

この原稿は、CHORUSを通じて公開されたアクセプト原稿です。として掲載されました：

Three-Qubit Randomized Benchmarking

デビッド・C・マッケイ、サラ・シェルドン、ジョン・A・スモリン、ジェリー・M・チョウ、ジェイ・M・ガンベッタ

Phys. Rev. Lett. **122**, 200502 — Published 23 May 2019

DOI: [10.1103/PhysRevLett.122.200502](https://doi.org/10.1103/PhysRevLett.122.200502)

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

Three-Qubit Randomized Benchmarking

David C. McKay, Sarah Sheldon, John A. Smolin, Jerry M. Chow, and Jay M. Gambetta

Phys. Rev. Lett. **122**, 200502 — Published 23 May 2019

DOI: [10.1103/PhysRevLett.122.200502](https://doi.org/10.1103/PhysRevLett.122.200502)

Three Qubit Randomized Benchmarking

David C. McKay,* Sarah Sheldon, John A. Smolin, Jerry M. Chow, and Jay M. Gambetta
IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

(Dated: May 3, 2019)

量子回路のサイズが大きくなるにつれて、スケーラブルなマルチ量子ビット忠実度メトリックを確立することが重要である。本論文では、マイクロ波による相互作用（交差共鳴）を持つ3つの固定周波数トランスモン量子ビットからなる量子デバイスにおいて、3量子ビットランダム化ベンチマーク（RB）を初めて調査した。その結果、クリフォードあたりの3量子ビット誤差は、全対全ゲート接続で0.106、線形ゲート接続で0.207と測定されました。さらに、混合次元の同時RB（1量子ビットと2量子ビットの同時RB）を導入することで、3量子ビット誤差が1量子ビットと2量子ビットの誤差から予測できることを示すことができた。しかし、ゲートに特定のコヒーレントな誤差を導入することで、3量子ビット誤差を0.302まで増加させることができました。この増加は、同時RBによる1量子ビットと2量子ビット誤差の増加に比例して予測されるものではありません。このことは、3量子ビットRBのような多量ビット評価指標が、デバイス全体の性能を評価する上で重要であることを示しています。

量子回路の規模が大きくなるにつれて、特性解析の問題はより深刻になる。量子ビットの数だけ状態空間が指数関数的に増大するため、システムを再構築するためのトモグラフィーの手法には指数関数的なリソースが必要になります。実際、量子プロセスのトモグラフィーに必要な測定数は、 16^n [1] (n は量子ビットの数)とスケーリングされることが分かっています。スケーリングの問題を回避するために、ユニバーサルゲートセットを構築するために使用されるゲートの原始的なセットを特徴付けることに重点を置いた方法があります。最低限、 n 個の量子ビットに対して、この集合は、すべての n 個の量子ビットに対するいくつかの1量子ビットゲートと $n-1$ 個の2量子ビットゲートを含んでいます[2]。しかし、多ビットアルゴリズムの忠実度は、これらのプリミティブゲートの忠実度によって予測されるという仮定は、どの程度良いのでしょうか？クロストークやアドレス指定エラーによって、この仮定が破綻することが強く示唆されています。例えば、サーフェスコードのアルゴリズムでは、多ビット回路に2クビットCNOTゲートを並列に順次適用し、ローカルな5クビットゲートを構築する必要があります。サーフェスコードはエラー訂正の閾値が高いと予測されていますが、通常、直接ゲートが存在する量子ビット間でのみ相関ノイズが発生するシミュレーションが行われています[3]。最近行われた論理量子ビットの5量子ビットテストでは、2量子ビットゲートの間に観客（すなわち、非参加の隣接量子ビット）に対するZZ項を補償することによって、忠実度が大幅に改善されました[4]。さらに、アルゴリズムとプリミティブゲートの忠実度が必ずしも一致しないことが、いくつかの研究で観察されています。例えば、2つの異なる5量子ビットプロセッサで4つのアルゴリズムを実行したところ、プリミティブからアルゴリズムへの忠実度への決定的な一致は見られませんでした[5]。2量子ビットゲートフィデリティが0.99の5量子ビットデバイスにおいて、4つの2量子ビットゲートを適用した後の5量子ビットGHZの状態フィデリティは0.82でした[6]。したがって、真のアルゴリズム忠実度を予測するためには、マルチクビット忠実度のメトリクスを測定する必要があります。

幸いなことに、平均ゲートフィデリティのような少数の指標に基づいてプロセスを特徴付けることを目的とする場合、スケーリングの問題は回避することができます。この考えに基づき、モンテカルロサンプリング[7, 8]、圧縮センシング[9]、行列積状態トモグラフィー[10]、ツワーリングプロトコル[11]などの手法が提案されており、光子[12]、NMR [13]、トラップイオン [14] など様々な多ビット系で適用されています。

さらに、10[15]および12[16]量子ビット状態で実証されたように、特定の多ビットもつれ状態の忠実度を効率的に測定することができる[16]。しかし、これらの技術に共通する欠点は、結果が準備誤差や測定誤差に敏感であること（時には指数関数的に大きくなる）、また、基礎となるゲートを完全に特徴付けることができないことです。これらの問題は、ランダムベンチマーク[17, 18]（RB）によって解決されます。これは、同一演算子に等しいランダムクリフォードゲートのシーケンスが、一連の量子ビットに適用されます。シーケンスの長さに対する量子ビットの分極の減衰は、準備や測定の誤差に依存しないクリフォードセットの平均的な忠実度を測定します。RBは、超伝導回路 [6, 19–21]、イオントラップ [17, 22–24]、中性原子トラップ [25]、NMRシステム [26]、固体スピン量子ビット [27] のゲート特性評価に広く用いられている方法である。また、インターリーブ[28]、純度[29, 30]、リーク[31, 32]による特定のゲートエラーを測定するためにRBの拡張が提案・実装されています。

RBは、2つの方法で多ビットシステムの忠実度に対処するように設計されている。一つは、RBは n -qubitクリフォード群からシーケンスを構築することで実行できる。さらに、 n -qubit空間をqubitの集合 $\{n_i\}$ に細分化し、各分割集合で同時に n_i -qubit RBを実行することができる[33]。どちらの方法も n -qubit空間における忠実度のメトリクスを与える。この2つの方法があるにもかかわらず、 $n > 2$ のRBについては、原始的なゲートセットのみを特徴付けることで十分であると考えられているため、これまで実証されていませんでした。本論文では、3量子ビット固定周波数超伝導デバイスにおいて、様々な3量子ビットRBの組み合わせを初めて示す。全ゲート接続の場合、クリフォードあたりの3量子ビット誤差（3Q EPC）は0.106であり、これは同時RBによる原始ゲート誤差からよく予測される。

Three Qubit Randomized Benchmarking

David C. McKay,^{*} Sarah Sheldon, John A. Smolin, Jerry M. Chow, and Jay M. Gambetta
IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

(Dated: May 3, 2019)

As quantum circuits increase in size, it is critical to establish scalable multiqubit fidelity metrics. Here we investigate, for the first time, three-qubit randomized benchmarking (RB) on a quantum device consisting of three fixed-frequency transmon qubits with pairwise microwave-activated interactions (cross-resonance). We measure a three-qubit error per Clifford of 0.106 for all-to-all gate connectivity and 0.207 for linear gate connectivity. Furthermore, by introducing mixed dimensionality simultaneous RB — simultaneous one- and two-qubit RB — we show that the three-qubit errors can be predicted from the one- and two-qubit errors. However, by introducing certain coherent errors to the gates we can increase the three-qubit error to 0.302, an increase that is not predicted by a proportionate increase in the one- and two-qubit errors from simultaneous RB. This demonstrates the importance of multiqubit metrics, such as three-qubit RB, on evaluating overall device performance.

As quantum circuits increase in size, the problem of characterization becomes more acute. Exponential growth of the state space with the number of qubits means that tomographic methods for reconstructing the system will require exponential resources. Indeed, the number of required measurements for quantum process tomography scales as 16^n [1] where n is the number of qubits. To avoid scaling issues, methods have focused on characterizing the primitive set of gates used to construct the universal gateset. At minimum, for n qubits, this set contains several one-qubit gates for all n qubits and $n-1$ two-qubit gates [2]. But how good is the assumption that multiqubit algorithmic fidelities will be predicted by these primitive gate fidelities? There are strong indications that this assumption fails due to crosstalk and addressability errors. For example, surface code algorithms require constructing local five-qubit gates via sequential application of two-qubit CNOT gates in parallel across a multiqubit circuit. Surface codes are predicted to have a high threshold for correcting errors, but they are typically simulated with correlated noise only between qubits for which there is a direct gate [3]. In a recent five-qubit test of a logical qubit, the fidelity was greatly improved by compensating for ZZ terms to spectator (i.e., non-participating neighboring) qubits during the two-qubit gate [4]. In addition, several studies have observed that algorithmic and primitive gate fidelity do not always agree. For example when four algorithms were run on two different five-qubit processors there was no definitive agreement from primitive to algorithmic fidelity [5]. In a five-qubit device with measured two-qubit gate fidelities of 0.99, the state fidelity of a five-qubit GHZ state was 0.82 after applying four two-qubit gates [6]. Therefore, to predict the true algorithmic fidelity we need to measure multiqubit fidelity metrics.

