PAPER • OPEN ACCESS

A randomized benchmarking suite for mid-circuit measurements

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

A randomized benchmarking suite for mid-circuit measurements

L C G Govia[1,*] , P Jurcevic[2] , C J Wood[2] , N Kanazawa[3], S T Merkel[2] and D C McKay[2]

1 IBM Quantum, IBM Almaden Research Center, San Jose, CA 95120, United States of America
2 IBM Quantum, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, United States of America
3 IBM Quantum, IBM Research Tokyo, 19-21 Nihonbashi Hakozaki-cho, Chuo-ku, Tokyo 103-8510, Japan
* Author to whom any correspondence should be addressed.

E-mail: lcggovia@ibm.com

Abstract

Mid-circuit measurements are a key component in many quantum information computing protocols, including quantum error correction, fault-tolerant logical operations, and measurement based quantum computing. As such, techniques to quickly and efficiently characterize or benchmark their performance are of great interest. Beyond the measured qubit, it is also relevant to determine what, if any, impact mid-circuit measurement has on adjacent, unmeasured, spectator qubits. Here, we present a mid-circuit measurement benchmarking suite developed from the ubiquitous paradigm of randomized benchmarking. We show how our benchmarking suite can be used to both detect as well as quantify errors on both measured and spectator qubits, including measurement-induced errors on spectator qubits and entangling errors between measured and spectator qubits. We demonstrate the scalability of our suite by simultaneously characterizing mid-circuit measurement on multiple qubits from an IBM Quantum Falcon device, and support our experimental results with numerical simulations. Further, using a mid-circuit measurement tomography protocol we establish the nature of the errors identified by our benchmarking suite.

1. Introduction

Steady progress towards quantum error correction and fault tolerant quantum computing has led to recent experimental demonstrations of small quantum error correcting codes [1–10]. An essential component to most implementations of quantum error correction is the ability to repeatedly measure stabilizers of the code, often achieved via a stabilizer check circuit that encodes the outcome of the stabilizer measurement into the state of an ancilla qubit. Whether via an ancilla or measured directly [11, 12], stabilizer checks require fast and accurate mid-circuit measurement. Thus, characterizing and benchmarking mid-circuit measurement is a key capability for the development and execution of fault tolerant quantum computing.

Going beyond the typically measured state-assignment fidelity, quantum detector tomography [13–17] can be used to characterize terminal measurements in terms of a positive operator-valued measure (POVM). However, for the characterization of mid-circuit measurements a POVM description is insufficient. In this case full characterization of the measurement action leading to each outcome is described by a quantum channel, and hence process tomography is required [18–20]. While one can imagine extensions of these protocols to small stabilizer check circuits, the exponential resource scaling of process tomography make such characterization approaches impractical to deploy for larger quantum codes. While providing less detailed information about the measurement operation, there is a need for a scalable benchmark that can quickly assess the performance of mid-circuit measurement, and how it impacts not only the measured qubit but also those qubits connected to it.

Here, we introduce the mid-circuit measurement randomized benchmarking (mcm-rb) suite as one such benchmark. Building off the well-studied family of randomized benchmarking (RB) protocols [21–37], the mcm-rb suite comprises a central protocol, `mcm-rb`, that interleaves measurements of an ancilla qubit between the gates of Clifford RB performed on a distinct control qubit. The two other protocols of our suite are control experiments that replace either the measurement (`delay-rb`) or the Clifford gates (`mcm-rep`)

with delays of equal time duration to the replaced operation. Comparison between the decay curves of `mcm-rb` with the control experiments allow for the identification, and in some cases quantification, of the error induced by measurement on the control and ancilla qubits under the standard RB assumptions (i.e. Markovianity [22]). The mcm-rb suite is highly scalable, as it can be applied to many control and ancilla qubits at once to simultaneously benchmark measurement induced error, including measurement cross-talk.

This manuscript is organized as follows. In section 2 we describe the procedure of the mcm-rb suite, and in section 3 we discuss a classification of errors that the suite can detect. In section 4 we demonstrate the mcm-rb suite on an IBM Quantum device, and in section 5 we present supporting numerical simulations. Finally, in section 6 we discuss a limitation and potential extension of the protocol, and in section 7 we make our concluding remarks. Further details of our experimental and numerical results can be found in the Appendices. While writing this manuscript, we became aware of [38], which demonstrates the `mcm-rb` protocol, but not the rest of the suite, on a trapped ion system to study the impact of both mid-circuit measurement and reset.

2. Mid-circuit measurement RB suite

The mcm-rb suite is defined by the set of RB-style protocols (`mcm-rb`, `delay-rb`, `mcm-rep`), for which example circuits are shown in figure 1. The first protocol, `mcm-rb`, interleaves ancilla-qubit mid-circuit measurements between Clifford gates performed on the control qubit. Similar to simultaneous RB [26], it performs a single-subsystem twirl on the potentially two-qubit error induced by ancilla measurement. The second protocol, `delay-rb`, is analogous to interleaved RB (IRB) [25] on the control qubit, with the interleaved gate a noisy identity corresponding to a delay of equal duration to the ancilla-qubit measurement.

Together, these two protocols form an IRB procedure designed to detect errors on the control qubit induced by the ancilla-qubit measurement. Though it contains an interleaved gate itself, `delay-rb` is the reference sequence, and `mcm-rb` is the interleaved error sequence. It is important to reference `mcm-rb` by a sequence that contains interleaved delays in order to remove the trivial T_1 and T_2 decay of the control qubit during the potentially long measurement time, as this error may otherwise dominate other errors induced by the ancilla-qubit measurement. For example, for the experimental results of section 4 the measurement time approaches $1\,\mu s$, which should be compared to the much shorter gate time on the order of 10s of ns and the several hundred μs T_1 and T_2 times.

The final protocol, `mcm-rep`, is included to detect errors on the ancilla due to its own measurement, and is a modification of quantum non-demolition (QND) tests, e.g. [7]. Delays of equal duration to the control-qubit Clifford gates are interleaved between repeated measurements to keep all three protocols of equal duration (for a given sequence length), and to detect measurement and logical basis misalignment. We discuss the latter point in more detail in section 4. Comparing `mcm-rep` to `mcm-rb` is also useful for detecting certain kinds of two-qubit errors that require one qubit to be excited, of which we show an example in sections 4 and 5.

The mcm-rb suite is implemented similar to any RB-style protocol. A set of sequence lengths, $\{N_i\}_i$, is chosen, and for `mcm-rb` and `delay-rb` each sequence of length N_i consists of N_i random single-qubit Clifford gates on the control qubit interleaved by either measurements on the ancilla, or delays of equal duration, respectively. A final Clifford gate that is meant to invert the action of the previous N_i Cliffords terminates every circuit. At each sequence length, many random Clifford circuits are executed. For `mcm-rep` the circuit consists of N_i ancilla measurements interleaved by delays of equal duration to the control qubit Clifford gates. In this manuscript we have chosen $N_{\max} = 150$.

