physrevlett.109.240504 (2012).
- [37] E. Magesan and J. M. Gambetta, Effective hamiltonian models of the cross-resonance gate, *Phys. Rev. A* **101**, 052308 (2020).