Fortunately, the issue of scaling can be circumvented if the goal is to characterize a process based on a few measures, e.g., average gate fidelity. Based on this idea, there have been several proposed techniques such as

Monte Carlo sampling [7, 8], compressed sensing [9], matrix product state tomography [10], and twirling protocols [11] which have been applied in a variety of multiqubit systems such as photons [12], NMR [13], and trapped ions [14]. Furthermore, the fidelity of certain multiqubit entangled states can be efficiently measured, as was demonstrated for 10 [15] and 12 [16] qubit states [16]. However, a common drawback to these techniques is that the result is sensitive to preparation and measurement errors (sometimes exponentially so) and/or does not fully characterize the underlying gates. These problems are addressed by randomized benchmarking [17, 18] (RB), where sequences of random Clifford gates equaling the identity operator are applied to a set of qubits. The decay of qubit polarization versus sequence length measures the average fidelity of the Clifford set independent of preparation and measurement errors. RB is a method widely used to characterize gates in superconducting circuits [6, 19–21], ion-traps [17, 22–24], neutral-atom-traps [25], NMR systems [26] and for solid-state spin qubits [27]. Extensions to RB have been proposed and implemented to measure specific gate error via interleaving [28], purity [29, 30] and leakage [31, 32].

RB is designed to address fidelities in multiqubit systems in two ways. For one, RB can be performed by constructing sequences from the n -qubit Clifford group. Additionally, the n -qubit space can be subdivided into sets of qubits $\{n_i\}$ and n_i -qubit RB performed in each subset simultaneously [33]. Both methods give metrics of fidelity in the n -qubit space. Despite the availability of these two methods, there has been no demonstration of RB with $n > 2$ since it is viewed as sufficient to characterize only the primitive gateset. Here we show, for the first time, a variety of three-qubit RB combinations in a three-qubit fixed-frequency superconducting device. For all-to-all gate connectivity we measure a three-qubit error per Clifford (3Q EPC) of 0.106, which is well-predicted by the primitive gate errors from simultaneous RB. However, we find a strong dependence on whether we perform

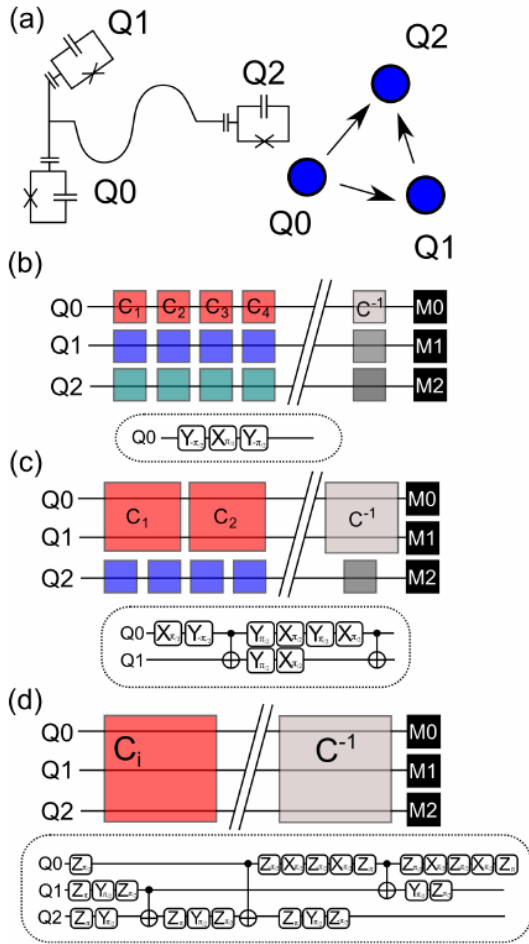


FIG. 1. (Color Online) (a) Schematic of the experimental setup and connectivity of the CNOT 2Q gates (control→target). (b) 1Q simultaneous RB $\{[0], [1], [2]\}$, (c) 2Q-1Q simultaneous RB $\{[0, 1], [2]\}$ and (d) 3Q RB $\{[0, 1, 2]\}$. Under each is a sample (b) 1Q (c) 2Q and (d) 3Q Clifford gate.

しかし、ゲートの較正を集団で行うか個人で行うかに強く依存することがわかり、ゲートを個人で較正すると誤差は0.302に増加する。重要なことは、この誤差の増大は、同時RBから測定される原始ゲート誤差の相応の増大によって予測されないことである。ゲートの一括校正の重要性は、最近の12量子ビットのクラスタ状態に関する文献[16]でも強調されています。[16]。また、デバイスのゲート接続を線形に制限した場合、3Q EPCは0.207に増加することから、デバイスの接続性の重要性を示しています。実験の詳細を説明する前に、まずRB法の概要を説明します。RBの詳細な議論は文献[34]にあります。RBの詳細な議論は文献[34]を参照してください。このアイデアは、ランダムなn-qubitクリフォード・ゲート $C_{n,i}^{Q_{m-1}}$ のm長シーケンスを構築することです。 $C_{n,i}^{Q_{m-1}}$ で、配列 $C_{n,m-1}^{-1}$ の逆数が付加されている。このような逆行列はGotteman-Knill theorem [35]によって効率的に計算される。

状態 $|0\rangle^{\otimes n}$ からスタートし、クリフォードゲートの全シーケンスを適用して、各クビットの $|0\rangle$ の個数を計測する。この手順を異なる乱数列に対して1回繰り返すと、1が大きい極限では、エラーマップが脱分極エラーマップ $\Lambda[\rho] = \alpha\rho + (1-\alpha)\mathbb{I}/d$ ($p = 1-\alpha$ は脱分極確率)に渦を巻く。配列長に対する $|0\rangle$ の集団は指数関数的減衰 $A\alpha^m + B$ に適合し、クリフォードゲート上の平均誤差は

$$EPC = \frac{2^n - 1}{2^n} (1 - \alpha), \quad (1)$$

(多種多様なノイズモデルに対して[36–38])。状態準備や測定誤差は崩壊定数に影響を与えない。クリフォード群のゲート数は超指数的に増加し、24個の1qubitゲート、11520個の2qubitゲート、92897280個の3qubitゲートが存在します[39]。しかし、この方法では、この集合から公正なサンプリングが必要なだけです。各ゲートはプリミティブゲートのセットから構築され、必要な1Qおよび2Qゲートの正確な数は、使用する基底に依存します。本研究では、2Qゲートは、iが制御、jがターゲットである制御NOT (CNOT_{ij}) とする

。1Qゲートの集合 $\{I, X_{m2}, X_{-m2}, Y_{m2}, Y_{-m2}\}$ ($P_\theta = e^{-i\theta\sigma_2 P}$) を使用して、私たちは1Qと2Qクリフォードゲートを生成します。このゲートセットでは、1Q Cliffordあたり2.2083個の1Qプリミティブゲート、2Q Cliffordあたり1.5個のCNOTゲートと12.2167個の1Qゲートが存在します。3Qクリフォードの生成には、1Qゲート $\{X_{m2}, X_{-m2}, Y_{-m2}\}$ に任意のZ回転を加えたセットを使用しますが、これはソフトウェアで定義されています[19]；これはQiskitコンパイラ[40]が使用しているセットです。全対全接続の場合、3.5個のCNOTゲートと11.6個の1Qゲート(XとYのみカウント)がある。Qiskitコンパイラを使用して、CNOTゲートの1つを削除して接続性を変更した結果、3Q Cliffordあたり平均7.7 CNOTゲートと18.4 1Qゲートになりました。1Q、2Q、3Qクリフォードのサンプルを図1に示す。多ビットシステムの場合、RBは(上記で詳述したように)完全なn-qubitsに対して行われることもあれば、システムのサブセットに対して行われることもある。例えば、CNOTゲートを定義する2量子ビットのサブセットに対して、他の量子ビットが静止している間に2Q RBを実行することが一般的です。文献[33]で説明されているように、このRBデータは、他の量子ビット部分空間が回転していないため、必ずしも指数関数的に減衰することはありません。サブセットは同時RBによってより厳密に特徴付けられ、すべての量子ビットがアクティブであるため、クロストーク誤差のあるレベルも測定されます。ここでは、 n_m 個の量子ビットのmセットが独立した n_m -量子ビットRBを実行しているベンチマークを示すために $\{[i, j], \dots, [k]\}$ という表記を使用します。例えば $\{[0], [1, 2]\}$ は、量子ビット0に1Q RB、量子ビット1、2に2Q RBを行うことを示す。3量子ビットの異なる組み合わせを図1に示します。