For all three protocols, the outcomes of the mid-circuit measurements are discarded, and the ground state probabilities for all control and ancilla qubits at the end of each circuit is estimated from the terminal measurement. These probabilities are averaged over the random Clifford circuits and for each control or ancilla qubit the decay as a function of sequence length is fit to the exponential function $P_0 = A\alpha^{N_i} + B$. The RB-decay parameter α defines the error per Clifford/measurement (EPC/M) for each qubit by EPC/M $= (1 - \alpha)/2$, while the other fit parameters A and B account for system preparation and measurement error (SPAM). Under the standard assumptions of RB [22] `mcm-rb` and `delay-rb` will follow an exponential decay curve from which SPAM can be isolated from EPC, but this is not guaranteed for `mcm-rep`. However, if the error on the measurement leads to monotonic decay then its EPM can be isolated from SPAM, though our fitting procedure makes the further (generically unnecessary) requirement that the decay is exponential. We have chosen not to Clifford twirl the measured ancilla qubits as doing so would require a final inverse gate that would be conditional on the full history of measurement outcomes.

While figure 1 shows as an example only a single control and ancilla qubit, the mcm-rb suite can be applied simultaneously to multiple control and ancilla qubits. This can be used, for example, to test the impact of measurement of a central ancilla on multiple control qubits, or to test the impact of the
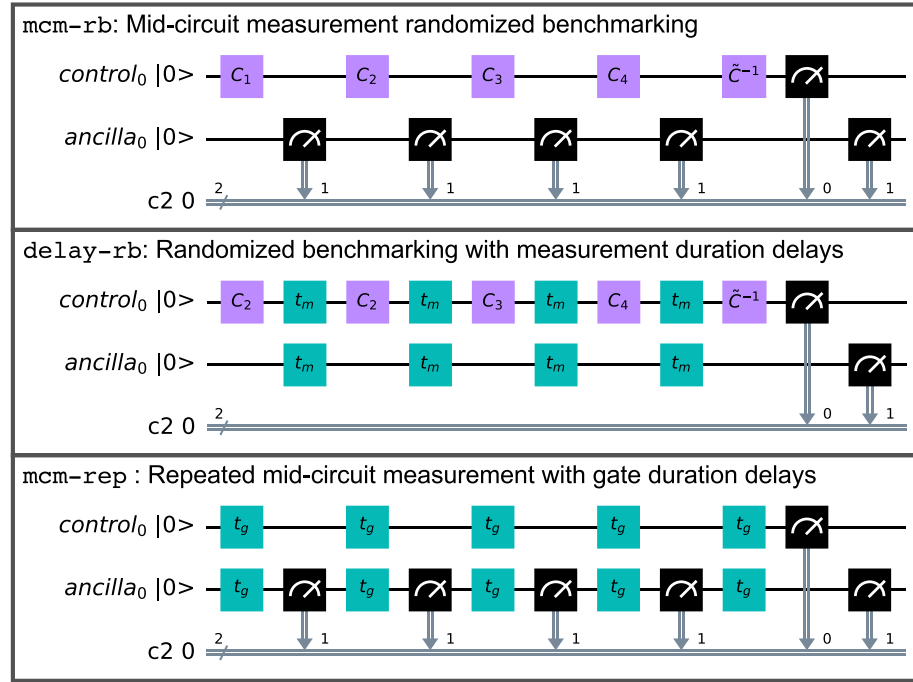
Figure 1. Example circuits of the mcm-rb suite protocols. Sequence length $N_i = 4$ for all three protocols. (Upper panel) `mcm-rb` circuit with Clifford gates on the control qubit interleaved by measurements on the ancilla. (Middle panel) `delay-rb` circuit where Clifford gates on the control qubit are interleaved with delays of duration t_m, the length of an ancilla measurement. (Lower panel) `mcm-rep` with repeated measurements on the ancilla qubit interleaved by delays of duration t_g, the length of a control qubit Clifford gate.

measurement of multiple ancilla qubits on a single control. In our experimental demonstrations of section 4 we study both applications by simultaneously performing the mcm-rb suite across 12 control and 5 ancilla qubits on our device. While simultaneous Clifford gates on control qubits can introduce cross-talk error, since `mcm-rb` and `delay-rb` operate under the same control conditions with simultaneous gates they equivalently experience cross-talk. Thus, we can still use `delay-rb` as the reference sequence to quantify the error induced by measurement in `mcm-rb`.

3. Error detection with mid-circuit measurement RB

In this section we demonstrate the capability of the mcm-rb suite (`mcm-rb`, `delay-rb`, `mcm-rep`) to detect, and in many case estimate the magnitude of, errors induced by mid-circuit measurement on either the control or (measured) ancilla qubit. To do so, rather than focus on the effects of specific errors, we classify the distinct *error signatures* that the mcm-rb suite's decay curves can exhibit, where each error signature can have more than one possible underlying physical error mechanism.

Error signatures are classified by comparing the EPC of the control and error per measurement (EPM) of the ancilla for the three components of the mcm-rb suite. From here one, we denote the EPC and EPM by ϵ_ν^q, with $q \in \{c, a\}$ for control and ancilla respectively, and $\nu \in \{\text{rb,del,rep}\}$ for `mcm-rb`, `delay-rb`, and `mcm-rep` respectively. Table 1 outlines the error signatures we consider, and the expected relationships between the various EPCs and EPMs.

It should be noted that multiple physical errors which result in distinct error signatures can occur simultaneously, such that the decay curves of the mcm-rb suite display a combination of the error signatures listed in table 1. In this case, one has to determine the likely underlying error signatures through a process of elimination given which ϵ_ν^q are nonzero. While it is impossible to distinguish some combinations, e.g. a non-QND measurement and either a measurement induced control or two-qubit error, there is sufficient information given by the mcm-rb suite to use either in a debugging cycle or to use with knowledge of the device physics to determine the likely error signatures and their underlying causes.

An important question is whether or not the mcm-rb suite can *quantify* the EPM or the added EPC due to mid-circuit measurement. Since we measure the EPM by exponential fit, to quantify the error due to mid-circuit measurement on the measured ancilla we require that the error process induce an exponential

Table 1. Error signatures detected by the mcm-rb suite that we consider in this paper. For each error signature, the expected relationships between the EPCs and EPMs are shown. Each ϵ_ν^q is an EPC/M with $q \in \{c, a\}$ for control and ancilla, and $\nu \in \{$rb,del,rep$\}$ for `mcm-rb`, `delay-rb`, and `mcm-rep`.