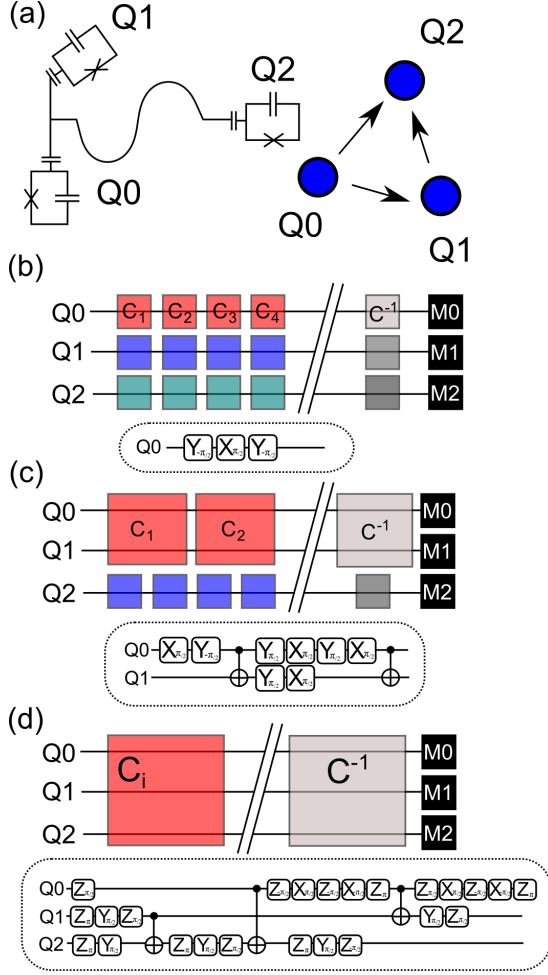


FIG. 1. (Color Online) (a) Schematic of the experimental setup and connectivity of the CNOT 2Q gates (control \rightarrow target). (b) 1Q simultaneous RB $\{[0], [1], [2]\}$, (c) 2Q-1Q simultaneous RB $\{[0, 1], [2]\}$ and (d) 3Q RB $\{[0, 1, 2]\}$. Under each is a sample 1Q (c) 2Q and (d) 3Q Clifford gate.

gate calibrations collectively or individually; the error increases to 0.302 when gates are calibrated individually. Importantly, this increase in error is not predicted by a commensurate increase in the primitive gate errors as measured from simultaneous RB. The importance of collective gate calibrations was also highlighted by the recent 12-qubit cluster state work of Ref. [16]. We also show the importance of connectivity in devices as the 3Q EPC increases to 0.207 when we limit the device to have linear gate connectivity.

Before describing our experiment in detail we first provide a brief summary of the RB method; a detailed discussion of RB can be found in Ref [34]. The idea is to construct an m -length sequence of random n -qubit Clifford gates $\prod_i^{m-1} \{C_{n,i}\} = \tilde{C}_{n,m-1}$ which is appended by the inverse of the sequence $\tilde{C}_{n,m-1}^{-1}$. Such an inverse is efficiently calculated by the Gottesman-Knill the-

orem [35]. Starting in the state $|0\rangle^{\otimes n}$ and applying the full sequence of Clifford gates, we then measure the population in $|0\rangle$ of each qubit. This procedure is repeated l times for different random sequences, which in the limit of large l , twirls the error map to a depolarizing error map $\Lambda[\rho] = \alpha\rho + (1 - \alpha)\mathcal{I}/d$ where $p = 1 - \alpha$ is the depolarizing probability. The population in $|0\rangle$ versus the sequence length fits to an exponential decay $A\alpha^m + B$ and the average error over the Clifford gates is

$$EPC = \frac{2^n - 1}{2^n} (1 - \alpha), \quad (1)$$

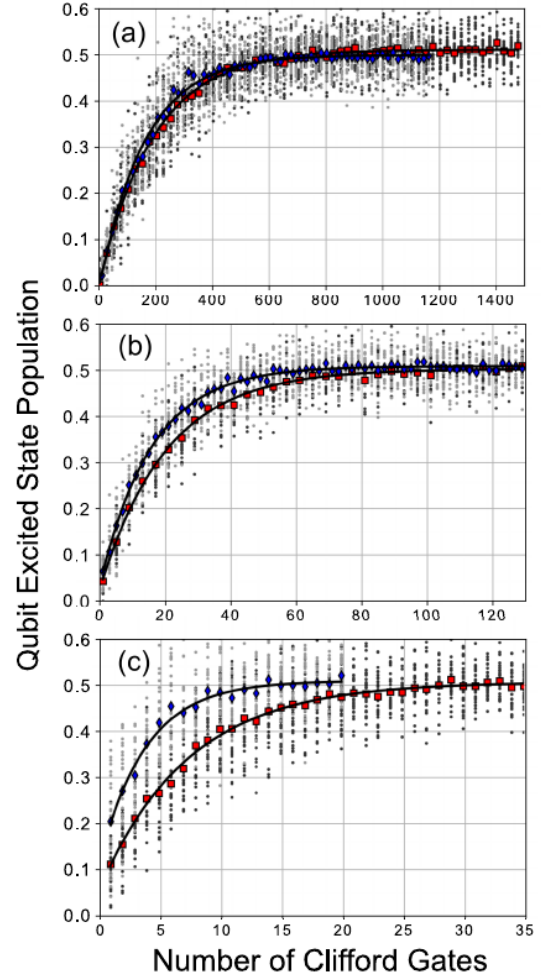
(for a wide variety of noise models [36–38]). State preparation and measurement errors do not affect the decay constant. The number of gates in the Clifford group grows superexponentially — there are 24 one-qubit gates, 11520 two-qubit gates and 92897280 three-qubit gates [39]. However, the method only requires fair sampling from this set. Each gate is constructed from a set of primitive gates and the exact number of 1Q and 2Q gates required depends on the basis used. In this work, our 2Q gate is a controlled NOT (CNOT $_{ij}$) where i is the control and j is the target. We generate our 1Q and 2Q Clifford gates using the set of 1Q gates $\{I, X_{\pi/2}, X_{-\pi/2}, Y_{\pi/2}, Y_{-\pi/2}\}$ where $P_\theta = e^{-i\theta/2P}$. With this gate set there are 2.2083 1Q primitive gates per 1Q Clifford and 1.5 CNOT gates and 12.2167 1Q gates per 2Q Clifford. To generate the 3Q Cliffords we use the set of 1Q gates $\{X_{\pi/2}, X_{-\pi/2}, Y_{-\pi/2}\}$ plus arbitrary Z rotations, which are software defined [19]; this is the set used by the Qiskit compiler [40]. For all-to-all connectivity there are 3.5 CNOT gates and 11.6 1Q gates (counting only X and Y). We use the Qiskit compiler to change the connectivity by removing one of the CNOT gates which results in an average of 7.7 CNOT gates and 18.4 1Q gates per 3Q Clifford. Sample 1Q, 2Q and 3Q Cliffords are shown in Fig. 1.

In the case of multiqubit systems, RB may be performed on the full n -qubits (as detailed above), or on subsets of the system. For example, it is common to perform 2Q RB on the subset of two-qubits defining a CNOT gate while the other qubits are quiescent. As explained in Ref [33], this RB data will not necessarily decay exponentially because the other qubit subspaces are not twirled. Subsets are more rigorously characterized by simultaneous RB, which also measures some level of crosstalk error since all qubits are active. Herein we will use the notation $\{[i, j], \dots, [k]\}$ to denote benchmarking where the m^{th} set of n_m qubits is performing independent n_m -qubit RB. For example $\{[0], [1, 2]\}$ would indicate 1Q RB on qubit 0 and 2Q RB on qubits 1 and 2. The different combinations for three-qubits are shown in Fig. 1.

To test 3Q RB we use a device comprised of three fixed-frequency superconducting transmon qubits (Q0, Q1, Q2) of frequencies (5.353, 5.291, 5.237) GHz coupled to a common 6.17GHz bus resonator. Our 1Q gates are 44.8 ns

3Q RBをテストするために、3つの固定周波数超伝導トランスモン量子ビット (Q0、Q1、Q2) の周波数 (5.353、5.291、5.237) GHzを、共通の6.17GHzバス共振器に結合した装置を使用します。1Qゲートは44.8ns幅のDRAG型マイクロ波パルス[41]である。2Qゲートは、ガウス平滑化された正方形のマイクロ波パルスを、他の量子ビット (ターゲット) の1つの周波数で量子ビット (コントロール) に適用します。これにより、交差共鳴相互作用が活性化され、240 nsの複合パルスCNOTゲートを構築するために調整することができます。詳細は文献[42]に記載されています。図1にデバイスとCNOTの接続の概略を示します。デバイスの詳細については、文献[43]に記載されています。[43]。我々の3量子ビットシステムでは、8つの可能なRBの組み合わせを考えています：同時1Q RB ($\{[0], [1], [2]\}$)、分離2Q RB ($\{[0], [1]\}, \{[0], [2]\}, \{[1], [2]\}$)、同時2Q RBと1Q RB (2Q-1Q RB) ($\{[0], [1], [2]\}, \{[0], [2], [1]\}, \{[1], [2], [0]\}$) そして最後に3Q RB ($\{[0], [1], [2]\}$)。それぞれの組み合わせについて、 $l = 30$ の平均を行った (ただし、2Q RBについては $l = 20$ とした)。同時RBでは、異なるサブシステムの配列長を一致させることを試み、2Q-1Q同時RBでは9:1Q:2Qのクリフォードゲートの比率を使用する。これらのRBシーケンスを、2つの異なる校正手順で実行します。手順Aでは、1Qゲートのパラメータ (量子ビット周波数、パルス振幅、ドラッグ振幅など) を同時に校正する。手順Bでは、1Qゲートのパラメータを個別に校正します。どちらの場合も、2Qゲートを個別に校正します。1Q、2Q、3QのRBから生成される曲線の種類を知るために、校正Aからのデータの一部を図2に示します。RBセット全体と両校正の誤差は、表Iにまとめられています。表Iから、2Q-1Q RBからの2Qゲート・エラーは、クロストークの増加と一致し、悪化していることがわかります。ただし、校正Aの CNOT_{12} は 2.8×10^{-2}

から 1.74×10^{-2} に減少しています。このことは、校正手順の違い、主に量子ビット周波数の校正値が異なることを強調しています。校正Aの量子ビット周波数は、ペア間の平均的なZZ相互作用 ($ZZ_{01} = 20$ kHz、 $ZZ_{02} = 352$ kHz、 $ZZ_{12} = 114$ kHz) によりシフトしていることがわかります。校正Aでは ZZ_{02} シフトをQ2の周波数に校正しているため、Q0が基底状態であれば CNOT_{12} のベンチマークではZ誤差が生じますが、校正Bではその逆なので単体の CNOT_{12} RB誤差は非常に小さい (0.92×10^2)。校正手順には微妙な違いがありますが、3Q RBの誤差には大きな違いがあり、3Q RBがいかにアルゴリズムの忠実度を校正手順から敏感に探ることができるかを示しています。全体として、平均的なZZを量子ビット周波数に校正することで、3Qの忠実度が最大になります。また、表Iのデータは、CNOTの1つを省略するとアルゴリズム誤差が著しく増大することから、接続性の重要性を示しています。つまり、1Qと2Qの誤差 (より具体的には、1Qと2Qの脱分極率) から3Qの誤差を予測することができるのか？この問いに答えるために、図2。(カラーオンライン) キャリブレーションAの異なるRBシーケンスによるQubit 0の実験データ。



黒線はデータへの指数関数フィットで、灰色の点は個々の試行によるものである。赤い四角 (青い菱形) は、明るい灰色 (暗い灰色) のポイントについて、これらの試行の平均値である。(a) 同時1Q (赤四角) および2Q-1Q RB (青菱形) からの1Q RB。(b) 01の組を単独で行った場合の2Q RB (赤の四角) とQ2の1Q RBと同時の2Q RB (青の菱形)。(c) 全対全接続の場合の3Q RB (赤の四角) と、限定的 (CNOT_{12} なし) 接続の場合の3Q RB (青の四角)。これらのフィットによる減衰パラメータはRef. [43]。

を計算し、予測される3Q崩壊パラメータ α (式 (1) を用いてEPCに変換する) を算出する、

$$\alpha_{3Q} = \frac{\alpha_{1Q}^{N_1/3} \alpha_{2Q}^{2N_2/3}}{7} \left(1 + 3\alpha_{1Q}^{N_1/3} \alpha_{2Q}^{N_2/3} + 3\alpha_{1Q}^{2N_1/3} \alpha_{2Q}^{N_2/3} \right) \quad (2)$$

ここで、 N_2 (N_1) は3Qクリフォードあたりの2Q (1Q) ゲートの数、 $p_1 = 1 - \alpha_1$ ($p_2 = 1 - \alpha_2$) は1Q (2Q) 分極除去確率である。

wide DRAG shaped microwave pulses [41]. Our 2Q gates are Gaussian smoothed square microwave pulses applied to a qubit (the control) at the frequency of one of the other qubits (the target). This activates a cross-resonance interaction, which can be tuned to build a composite pulse CNOT gate of 240 ns; details are found in Ref [42]. A schematic of the device and CNOT connectivity is shown in Fig. 1. More device details are given in Ref. [43].

For our three-qubit system we consider 8 possible RB combinations: simultaneous 1Q RB ($\{[0], [1], [2]\}$), separate 2Q RB ($\{[0, 1]\}, \{[0, 2]\}, \{[1, 2]\}$), simultaneous 2Q RB and 1Q RB (2Q-1Q RB) ($\{[0, 1], [2]\}, \{[0, 2], [1]\}, \{[1, 2], [0]\}$) and, finally, 3Q RB ($\{[0, 1, 2]\}$). For each combination we perform $l = 30$ averages (except for separate 2Q RB where $l = 20$). For simultaneous RB we attempt to match the sequence lengths on the different subsystems, so we use a ratio of 9:1 1Q:2Q Clifford gates for 2Q-1Q simultaneous RB. We perform these RB sequences under two different calibration procedures. In procedure A we calibrate the 1Q gate parameters simultaneously, e.g., qubit frequency, pulse amplitude and drag amplitude. In procedure B we calibrate the 1Q gate parameters individually. In both cases we calibrate the 2Q gates separately. To give a sense of the types of curves produced from 1Q, 2Q and 3Q RB, a subset of the data from calibration A is shown in Fig. 2. The errors from the full RB set and for both calibrations are summarized in Table I.

The data from Table I demonstrate that 2Q gate errors from 2Q-1Q RB are worse, consistent with increased crosstalk. There is one exception, CNOT_{12} , for calibration A which decreases from 2.8×10^{-2} to 1.74×10^{-2} . This highlights the difference between the calibration procedures, mainly that they result in different calibrated values for the qubit frequency. The qubit frequencies in calibration A are shifted by the average ZZ interaction between pairs ($\text{ZZ}_{01}=20$ kHz, $\text{ZZ}_{02}=352$ kHz and $\text{ZZ}_{12}=114$ kHz). Since the ZZ_{02} shift is calibrated into the frequency of Q2 for calibration A, there is a Z error when benchmarking CNOT_{12} if Q0 is in the ground state; the opposite is true for calibration B and so the standalone CNOT_{12} RB error is very low (0.92×10^{-2}). Although there is only a subtle difference between the calibration procedures, there is a large difference between the 3Q RB errors illustrating how 3Q RB can be a sensitive probe of such calibration procedures on algorithmic fidelity. Overall, calibrating the average ZZ into the qubit frequencies maximizes 3Q fidelity. The Table I data also show the importance of connectivity as omitting one of the CNOTs causes the algorithmic error to increase appreciably.