Error signature	EPC/M
No measurement induced error	$\epsilon_\nu^a \approx 0 \ \forall \nu$ $\epsilon_{\rm rb}^c \approx \epsilon_{\rm del}^c$, $\epsilon_{\rm rep}^c \approx 0$
Non-QND measurment error	$\epsilon_{\rm del}^a \approx 0$, $\epsilon_{\rm rb}^a > 0$, $\epsilon_{\rm rep}^a > 0$ $\epsilon_{\rm rb}^c \approx \epsilon_{\rm del}^c$, $\epsilon_{\rm rep}^c \approx 0$
Measurement induced control error	$\epsilon_\nu^a \approx 0 \ \forall \nu$ $\epsilon_{\rm rb}^c > \epsilon_{\rm del}^c$, $\epsilon_{\rm rep}^c \geqslant 0$
Measurement induced 2-qubit error	$\epsilon_{\rm del}^a \approx 0$, $\epsilon_{\rm rb}^a \geqslant 0$, $\epsilon_{\rm rep}^a \geqslant 0$ $\epsilon_{\rm rb}^c > \epsilon_{\rm del}^c$, $\epsilon_{\rm rep}^c \geqslant 0$
RB cross-talk error	$\epsilon_{\rm rep}^a \approx 0$, $\epsilon_{\rm rb}^a > 0$, $\epsilon_{\rm del}^a > 0$ $\epsilon_{\rm rb}^c \approx \epsilon_{\rm del}^c$, $\epsilon_{\rm rep}^c \approx 0$

decay of ground-state probability with sequence length . While this is not generically guaranteed, we argue in the following non-QND error subsection that it applies to a wide class of error models in the small error limit.

As for quantifying the added EPC due to mid-circuit measurement, as mentioned previously from the control qubit's perspective the pair of experiments `mcm-rb` and `delay-rb` together form an interleaved RB (IRB) protocol. `mcm-rb` interleaves a noisy identity operation on the control qubit, with the error induced by mid-circuit measurement on the ancilla. Thus, if the measurement induced error satisfies the necessary assumptions of IRB [25] we can quantify the error due to mid-circuit measurement as we would for any IRB procedure, with the error induced by measurement given by $\epsilon_{\rm IRM} = (1 - \alpha_{\rm rb}/\alpha_{\rm del})/2$. We note that as with any IRB estimate of the added error, one must exercise caution with respect to quantitative accuracy as the accuracy of IRB estimates is very sensitive to the underlying errors of the reference sequence (in our case `delay-rb`), as well as the nature of the interleaved error.

4. MCM-RB in experiment

In this section, we demonstrate the practical application of the mcm-rb suite on the IBM Quantum Falcon R8 device `ibm_peekskill`. To showcase the scalability of simultaneous mcm-rb, the mcm-rb suite experiments were performed in parallel on 5 ancilla-control qubit sets for a total of 17 qubits operating simultaneously. Two distinct 17-qubit configurations on `ibm_peekskill` were considered, such that 23 of the 27 qubits on `ibm_peekskill` were studied. For further details see appendix A, and for complete mcm-rb suite data on all 23 qubits see appendix C. Experimental data and Jupyter notebooks to reproduce the figures are available at [39].

Generally, all measured ancilla qubits had weak non-QND error, with two showing considerable decay after the longest sequence (150 mid-circuit measurements). Most control qubits showed some measurement induced control error, and for many control-ancilla pairs there was evidence of measurement induced two-qubit error. As such it was not always possible to distinguish between these two error signatures, but in a few cases the distinction was significant enough to be conclusive. There were no consistent patterns observed across the device. In the following, where possible we present an example of each error signature from table 1 using mcm-rb data taken on `ibm_peekskill`.
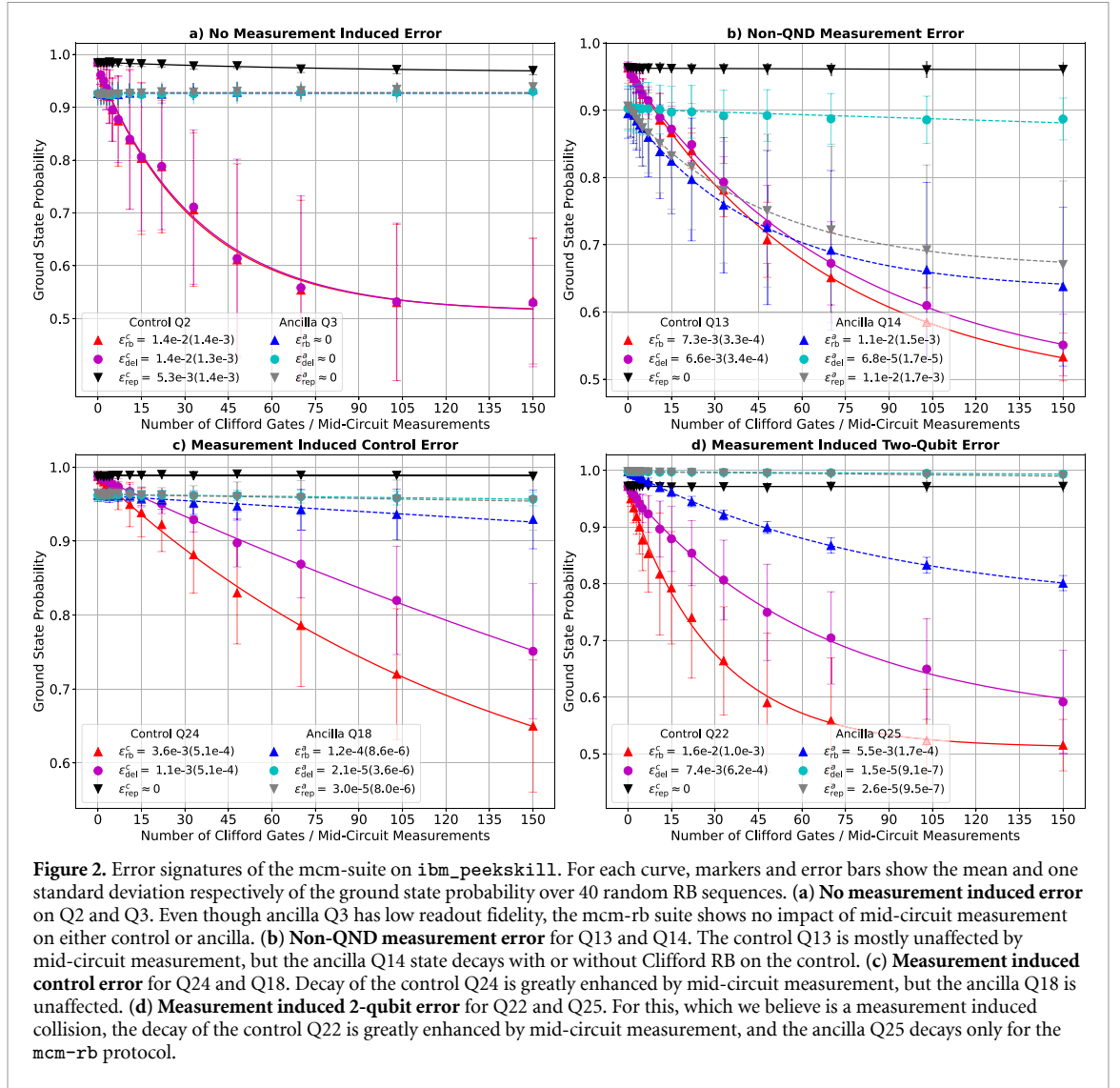
4.1. No measurement induced error
The trivial error signature occurs when the EPC for `mcm-rb` and `delay-rb` are indistinguishable from one another, the EPC is zero for `mcm-rep`, and the EPM is zero for all three experiments in the mcm-rb suite. In this case, interleaving mid-circuit measurements has no effect on either the control or ancilla qubit, which is the desired outcome for most applications. An example of no measurement induced error from `ibm_peekskill` is shown in figure 2(a). This pair of qubits was specifically chosen to also highlight that non-unity readout fidelity on the ancilla qubit does not impact the mcm-rb suite.