One of the main questions about 3Q RB is how much new information does it convey, i.e., can 3Q errors be predicted from the 1Q and 2Q errors (more specifically, the 1Q and 2Q depolarizing rates)? To answer this ques-

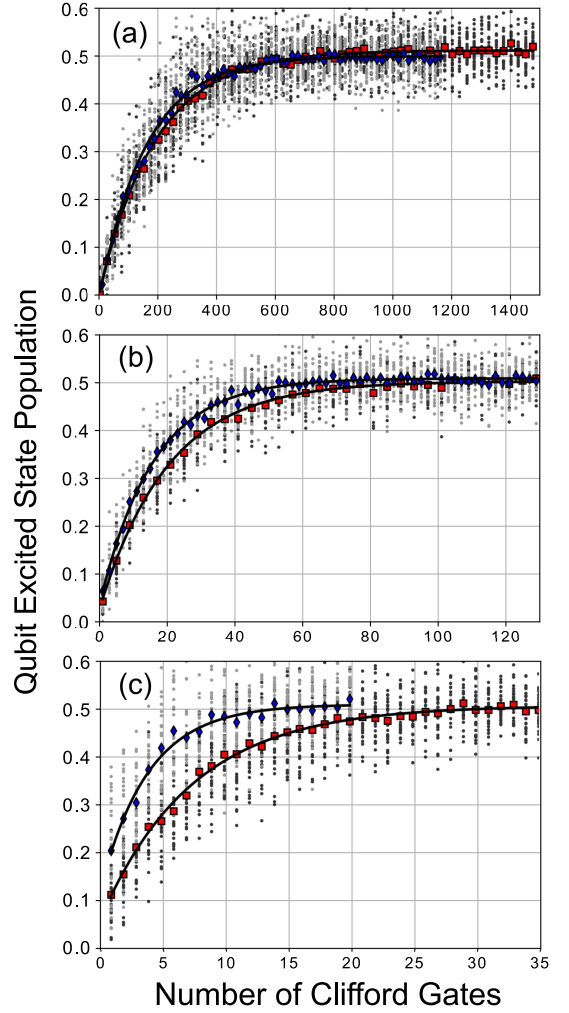


FIG. 2. (Color Online) Qubit 0 experimental data from different RB sequences for calibration A. Black lines are exponential fits to the data and the gray points are from the individual trials. Red squares (blue diamonds) are the averages over these trials for the light gray (dark gray) points. (a) 1Q RB from simultaneous 1Q (red squares) and 2Q-1Q RB (blue diamonds). (b) 2Q RB for the 01 pair performed in isolation (red square) and simultaneously with 1Q RB on Q2 (blue diamonds). (c) 3Q RB for all-to-all connectivity (red squares) and for limited (no CNOT_{12}) connectivity (blue squares). The decay parameters from these fits are given in Ref. [43].

tion we calculate the predicted 3Q decay parameter α (converting to EPC using Eqn. 1),

$$\alpha_{3Q} = \frac{\alpha_{1Q}^{N_1/3} \alpha_{2Q}^{2N_2/3}}{7} \left(1 + 3\alpha_{1Q}^{N_1/3} \alpha_{2Q}^{N_2/3} + 3\alpha_{1Q}^{2N_1/3} \alpha_{2Q}^{N_2/3} \right) \quad (2)$$

where N_2 (N_1) is the number of 2Q (1Q) gates per 3Q Clifford, and $p_1 = 1 - \alpha_1$ ($p_2 = 1 - \alpha_2$) are the 1Q (2Q) depolarizing probabilities. For simplicity we assume that all 1Q gates and 2Q gates have the same depolarizing

	Cal. A	Cal. B
T_1	$[29,50,39] \mu s$	$[42,47,35] \mu s$
T_2	$[39,75,59] \mu s$	$[61,74,46] \mu s$
1Q EPG Coherence Limit	$[6.5,3.5,4.4] \times 10^{-4}$	$[4.2,3.6,5.4] \times 10^{-4}$
1Q EPG from $\{[0], [1], [2]\}$ RB	$[1.12(2), 0.86(1), 1.22(2)] \times 10^{-3}$	$[1.40(5), 0.81(1), 1.66(4)] \times 10^{-3}$
1Q EPG from $\{[i], [j, k]\}$ RB	$[1.41(3), 0.95(2), 1.35(2)] \times 10^{-3}$	$[1.68(4), 0.95(2), 1.54(3)] \times 10^{-3}$
2Q EPG Coherence Limit	$[6,7,5] \times 10^{-3}$	$[5,6,6] \times 10^{-3}$
2Q EPG from $\{[i, j]\}$ RB	$[1.26(7), 1.15(8), 2.8(2)] \times 10^{-2}$	$[0.86(5), 2.8(1), 0.92(7)] \times 10^{-2}$
2Q EPG from $\{[i, j], [k]\}$ RB	$[1.89(6), 1.62(6), 1.74(7)] \times 10^{-2}$	$[2.45(8), 4.2(2), 4.3(2)] \times 10^{-2}$
3Q EPC from $\{[0, 1, 2]\}$ RB (all-to-all)	0.106(2)	0.302(6)
3Q EPC from $\{[0, 1, 2]\}$ RB (omit CNOT ₁₂)	0.207(3)	N/A

表1. 1Q (1qubit) EPGでは $[Q_0, Q_1, Q_2]$ 順、2Q (2qubit) EPGでは $[CNOT_{01}, CNOT_{02}, CNOT_{12}]$ 順で異なるRB実験によるEPG (ゲートあたりのエラー) とEPC (Cliffordあたりのエラー)。1Q EPGは本文で示したセットで平均化したゲートあたりの誤差です。2Q EPGは $\{[0], [1], [2]\}$ ベンチマークによる1Q EPGを仮定した2Q EPCから計算される (この計算の詳細については[43]を参照)。校正BのCNOT₁₂を省略した3Q EPCは、誤差が大きすぎてデータに適切にフィットしないためN/A。コヒーレンス制限誤差は、 T_1 と T_2 の誤差のみを仮定して計算されています。 T_1 と T_2 の校正間の変動は、実験と実験の間の約3日間のドリフトによるものである。誤差はフィットパラメーターの不確実性を反映している。

	Cal. A		Cal. B
	All-to-All	Omit CNOT ₁₂	All-to-All
3Q EPC from RB	0.106(2)	0.207(3)	0.302(6)
Coherence Limit	0.044	0.094	0.041
3Q EPC Predicted from $\{[i], [j, k]\}$ RB	0.115(4)	0.226(6)	0.187(7)

簡単のため、すべての1Qゲートと2Qゲートが同じ脱分極確率であると仮定する 表II. 式2を適用して表Iに示した1Qと2QのEPG番号から予測される3Q EPC。計算の詳細については本文を参照のこと。

確率; 式2の一般形と導出の詳細については、[43]を参照。 N_1 と N_2 について先に議論した値は、ゲートの有限時間を考慮していない。現実には、いくつかの量子ビットにアイドル期間が存在し、アイドル期間を1量子ビットゲートとして特徴付けると、全対全 (限定) 接続では $N_1 = 34.7$ ($N_1 = 67.9$) になる。これは3Q EPCの予測に使われる数値である。式(2)の1Qと2Qの脱分極確率については、表Iから2組の数値を使用して、表IIに示す3Q EPCの予測値を算出する。最初のセットはコヒーレンス制限EPGで、意外にも3Q EPCの予測値は測定値よりもはるかに低く、誤差の大部分はクロストークなどのハミルトニアン内の不要で補償されない項によるものであることを示しています。2つ目の数値は、2Q-1Q同時RBによるもので、プリミティブゲートエラーを最も正確に測定できるはずです。実際、キャリブレーションAの場合、予測される3Q EPCは、all-to-allとlimited connectivityの両方において正確である。しかし、校正Bの場合、予測された3Q EPCと測定された3Q EPCの間にはほとんど一致がなく、3Q RB忠実度は、原始ゲートのベンチマークでは十分に明らかにならない微妙なエラーに敏感な新しいマルチビット指標として有用であることが示されました。

キャリブレーションBでは、他の量子ビットが2Qゲートを実行する間、観客量子ビットにアイドル時間があるため、補償されないZZエラーが3Qクリフォードゲートの特定の構造によって増幅されます (これは図1 d. に模式的に示されている)。測定された1Q/2Q誤差とZZを含むシミュレーションは、観測された3Q RBデータをよく予測します (文献[43]を参照)。[43]. 3Qクリフォードゲートの実装は一意ではないので、ある構造では異なる誤差項を増幅または減衰させる可能性があるため、そのような構造を詳細に調べることは今後の研究に委ねられる。