4.2. Non-QND measurement
A QND measurement is one for which the measurement operator commutes with the system Hamiltonian, and leaves the system in the logical eigenstate corresponding to the reported measurement outcome [40]. As such, non-QND measurement is commonly used as a catch-all term to describe any error that changes the state of the system from that reported, though this is only a subset of possible non-QND errors. One example is an error process that has a finite probability of flipping the state of the qubit after measurement.

However, measurements that project the system onto an eigenbasis that is not the logical basis, i.e. the logical and measurement bases are misaligned, are also non-QND. A protocol based on repeated measurements with no delays can detect non-QND errors such as measurement-induced state flips [7], but

Figure 2. Error signatures of the mcm-suite on `ibm_peekskill`. For each curve, markers and error bars show the mean and one standard deviation respectively of the ground state probability over 40 random RB sequences. (**a**) **No measurement induced error** on Q2 and Q3. Even though ancilla Q3 has low readout fidelity, the mcm-rb suite shows no impact of mid-circuit measurement on either control or ancilla. (**b**) **Non-QND measurement error** for Q13 and Q14. The control Q13 is mostly unaffected by mid-circuit measurement, but the ancilla Q14 state decays with or without Clifford RB on the control. (**c**) **Measurement induced control error** for Q24 and Q18. Decay of the control Q24 is greatly enhanced by mid-circuit measurement, but the ancilla Q18 is unaffected. (**d**) **Measurement induced 2-qubit error** for Q22 and Q25. For this, which we believe is a measurement induced collision, the decay of the control Q22 is greatly enhanced by mid-circuit measurement, and the ancilla Q25 decays only for the `mcm-rb` protocol.

will be insensitive to errors due to logical and measurement basis misalignment. This insensitivity is due to the fact that from the measurement's perspective misalignment errors are not an error, and repeated measurement with no delay will repeatedly project the system into the same measurement basis state, with no probability of a state flip. Thus, there would be no decay of the 'ground-state probability'. Only the first measurement, with preparation in the logical ground state, shows any evidence of the basis misalignment, but the error induced at this sequence length is functionally indistinguishable from SPAM.

On the other hand, as mcm-rb and mcm-rep have delays on the ancilla between repeated measurements to accommodate control Clifford gates, there is time for the ancilla logical Hamiltonian to evolve the system out of a measurement basis state. This results in a finite probability of a measurement basis state flip after each mid-circuit measurement, and thus nonzero $\epsilon_{\rm rb}^a$ and $\epsilon_{\rm rep}^a$, such that the mcm-rb suite can detect both kinds of non-QND error. Nevertheless, in our experimental system the mid-circuit measurements have been tuned up to mitigate the impact of Stark shifts and dephasing due to residual photons in the measurement resonator [41]. This removes a major source of misalignment errors present in our system [42], though others may still persist [43].

For a generic non-QND error, as there is no unitary twirl applied to the ancilla qubit we cannot guarantee exponential decay of its ground-state probability. As a result, the EPM estimated by an exponential fit may not be a faithful quantifier of the true EPM. However, in the appropriate limit exponential decay can be obtained if the error model after each mid-circuit measurement is the same, and results in a finite probability of the qubit leaving its initial state. Such error models are common, e.g. arising due to spurious coherent or incoherent qubit transitions driven by the measurement pulse, or the action of the logical Hamiltonian during the delay time between measurements with misaligned logical and measurement bases.

While transitions out of the computational subspace, i.e. leakage [44, 45], often lead to a similar exponential decay of the ancilla ground state probability, care must be taken to ensure that the qubit reset between shots is able to remove leaked population.

If the probability of a state flip after each measurement is p, then for an `mcm-rb` and `mcm-rep` sequence of length N the ground state probability for the terminal measurement is the probability of an even number of state flips during the sequence, which is given by

$$P_{\mathrm{GS}} = \frac{(1-p)^{N+2} - p^{N+2-w}(1-p)^w}{1-2p} \qquad (1)$$

with $w = 0$ for even N and $w = 1$ for odd N (see appendix A for a derivation). In the limit of a small error such that $p \ll 1$ this becomes approximately a single exponential decay with $P_{\mathrm{GS}} \propto (1-p)^N$. When this is the case, an exponential fit to the data allows us to quantify the amount of induced EPM.

An example of a non-QND measurement error signature from `ibm_peekskill` is shown in figure 2(b). The ancilla decay clearly shows the tell-tale signature of a non-QND error: $\epsilon_{\mathrm{rb}}^a \approx \epsilon_{\mathrm{rep}}^a > 0$, and we note that these decay curves are well fit by an exponential. While there is weak decay of the ancilla for long `delay-rb` sequences, with $\epsilon_{\mathrm{del}}^a \neq 0$, this is three orders of magnitude less than ϵ_{rb}^a, such that it is clear the measurement is negatively impacting the ancilla. The control EPCs with and without mid-circuit measurements are not indistinguishable, indicating that there may also be some measurement induced error on the control qubit for this qubit pair.

4.3. Measurement induced control error

The mcm-rb suite was designed to detect any errors induced by the measurement on the control qubit, and this error signature is indicative of an induced error that impacts *only* the control qubit, such that the ancilla qubit can be ignored as it is unaffected. This error signature can be understood as an additional error interleaved between the non-ideal Clifford gates on the control qubit. Assuming the usual requirements of IRB hold for the measurement induced error [25], e.g. it is Markovian and at most weakly gate dependent, then the standard IRB procedure can be used to compare the decay of the `mcm-rb` and `delay-rb` experiments to quantify the error induced on the control qubit by mid-circuit measurement. Note that it is important to use `delay-rb` as the reference sequence to capture the control-error due to interleaving a long delay between Clifford gates, and the magnitude of this delay error can be quantified using IRB comparing `delay-rb` to a standard RB experiment with no delay.

An example of a measurement induced control error signature from `ibm_peekskill` is shown in figure 2(c). For the control qubit EPCs, ϵ_{rb}^c is more than a factor of three larger than $\epsilon_{\mathrm{del}}^c$, indicating a significant impact of mid-circuit measurement on the control qubit. The EPMs for the ancilla qubit are all almost negligible, except for ϵ_{rb}^a, which indicates there may also be evidence for a weak two-qubit measurement induced error or RB cross-talk error.
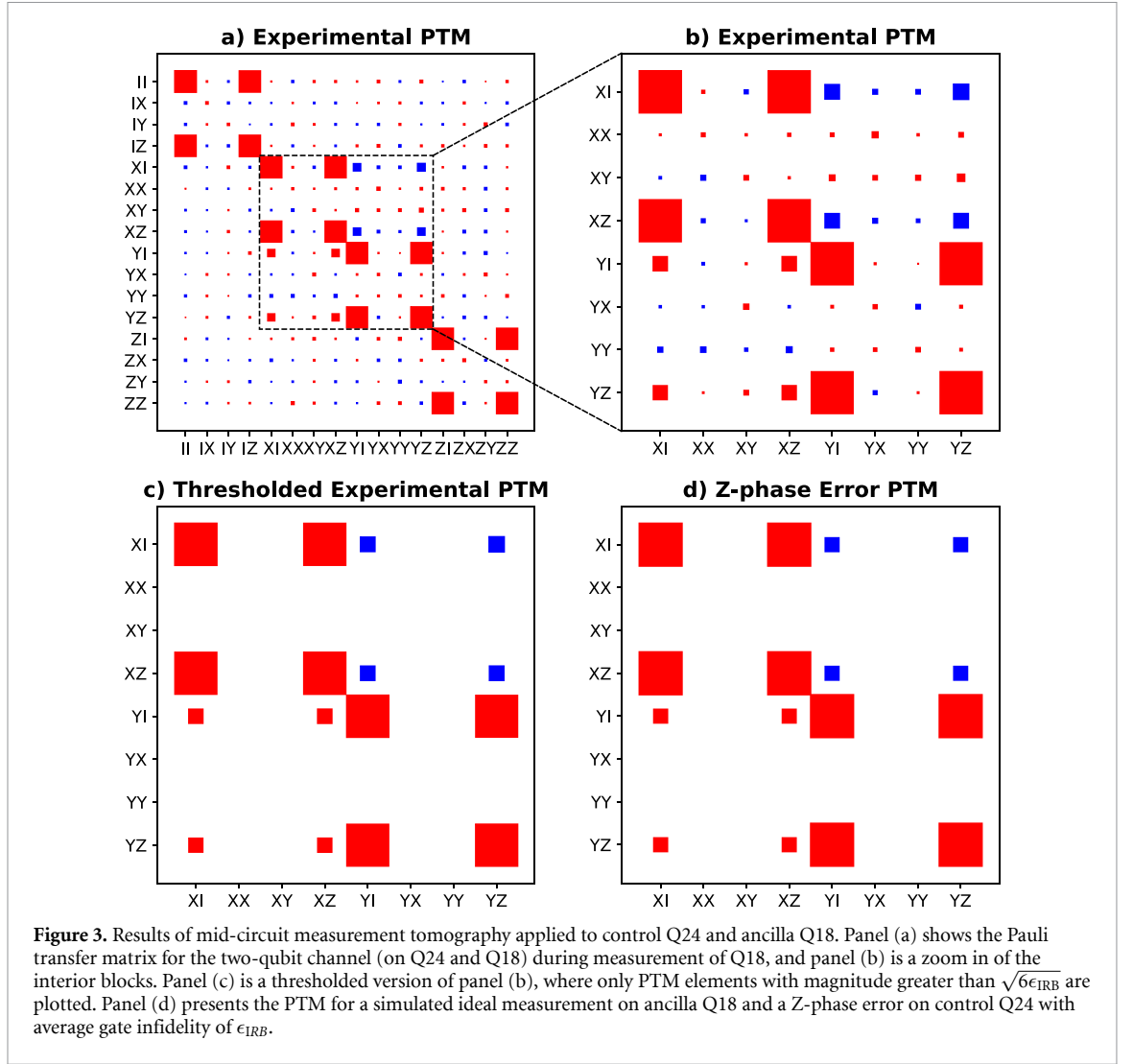
While the physical origin of this error cannot be determined by the mcm-rb suite, given the nature of the hardware platform and the fact that $\epsilon_{\mathrm{rep}}^c \approx 0$, it is likely due to either a measurement induced Stark shift, or weak cross-dephasing on the control qubit. The power of the mcm-rb suite is that it can quickly identify all such issues across an entire chip, which can then be explored in more detail with slower techniques to determine their origin.

In the case of this qubit pair, we performed a follow-up experiment that interleaved data collection for the mcm-rb suite protocols and a mid-circuit measurement tomography protocol. From the results of the mcm-rb suite we obtained an estimate of the control error induced by measurement of $\epsilon_{\mathrm{IRB}} = 1.7 \times 10^{-3} \pm 1.0 \times 10^{-3}$, which unfortunately is a good example of the large uncertainty bounds on IRB estimates. From the results of the measurement tomography we obtained the Pauli transfer matrix (PTM) shown in figure 3(a), which shows the expected signal for the ideal channel (the dominant block diagonal structure in red) along with many spurious non-zero elements arising from the error channel.

Zooming into the interior blocks in figure 3(b), the largest error source has a structure indicative of coherent Z-phase error induced by Stark shift. For a Z-phase error of angle θ induced on the control qubit by ancilla measurement, we would have that $\epsilon_{\mathrm{IRB}} = (1 - 4\cos(2\theta))/3$. For the PTM the non-zero elements due to the Z-phase error all have the same magnitude, given by

$$\left| R_{\hat{\mathbb{I}}\hat{Y}, \hat{\mathbb{I}}\hat{X}} \right| = |\sin(2\theta)| \approx \sqrt{6\epsilon_{\mathrm{IRB}}}, \qquad (2)$$

where in the last expression we have used the second order in θ series expansions of ϵ_{IRB} and $R_{\hat{\mathbb{I}}\hat{Y}, \hat{\mathbb{I}}\hat{X}}$ to relate the two quantities.