結論として、我々は初めて3Q RBとサブセット2Q-1Q同時RBを実証した。真の原始的な3量子ビットゲートは存在しませんが、3Q RBは、1量子ビットゲートや2量子ビットゲートの測定基準では捉えられない忠実度を測定します。システムが大型化し、クロストーク項が誤差を支配するようになると、3Q RBのようなメトリクスは、こうした大型システムの真のアルゴリズム的忠実度をベンチマークする上で重要な役割を果たすようになります。3Q RBはクロストーク・エラーの修正方法を示すものではありませんが、緩和策を検証する上で重要な役割を果たすことになるでしょう。クロストークを抑制するソフトウェアおよびハードウェアの方法は、活発な研究分野であり、チューナブルカブラのようなアクティブ素子の使用が必要になるかもしれません[44, 45]。

Firat Solgun, Markus Brink, Sami Rosenblatt, George Keefeには、デバイスのモデリングと作製に協力してもらいました。Lev Bishop, Andrew Cross, Easwar Magesan, Antonio Corcolesには、議論と原稿のコメントをいただきました。クリフォード・ゲートの生成に協力してくれたChristopher WoodとSergey Bravyiに感謝します。この研究は、契約W911NF-14-1-0124のもと、陸軍研究所の支援を受けたものである。

	Cal. A	Cal. B
T_1	[29,50,39] μs	[42,47,35] μs
T_2	[39,75,59] μs	[61,74,46] μs
1Q EPG Coherence Limit	$[6.5, 3.5, 4.4] \times 10^{-4}$	$[4.2, 3.6, 5.4] \times 10^{-4}$
1Q EPG from $\{[0], [1], [2]\}$ RB	$[1.12(2), 0.86(1), 1.22(2)] \times 10^{-3}$	$[1.40(5), 0.81(1), 1.66(4)] \times 10^{-3}$
1Q EPG from $\{[i], [j, k]\}$ RB	$[1.41(3), 0.95(2), 1.35(2)] \times 10^{-3}$	$[1.68(4), 0.95(2), 1.54(3)] \times 10^{-3}$
2Q EPG Coherence Limit	$[6, 7, 5] \times 10^{-3}$	$[5, 6, 6] \times 10^{-3}$
2Q EPG from $\{[i, j]\}$ RB	$[1.26(7), 1.15(8), 2.8(2)] \times 10^{-2}$	$[0.86(5), 2.8(1), 0.92(7)] \times 10^{-2}$
2Q EPG from $\{[i, j], [k]\}$ RB	$[1.89(6), 1.62(6), 1.74(7)] \times 10^{-2}$	$[2.45(8), 4.2(2), 4.3(2)] \times 10^{-2}$
3Q EPC from $\{[0, 1, 2]\}$ RB (all-to-all)	0.106(2)	0.302(6)
3Q EPC from $\{[0, 1, 2]\}$ RB (omit CNOT ₁₂)	0.207(3)	N/A

TABLE I. EPG (error per gate) and EPC (error per Clifford) from different RB experiments in [Q0,Q1,Q2] order for 1Q (one-qubit) EPG and in order [CNOT₀₁, CNOT₀₂, CNOT₁₂] for the 2Q (two-qubit) EPG. 1Q EPG is the error per gate averaged over the set indicated in the main text. 2Q EPG is calculated from the 2Q EPC assuming the 1Q EPG from $\{[0], [1], [2]\}$ benchmarking (see [43] for details of this calculation). 3Q EPC omitting CNOT₁₂ for calibration B is N/A because the error was too high to properly fit the data. The coherence limited errors are calculated assuming only errors from T_1 and T_2 . Variability in T_1 and T_2 between the calibrations is due to drift over the approximately three days between experiments. Errors reflect the uncertainty in the fit parameters.

	Cal. A		Cal. B
	All-to-All	Omit CNOT ₁₂	All-to-All
3Q EPC from RB	0.106(2)	0.207(3)	0.302(6)
Coherence Limit	0.044	0.094	0.041
3Q EPC Predicted from $\{[i], [j, k]\}$ RB	0.115(4)	0.226(6)	0.187(7)

TABLE II. Predicted 3Q EPC from 1Q and 2Q EPG numbers listed in Table I by applying Eqn. 2. See the main text for a detailed discussion of the calculation.

probability; see [43] for the general form of Eqn. 2 and details of the derivation. The values discussed previously for N_1 and N_2 did not consider the finite duration of gates. In reality, there will be idle periods on some qubits and characterizing idle periods as one-qubit gates, $N_1 = 34.7$ ($N_1 = 67.9$) for all-to-all (limited) connectivity. This is the number used for predicting the 3Q EPC.

For the 1Q and 2Q depolarizing probabilities in Eqn (2) we use two sets of numbers from Table I to calculate the predicted 3Q EPC shown in Table II. The first set are the coherence limited EPGs which unsurprisingly predict a much lower than measured 3Q EPC indicating that the majority of errors are due to unwanted and uncompensated terms in the Hamiltonian such as crosstalk. The second set of numbers are from 2Q-1Q simultaneous RB, which should be the most accurate measure of primitive gate errors. Indeed, for calibration A the predicted 3Q EPC is accurate for both all-to-all and limited connectivity. However, in the case of calibration B, there is very little agreement between the predicted and measured 3Q EPC, demonstrating the utility of the 3Q RB fidelity as a novel multiqubit metric sensitive to subtle errors that are

not fully revealed by benchmarking the primitive gates. In calibration B the uncompensated ZZ errors are amplified by the specific structure of the 3Q Clifford gate since there are idle periods on the spectator qubits while the other qubits perform the 2Q gate (this is schematically illustrated in Fig. 1 d.). Simulations including the measured 1Q/2Q errors and ZZ predict well the observed 3Q RB data, see Ref. [43]. Since the implementation of the 3Q Clifford gate is not unique, certain constructions may amplify or attenuate different error terms; investigating such constructions in detail is left for future study.

In conclusion, we demonstrate, for the first time, 3Q RB and subset 2Q-1Q simultaneous RB. Although there is no true primitive three-qubit gate, 3Q RB measures a fidelity that is not captured by the one- and two-qubit gate metrics. As systems continue to increase in size and crosstalk terms dominate error, metrics such as 3Q RB will play an important role in benchmarking the true algorithmic fidelity of these large systems. Although 3Q RB does not indicate how to correct crosstalk errors, it will play an important role in validating mitigation strategies. Software and hardware methods to suppress crosstalk are an active area of research and may require the use of active elements such as tunable couplers [44, 45].

We thank Firat Solgun, Markus Brink, Sami Rosenblatt and George Keefe for modeling and fabricating the device. We thank Lev Bishop, Andrew Cross, Easwar Magesan and Antonio Corcoles for discussions and manuscript comments. We thank Christopher Wood and Sergey Bravyi for help generating the Clifford gates. This work was supported by the Army Research Office under contract W911NF-14-1-0124.

- {dcmckay@us.ibm.com} [1] Isaac L. Chuang and M.A. Nielsen, "Prescription for experimental determination of a dynamics of a quantum black box," *J. Mod. Opt.* 44, 2455 (1997). [2] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information* (Cambridge, 2000). [3] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland, "Surface codes: 実用的な大規模量子計算に向けて" *Phys. Rev. A* 86, 032324 (2012). [4] 滝田舞香、アンドリュー・W・クロス、A・D・コルコレス、ジェリー・M・チョウ、ジェイ・M・ガンベッタ、「超伝導量子ビットによる耐障害性状態準備の実験的実証」、*Phys. Rev. Lett.* 119, 180501 (2017). [5] Norbert M. Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A. Landsman, Kenneth Wright, and Christopher Monroe, "Experimental comparison of two quantum computing architectures," *Proc. Nat. Acad. Sci.* 114, 3305 (2017). [6] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and John M. Martinis, "Superconducting quantum circuits at the surface code threshold for fault tolerance," *Nature* 508, 500 (2014). [7] Steven T. Flammia and Yi-Kai Liu, "Direct fidelity estimation from few pauli measurements," *Phys. Rev. Lett.* 106, 230501 (2011). [8] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin, "Practical characterization of quantum devices without tomography," *Phys. Rev. Lett.* 107, 210404 (2011). [9] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert, "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.* 105, 150401 (2010). [10] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu, "Efficient quantum state tomography," *Nat. Comm.* 1, 149 (2010). [11] Osama Moussa, Marcus P. da Silva, Colm A. Ryan, and Raymond Laflamme, "Practical experimental certification of computational quantum gates using a twirling procedure," *Phys. Rev. Lett.* 109, 070504 (2012). [12] Christian Schwemmer, Géza Tóth, Alexander Niggelbaum, Tobias Moroder, David Gross, Otfried Gühne, and Harald Weinfurter, "Experimental comparison of efficient tomography schemes for a six-qubit state," *Phys. Rev. Lett.* 113, 040503 (2014). [13] Dawei Lu, Hang Li, Denis-Alexandre Trottier, Jun Li, Aharon Brodutch, Anthony P. Krismanich, Ahmad Ghavami, Gary I. Dmitrienko, Guilu Long, Jonathan Baugh, and Raymond Laflamme, "Experimental estimation of average fidelity of a clifford gate on a 7-qubit quantum processor," *Phys. Rev. Lett.* 114, 140505 (2015). [14] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, "Efficient tomography of a quantum many-body system," *Nature Physics* 13, 1158 (2017). Chao Song, Kai Xu, Wuxin Liu, Chui-ping Yang, ShiBiao Zheng, Hui Deng, Qiwei Xie, Keqiang Huang, Qiujiang Guo, Libo Zhang, Pengfei Zhang, Da Xu, Dongning Zheng, Xiaobo Zhu, H. Wang, Y.-A. Cheon, C.-Y. Lu, Siyuan Han, and Jian-Wei Pan, "10-qubit entanglement and parallel logic operations with a superconducting circuit," *Phys. Rev. Lett.* 119, 180511 (2017). [15] Ming Gong, Ming-Cheng Chen, Yarui Zheng, Shiyu Wang, Chen Zha, Hui Deng, Zhiguang Yan, Hao Rong, Yulin Wu, Shaowei Li, Fusheng Chen, Youwei Zhao, Futian Liang, Jin Lin, Yu Xu, Cheng Guo, Lihua Sun, Anthony D. Castellano, Haohua Wang, Chengzhi Peng, Chao-Yang Lu, Xiaobo Zhu, and Jian-Wei Pan, "Genuine 12-qubit entanglement on a superconducting quantum processor," *Phys. Rev. Lett.* 122, 110501 (2019). E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, "Randomized benchmarking of quantum gates," *Phys. Rev. A* 77, 012307 (2008). Easwar Magesan, J. M. Gambetta, and Joseph Emerson, "Scalable and robust randomized benchmarking of quantum processes," *Phys. Rev. Lett.* 106, 180504 (2011). David C. McKay, Christopher J. Wood, Sarah Sheldon, Jerry M. Chow, and Jay M. Gambetta, "Efficient z gates for quantum computing," *Phys. Rev. A* 96, 022330 (2017). J. M. Chow, J. M. Gambetta, L. Tornberg, Jens Koch, Lev S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Randomized benchmarking and process tomography for gate errors in a solid-state qubit," *Phys. Rev. Lett.* 102, 090502 (2009). A. D. Córcoles, Jay M. Gambetta, Jerry M. Chow, John A. Smolin, Matthew Ware, Joel Strand, B. L. T. Plourde, and M. Steffen, "Process verification of two-qubit quantum gates by randomized benchmarking," *Phys. Rev. A* 87, 030301 (2013). J. P. Gaebler, T. R. Tan, Y. Lin, Y. Wan, R. Bowler, A. C. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. J. Wineland, "High-fidelity universal gate set for $^9\text{Be}^+$ ion qubit," *Phys. Rev. Lett.* 117, 060505 (2016). C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, "High-fidelity quantum logic gates using trapped-ion hyperfine qubits," *Phys. Rev. Lett.* 117, 060504 (2016). J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, "Randomized benchmarking of multiqubit gates," *Phys. Rev. Lett.* 108, 260503 (2012). S. Olmschenk, R. Chircireanu, K. D. Nelson, and J. V. Porto, "Randomized benchmarking of atomic qubits in an optical lattice," *New J. Phys.* 12, 113007 (2010). C. A. Ryan, M. Laforest, and R. Laflamme, "Randomized benchmarking of single and multi-qubit control in liquid-state nmr quantum information processing," *New J. Phys.* 11, 013034 (2009). M. Veldhorts, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, "An Addressable quantum dot qubit with fault-tolerant control-fidelity," *Nat. Nano.* 9, 981 (2014).

* dcmckay@us.ibm.com

- [1] Isaac L. Chuang and M.A. Nielsen, “Prescription for experimental determination of the dynamics of a quantum black box,” *J. Mod. Opt.* **44**, 2455 (1997).
- [2] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information* (Cambridge, 2000).
- [3] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland, “Surface codes: Towards practical large-scale quantum computation,” *Phys. Rev. A* **86**, 032324 (2012).
- [4] Maika Takita, Andrew W. Cross, A. D. Córcoles, Jerry M. Chow, and Jay M. Gambetta, “Experimental demonstration of fault-tolerant state preparation with superconducting qubits,” *Phys. Rev. Lett.* **119**, 180501 (2017).
- [5] Norbert M. Linke, Dmitri Maslov, Martin Roetteler, Shantanu Debnath, Caroline Figgatt, Kevin A. Landsman, Kenneth Wright, and Christopher Monroe, “Experimental comparison of two quantum computing architectures,” *Proc. Nat. Acad. Sci.* **114**, 3305 (2017).
- [6] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O’Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and John M. Martinis, “Superconducting quantum circuits at the surface code threshold for fault tolerance,” *Nature* **508**, 500 (2014).
- [7] Steven T. Flammia and Yi-Kai Liu, “Direct fidelity estimation from few pauli measurements,” *Phys. Rev. Lett.* **106**, 230501 (2011).
- [8] Marcus P. da Silva, Olivier Landon-Cardinal, and David Poulin, “Practical characterization of quantum devices without tomography,” *Phys. Rev. Lett.* **107**, 210404 (2011).
- [9] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert, “Quantum state tomography via compressed sensing,” *Phys. Rev. Lett.* **105**, 150401 (2010).
- [10] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu, “Efficient quantum state tomography,” *Nat. Comm.* **1**, 149 (2010).
- [11] Osama Moussa, Marcus P. da Silva, Colm A. Ryan, and Raymond Laflamme, “Practical experimental certification of computational quantum gates using a twirling procedure,” *Phys. Rev. Lett.* **109**, 070504 (2012).
- [12] Christian Schwemmer, Géza Tóth, Alexander Niggelbaum, Tobias Moroder, David Gross, Otfried Gühne, and Harald Weinfurter, “Experimental comparison of efficient tomography schemes for a six-qubit state,” *Phys. Rev. Lett.* **113**, 040503 (2014).
- [13] Dawei Lu, Hang Li, Denis-Alexandre Trottier, Jun Li, Aharon Brodutch, Anthony P. Krismanich, Ahmad Ghavami, Gary I. Dmitrienko, Guilu Long, Jonathan Baugh, and Raymond Laflamme, “Experimental estimation of average fidelity of a clifford gate on a 7-qubit quantum processor,” *Phys. Rev. Lett.* **114**, 140505 (2015).
- [14] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, “Efficient tomography of a quantum many-body system,” *Nature Physics* **13**, 1158 (2017).
- [15] Chao Song, Kai Xu, Wuxin Liu, Chui-ping Yang, Shi-Biao Zheng, Hui Deng, Qiwei Xie, Keqiang Huang, Qiu-jiang Guo, Libo Zhang, Pengfei Zhang, Da Xu, Dongning Zheng, Xiaobo Zhu, H. Wang, Y.-A. Chen, C.-Y. Lu, Siyuan Han, and Jian-Wei Pan, “10-qubit entanglement and parallel logic operations with a superconducting circuit,” *Phys. Rev. Lett.* **119**, 180511 (2017).
- [16] Ming Gong, Ming-Cheng Chen, Yarui Zheng, Shiyu Wang, Chen Zha, Hui Deng, Zhiguang Yan, Hao Rong, Yulin Wu, Shaowei Li, Fusheng Chen, Youwei Zhao, Futing Liang, Jin Lin, Yu Xu, Cheng Guo, Lihua Sun, Anthony D. Castellano, Haohua Wang, Chengzhi Peng, Chao-Yang Lu, Xiaobo Zhu, and Jian-Wei Pan, “Generate 12-qubit entanglement on a superconducting quantum processor,” *Phys. Rev. Lett.* **122**, 110501 (2019).
- [17] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, “Randomized benchmarking of quantum gates,” *Phys. Rev. A* **77**, 012307 (2008).
- [18] Easwar Magesan, J. M. Gambetta, and Joseph Emerson, “Scalable and robust randomized benchmarking of quantum processes,” *Phys. Rev. Lett.* **106**, 180504 (2011).
- [19] David C. McKay, Christopher J. Wood, Sarah Sheldon, Jerry M. Chow, and Jay M. Gambetta, “Efficient z gates for quantum computing,” *Phys. Rev. A* **96**, 022330 (2017).
- [20] J. M. Chow, J. M. Gambetta, L. Tornberg, Jens Koch, Lev S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, “Randomized benchmarking and process tomography for gate errors in a solid-state qubit,” *Phys. Rev. Lett.* **102**, 090502 (2009).
- [21] A. D. Córcoles, Jay M. Gambetta, Jerry M. Chow, John A. Smolin, Matthew Ware, Joel Strand, B. L. T. Plourde, and M. Steffen, “Process verification of two-qubit quantum gates by randomized benchmarking,” *Phys. Rev. A* **87**, 030301 (2013).
- [22] J. P. Gaebler, T. R. Tan, Y. Lin, Y. Wan, R. Bowler, A. C. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. J. Wineland, “High-fidelity universal gate set for $^9\text{Be}^+$ ion qubits,” *Phys. Rev. Lett.* **117**, 060505 (2016).
- [23] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, “High-fidelity quantum logic gates using trapped-ion hyperfine qubits,” *Phys. Rev. Lett.* **117**, 060504 (2016).
- [24] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, “Randomized benchmarking of multi-qubit gates,” *Phys. Rev. Lett.* **108**, 260503 (2012).
- [25] S. Olmschenk, R. Chicireanu, K. D. Nelson, and J. V. Porto, “Randomized benchmarking of atomic qubits in an optical lattice,” *New. J. Phys.* **12**, 113007 (2010).
- [26] C. A. Ryan, M. Laforest, and R. Laflamme, “Randomized benchmarking of single- and multi-qubit control in liquid-state nmr quantum information processing,” *New. J. Phys.* **11**, 013034 (2009).
- [27] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, “An addressable quantum dot qubit with fault-tolerant control-fidelity,” *Nat. Nano.* **9**, 981 (2014).