Figure 3. Results of mid-circuit measurement tomography applied to control Q24 and ancilla Q18. Panel (a) shows the Pauli transfer matrix for the two-qubit channel (on Q24 and Q18) during measurement of Q18, and panel (b) is a zoom in of the interior blocks. Panel (c) is a thresholded version of panel (b), where only PTM elements with magnitude greater than $\sqrt{6\epsilon_{\mathrm{IRB}}}$ are plotted. Panel (d) presents the PTM for a simulated ideal measurement on ancilla Q18 and a Z-phase error on control Q24 with average gate infidelity of ϵ_{IRB}.

Only plotting PTM elements with magnitude larger than $\sqrt{6\epsilon_{\mathrm{IRB}}}$ (using the mean of the estimate for ϵ_{IRB}) produces figure 3(c). This closely matches the PTM shown in figure 3(d) for a simulated Z-phase error with a θ calculated from ϵ_{IRB}. This demonstrates how the quantitative error benchmarking of the mcm-rb protocol can be used to help interpret the results of the more detailed mid-circuit measurement tomography, and in so doing obtain both the error magnitude and its nature.

4.4. Measurement induced two-qubit error

Unlike the error signatures we have thus far considered, there are sufficiently diverse measurement induced two-qubit errors that they will not all result in the same error signature. We do expect that any measurement induced two-qubit error will result in a faster decay of the control qubit for `mcm-rb` compared to `delay-rb`. Unfortunately, the control `mcm-rb` decay is not guaranteed to be exponential as only the control-qubit is twirled. However, from the simultaneous RB protocol [26] we know that in the limit of small two-qubit error the decay of a single-subsystem Clifford twirl will still be approximately exponential, such that we can again quantify the added error on the control using an IRB procedure comparing `mcm-rb` and `delay-rb`.

The signature of the ancilla decay curves is not consistent across the various error models that fall under measurement-induced two-qubit error. For example, if the error is a coherent excitation exchange between control and ancilla, then we would expect to see decay of the ancilla ground state probability only for the `mcm-rb` protocol but not the `mcm-rep` protocol, since the ancilla and control are both initialized in the ground state. On the other hand, a double excitation error (i.e. an *XX*-gate) induced by measurement would result in finite ancilla EPM for both `mcm-rb` and `mcm-rep`, while a correlated phase error (i.e. a *ZZ*-gate) would not impact the ancilla state.

An example of a measurement induced two-qubit error signature from `ibm_peekskill` is shown in figure 2(d). In this case, it is clear that this is a two-qubit error as we have both a substantial $\epsilon_{\mathrm{rb}}^a > 0$ and $\epsilon_{\mathrm{rb}}^c > \epsilon_{\mathrm{del}}^c$, with all three decay curves well fit by exponential functions. For this particular two-qubit error,

$\epsilon_{\text{rep}}^{a/c}$ is negligible compared to ϵ_{rb}^{a}. As we explain in more detail with a numerical simulation model in section 5, we attribute this particular kind of two-qubit error signature to a measurement induced collision. The ancilla qubit is Stark-shifted by the photons in the measurement cavity to a frequency close enough to the control qubit such that near-resonant excitation exchange can occur. The tell-tale signature that this is likely a collision is the fact that $\epsilon_{\text{rb}}^{a} \gg \epsilon_{\text{rep}}^{a}$, indicating that for a control qubit in its ground state the ancilla is not impacted.

4.5. RB cross-talk

Finally, we briefly discuss an error signature detectable by the mcm-rb suite, but which has nothing to do with mid-circuit measurement. If the implementation of the Clifford gates on the control qubit impacts the ancilla qubit, i.e. if there is cross-talk between the qubits for single qubit gates, then the `mcm-rb` and `delay-rb` curves for the ancilla will likely decay. However, due to the lack of a Clifford twirl on the ancilla, the mcm-rb suite cannot say much quantitatively about this error. By design the mcm-rb suite is meant to benchmark errors induced by measurement on the control qubit(s), and other protocols exist to benchmark [26, 35] or characterize [46] cross-talk.

5. Numerical simulation of MCM-RB

In the following subsections, for each non-trivial error signature of table 1 we use numerical simulation to study example physical error mechanisms that lead to the error signature. Our simulations are performed using the quantum circuit simulator in Qiskit Aer [47], which natively supports error processes such as the depolarizing channel and decoherence generated by qubit T_1 decay and T_2 dephasing. Additionally, one can add custom error processes either as unitary gates or via their Kraus decomposition, and we make use of this functionality for several of the measurement induced errors considered in the subsequent subsections.

In addition to the measurement induced error, our simulations add depolarizing error to each single qubit gate on the control qubit, as well as an amplitude and phase damping channel to the control qubit for each ancilla measurement. The latter is equivalent to the decoherence generated by control qubit T_1 and T_2 decoherence during the measurement, with parameters chosen to be representative of our experimental setup. For further details see appendix B.

5.1. Non-QND measurement

To model non-QND measurement, we take a simple approach and consider the application of a depolarizing channel after each mid-circuit measurement. A single-qubit depolarizing channel acts as

$$\mathcal{E}_{\text{dep}}(\rho) = (1-\eta)\rho + \eta\frac{\hat{\mathbb{I}}}{2}, \tag{3}$$

with η the depolarizing probability. For our simulations we scan η from 2% to 20%, as shown in figure 4. The lower panel shows an example of the expected mcm-rb suite decay curves for a non-QND measurement error, in this case for $\eta = 2\%$. The upper panel shows that the EPM for both `mcm-rb` and `mcm-rep` accurately estimate the simulated EPM, which for the depolarizing channel is $\eta/2$.

5.2. Measurement induced control error

We consider two models for physical error mechanisms that induce control qubit errors. The first is a measurement-induced Stark shift, which adds a Z-phase error to the control qubit after every measurement via the unitary $\hat{U}_{\text{Stark}} = e^{-i\phi\hat{\sigma}_z}$. For a cQED system such as `ibm_peekskill`, this can occur when readout photons intended for the ancilla-qubit resonator populate the control-qubit resonator. Then, via the dispersive interaction $\hat{H} = \chi\hat{\sigma}_z^c\hat{n}$ between the control-qubit and its readout resonator, the control frequency is Stark shifted by $2\chi\bar{n}$ during measurement, where $\bar{n}$ is the average number of photons in the resonator. For a mid-circuit measurement of duration t_m this leads to a Stark-phase error with $\phi = 2\chi\bar{n}t_m$.