- [28] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, Mark B. Ketchen, and M. Steffen, "Efficient measurement of quantum gate error by interleaved randomized benchmarking," *Phys. Rev. Lett.* **109**, 080505 (2012). [29] David C. McKay, Stefan Filipp, Antonio Mezzacapo, Easwar Magesan, Jerry M. Chow, and Jay M. Gambetta, "Universal gate for fixed-frequency qubits via a tunable bus," *Phys. Rev. Applied* **6**, 064007 (2016). [30] Joel Wallman, Chris Granade, Robin Harper, and Steven T Flammia, "Estimating the coherence of noise," *New. J. Phys.* **17**, 113020 (2015). [31] Christopher J. Wood and Jay M. Gambetta, "Quantification and characterization of leakage errors," (2017), [arxiv:1704.03081](#). [32] Joel J Wallman, Marie Barnhill, and Joseph Emerson, "Robust characterization of leakage errors," *New. J. Phys.* **18**, 043021 (2016). [33] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, John A. Smolin, Jerry M. Chow, Colm A. Ryan, Chad Rigetti, S. Poletto, Thomas A. Ohki, Mark B. Ketchen, and M. Steffen, "Characterization of addressability by simultaneous randomized benchmarking" *Phys. Rev. Lett.* **109**, 240504 (2012). [34] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson, "Characterizing quantum gates via randomized benchmarking," *Phys. Rev. A* **85**, 042311 (2012). [35] Daniel Gottesman, "The Heisenberg Representation of quantum computers," (1998), [arxiv:quant-ph/9807006](#). [36] Jeffrey M. Epstein, Andrew W. Cross, Easwar Magesan, and Jay M. Gambetta, "Investigating the limits of randomized benchmarking protocols," *Phys. Rev. A* **89**, 062321 (2014).
- [37] Joel J. Wallman, "Randomized benchmarking with gate-dependent noise," (2017), [arxiv:1703.09835](#).
- [38] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout, "What randomized benchmarking actually measures," *Phys. Rev. Lett.* **119**, 130502 (2017).
- [39] Maris Ozols, "Clifford group," (2008).
- [40] "Qiskit SDK," Online (2017).
- [41] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, "Simple pulses for elimination of leakage in weakly nonlinear qubits," *Phys. Rev. Lett.* **103**, 110501 (2009).
- [42] Sarah Sheldon, Easwar Magesan, Jerry M. Chow, and Jay M. Gambetta, "Procedure for systematically tuning up cross-talk in the cross-resonance gate," *Phys. Rev. A* **93**, 060302 (2016).
- [43] David McKay, "See supplementary material [url] which includes refs. [33, 46]," (2019).
- [44] Gengyan Zhang, Pranav S. Mundada, and Andrew A. Houck, "Suppression of qubit crosstalk in a tunable coupling superconducting circuit," (2018), [arxiv:1810.04182](#).
- [45] R. C. Bialczak, M. Ansmann, M. Hofheinz, M. Lenander, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis, "Fast tunable coupler for superconducting qubits," *Phys. Rev. Lett.* **106**, 060501 (2011).
- [46] Scott Aaronson and Daniel Gottesman, "Improved simulation of stabilizer circuits," *Phys. Rev. A* **70**, 052328 (2004).

- [28] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, Mark B. Ketchen, and M. Steffen, “Efficient measurement of quantum gate error by interleaved randomized benchmarking,” *Phys. Rev. Lett.* **109**, 080505 (2012).
- [29] David C. McKay, Stefan Filipp, Antonio Mezzacapo, Easwar Magesan, Jerry M. Chow, and Jay M. Gambetta, “Universal gate for fixed-frequency qubits via a tunable bus,” *Phys. Rev. Applied* **6**, 064007 (2016).
- [30] Joel Wallman, Chris Granade, Robin Harper, and Steven T Flammia, “Estimating the coherence of noise,” *New. J. Phys.* **17**, 113020 (2015).
- [31] Christopher J. Wood and Jay M. Gambetta, “Quantification and characterization of leakage errors,” (2017), [arxiv:1704.03081](#).
- [32] Joel J Wallman, Marie Barnhill, and Joseph Emerson, “Robust characterization of leakage errors,” *New. J. Phys.* **18**, 043021 (2016).
- [33] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, John A. Smolin, Jerry M. Chow, Colm A. Ryan, Chad Rigetti, S. Poletto, Thomas A. Ohki, Mark B. Ketchen, and M. Steffen, “Characterization of addressability by simultaneous randomized benchmarking,” *Phys. Rev. Lett.* **109**, 240504 (2012).
- [34] Easwar Magesan, Jay M. Gambetta, and Joseph Emerson, “Characterizing quantum gates via randomized benchmarking,” *Phys. Rev. A* **85**, 042311 (2012).
- [35] Daniel Gottesman, “The heisenberg representation of quantum computers,” (1998), [arxiv:quant-ph/9807006](#).
- [36] Jeffrey M. Epstein, Andrew W. Cross, Easwar Magesan, and Jay M. Gambetta, “Investigating the limits of randomized benchmarking protocols,” *Phys. Rev. A* **89**, 062321 (2014).
- [37] Joel J. Wallman, “Randomized benchmarking with gate-dependent noise,” (2017), [arxiv:1703.09835](#).
- [38] Timothy Proctor, Kenneth Rudinger, Kevin Young, Mohan Sarovar, and Robin Blume-Kohout, “What randomized benchmarking actually measures,” *Phys. Rev. Lett.* **119**, 130502 (2017).
- [39] Maris Ozols, “Clifford group,” (2008).
- [40] “Qiskit SDK,” Online (2017).
- [41] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, “Simple pulses for elimination of leakage in weakly nonlinear qubits,” *Phys. Rev. Lett.* **103**, 110501 (2009).
- [42] Sarah Sheldon, Easwar Magesan, Jerry M. Chow, and Jay M. Gambetta, “Procedure for systematically tuning up cross-talk in the cross-resonance gate,” *Phys. Rev. A* **93**, 060302 (2016).
- [43] David McKay, “See supplementary material [url] which includes refs. [33, 46],” (2019).
- [44] Gengyan Zhang, Pranav S. Mundada, and Andrew A. Houck, “Suppression of qubit crosstalk in a tunable coupling superconducting circuit,” (2018), [arxiv:1810.04182](#).
- [45] R. C. Bialczak, M. Ansmann, M. Hofheinz, M. Lenander, E. Lucero, M. Neeley, A. D. O’Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis, “Fast tunable coupler for superconducting qubits,” *Phys. Rev. Lett.* **106**, 060501 (2011).
- [46] Scott Aaronson and Daniel Gottesman, “Improved simulation of stabilizer circuits,” *Phys. Rev. A* **70**, 052328 (2004).