The second model is cross-measurement, where with some probability p_m the measurement of the ancilla also strongly measures the control qubit. This error is described by the quantum channel, $\mathcal{E}_{p_m}^c$, that completely dephases the control qubit with probability p_m. The Kraus representation of this channel is given by

$$\hat{K}_0 = \sqrt{p_m}|0\rangle\langle0|, \quad \hat{K}_1 = \sqrt{p_m}|1\rangle\langle1|, \quad \hat{K}_2 = \sqrt{1-p_m}\hat{\mathbb{I}}. \tag{4}$$

In a continuous time model this error could also be described by dephasing on the control qubit during the measurement with a rate γ defined by $e^{-\gamma t_m} = 1 - p_m$. In a cQED system, this error can occur due to readout
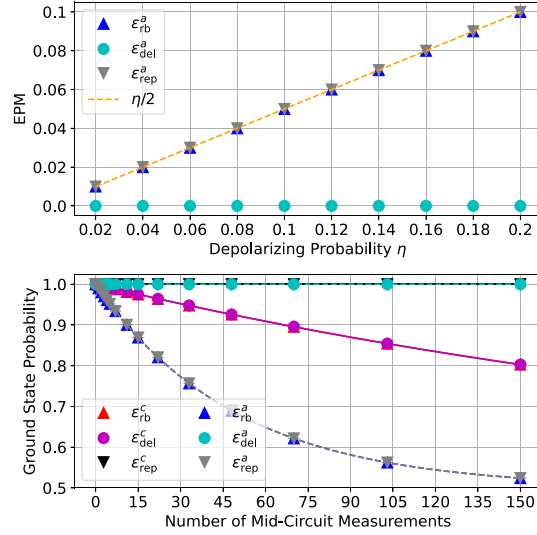
Figure 4. Non-QND Measurement Simulations. (Upper panel) EPM for the ancilla qubit for the three mcm-rb suite experiments. The EPM for `mcm-rb` and `mcm-rep` accurately estimates the simulated EPM of $\eta/2$, shown in the dashed orange line. (Lower panel) Example mcm-rb suite decay curves for a non-QND measurement error due to a depolarizing channel with depolarizing probability $\eta = 2\%$. In addition to the ancilla depolarizing channel, we apply an amplitude and phase damping channel corresponding to dissipation with $T_1 = 345\,\mu s$, $T_2 = 280\,\mu s$, and duration $t_m = 0.71\,\mu s$ on the control qubit.
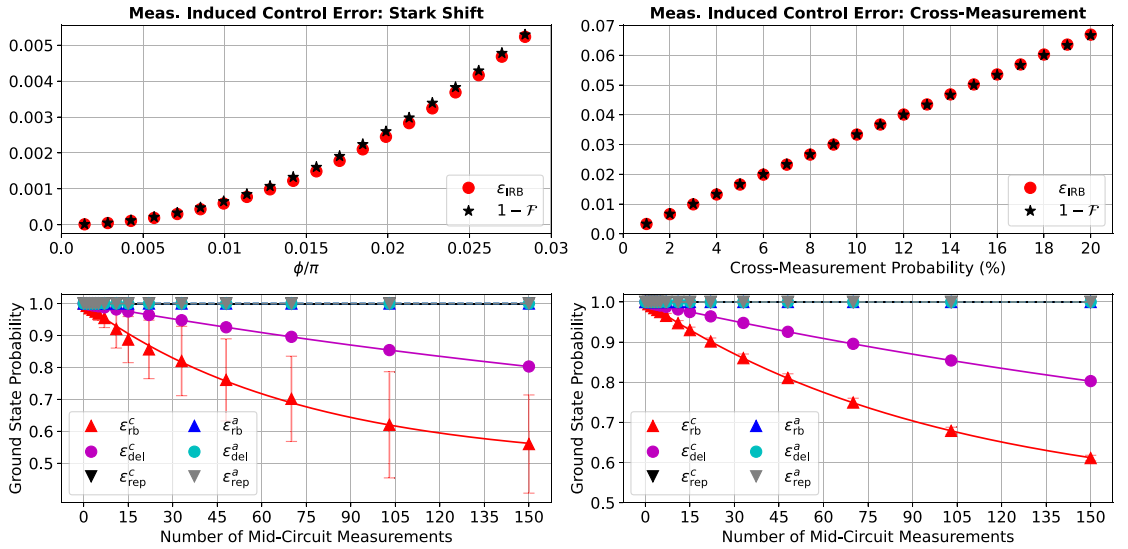
Figure 5. Measurement induced control error. Left column: Z-phase error due to control-qubit Stark shift. Right column: error due to a cross-measurement on the control qubit as described by the channel of equation (4). For both simulations in addition to the aforementioned errors we apply an amplitude and phase damping channel corresponding to dissipation with $T_1 = 345\,\mu s$, $T_2 = 280\,\mu s$, and duration $t_m = 0.71\,\mu s$ on the control qubit. The upper panels show the added error due to measurement as estimated by IRB, $\epsilon_{\mathrm{IRB}} = (1 - \alpha_{\mathrm{rb}}/\alpha_{\mathrm{del}})/2$, compared to the average gate infidelity $1 - \mathcal{F}$ of the simulated process. In both cases there is good agreement between the estimated and exact values. The lower panels show example mcm-rb suite decay curves, for $\phi/\pi \approx 0.03$ and $p_m = 0.01$. Despite their similar decay rates, note the considerable spread in the control-qubit `mcm-rb` decay curve for the coherent error compared to the incoherent error.

photons that leak into the control-qubit resonator, and is also possible in an ion trap system due to scattering of the readout laser pulse or photons fluoresced by the ancilla-qubit [38].

The results of our simulations are shown in figure 5, with the left column showing the results for the Stark shift error model and the right column for the cross-measurement error model. The lower panels show examples of the mcm-rb suite decay curves for these error models, and it is important to highlight that despite very different underlying physics, they produce the same error signature: $\epsilon_{\mathrm{rb}}^c \gg \epsilon_{\mathrm{del}}^c$ with all other EPC/M zero. This is to be expected as they are both errors that impact only the control qubit.

The upper panels show the error induced on the control qubit per mid-circuit measurement as estimated by IRB (ϵ_{IRB}), and compare to the average gate infidelity, i.e. $1 - \mathcal{F}$, with $\mathcal{F}$ the average gate fidelity which can be calculated exactly for these error channels. In both cases the IRB estimate is reasonably accurate, but it
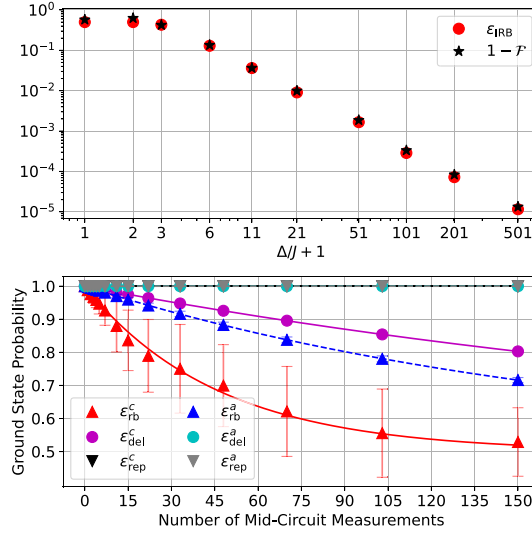
Figure 6. Measurement induced two-qubit error. The measurement induced collision error model of equation (5) parameterized by the ratio of the detuning and coupling, Δ/J. We also apply an amplitude and phase damping channel corresponding to dissipation with $T_1 = 345\,\mu$s, $T_2 = 280\,\mu$s, and duration $t_m = 0.71\,\mu$s on the control qubit. (Upper panel) Added error due to measurement as estimated by IRB, ϵ_{IRB}, compared to the average gate infidelity $1 - \mathcal{F}$. Note that the values of the *x*-axis are shifted by one to accommodate the log-log scale. (Lower panel) Example mcm-rb suite decay curves for a measurement induced collision with $\Delta/J = 20$.

is noticeably less so for the coherent Stark shift error model. This highlights the caution necessary when using IRB, especially for coherent error and if a high degree of accuracy is required. Comparing the mcm-rb suite decay curves of the lower panels, it is unsurprising that the IRB estimate is less accurate for the coherent Stark-shift error, given the much larger spread in the control-qubit decay curves for `mcm-rb` observed for that error model.

5.3. Measurement induced two-qubit error

From the broad class of possible two-qubit errors we consider one that is physically motivated by the superconducting hardware platform, and has an interesting error signature. In particular, we explore the impact of a measurement induced collision, where the measurement induced Stark shift on the ancilla qubit brings it close to resonance with the control qubit. We consider a minimal model for this system that is platform agnostic, described by a coupled control-ancilla system. The Hamiltonian describing their interaction is

$$\hat{H} = \frac{\Delta}{2}\hat{\sigma}_z^a + J\left(\hat{\sigma}_-^a \hat{\sigma}_+^c + \hat{\sigma}_+^a \hat{\sigma}_-^c\right), \tag{5}$$

where the qubits are in a frame rotating at the frequency of the control qubit such that $\Delta = \omega_a - \omega_c$, with ω_a the Stark-shifted frequency of the ancilla qubit. To implement our error model in simulation, after every measurement we add the unitary error $\hat{U}_{\text{Col}} = e^{-i\hat{H}t_m}$, which corresponds to evolving the system for a time t_m under the evolution of the collision Hamiltonian.

Figure 6 shows the results of these simulations. As in our other simulations, the upper panel shows that the error added by measurement can be accurately estimated using an IRB procedure to compare ϵ_{rb}^c and ϵ_{del}^c. See appendix B for further information on our calculation of the average gate infidelity, $1 - \mathcal{F}$, of the effective single qubit channel on the control qubit induced by this two-qubit error channel. We note that even though the IRB prediction is accurate, for $\Delta/J < 5$ the decay curves are not well fit by exponential functions, as the error induced by the two-qubit channel is too large for the single-subsystem twirl that is performed [26].

The lower panel shows an example of the expected mcm-rb suite decay curves ($\Delta/J = 20$), with $\epsilon_{\text{rb}}^c \gg \epsilon_{\text{del}}^c$ and finite ϵ_{rb}^a, all of which display exponential decay. It is the finite ϵ_{rb}^a that distinguishes a two-qubit error from an error only on the control qubit. As it requires one qubit to be at least partially excited, $\epsilon_{\text{rep}}^a = 0$ for a measurement induced collision, but other two-qubit error sources may have finite ϵ_{rep}^a.

6. Discussion

For our demonstrations of the mcm-rb suite we have exclusively focused on the scenario where the ancilla qubit is initially prepared in the ground state. An equivalent set of experiments could be performed with the

ancilla qubit prepared in the excited state, and this would return different results if the effective error channel on the control qubit depends on the ancilla state. On its own this is not problematic, and one could simply repeat the mcm-rb suite for both ancilla initial states, or randomize if the average channel is of more interest. However, care must be taken if measurement can induce ancilla state flips, as the change in the instantaneous control-qubit error channel due to the change in ancilla state would result in an overall non-Markovian control-qubit error channel.

As an example, consider the experimentally relevant situation of an ancilla with relaxation characterized by a timescale T_1. If the ancilla is initialized in the excited state, then for the initial duration of an RB sequence ($t \ll T_1$) the effective error channel on the control qubit is approximately static, given by the error channel conditioned on the ancilla in the excited state. For sequences with duration longer than T_1, near the end of the sequence ($t \gg T_1$) the effective control-qubit error is again approximately static, but now given by the error channel conditioned on the ancilla in the ground state[4].

Crucially, at some point during the sequence the control-qubit error channel changes, and thus across the total sequence the control-qubit error cannot be consistently defined by one single-qubit quantum channel. The control-qubit error is temporally correlated across the sequence in a non-trivial way, with quasi-static error that exhibits at most one switch during a given sequence. The impact of such temporal correlations on RB has been previously studied [48], and in practice it does not preclude EPC estimation, but may make the estimates less reliable and require more random sequences for convergence.

In our experimental system, the T_1 time is a factor of 2 or 3 longer than the longest sequences we use in our mcm-rb experiments. Our system exhibits ancilla-state-dependent control-qubit error due to the presence of weak *ZZ*-coupling between many of the qubits on the device. Due to the fact that gates are calibrated with all spectator qubits in the ground state, this results in a coherent error on the control qubit only for an excited ancilla, $\hat{U}_{ZZ} = e^{-i\hat{H}_{ZZ}t_m}$, described by the Hamiltonian

$$\hat{H}_{ZZ} = \nu \, |e\rangle\langle e|_a \otimes \hat{\sigma}_z^c, \tag{6}$$

where ν is the *ZZ*-coupling rate.

We simulate the impact of this error channel combined with relaxation of the ancilla-qubit on the `mcm-rb` protocol with the ancilla initialized in the excited state, and the results of these simulations are shown in figure 7. When T_1 is very short (e.g. 0.1 or 1.0 μs) the ancilla relaxes almost immediately, such that only the first gates in a given sequence experience the error $\hat{U}_{ZZ}$. All but the first few sequence lengths are well fit by an exponential with an ϵ_{rb}^c calculated from the control-qubit error model for the ancilla in the ground state. For very long T_1 (e.g. 100 μs), only the longest sequences are likely to experience ancilla relaxation. Aside from small deviations at the end, the full decay curve is well fit by an exponential with ϵ_{rb}^c calculated from $\hat{U}_{ZZ}$. For intermediate T_1 (e.g. 10 μs) the fit quality decreases significantly for the longer sequences where an ancilla relaxation, and thus an inconsistency in the control-qubit error, is likely to occur.

As these simulations show, the non-Markovian characteristic of the control-qubit error induced by the combination of *ZZ*-coupling and ancilla relaxation reduce the reliability of the EPCs obtained from RB fitting. Though our ancilla qubits have an average $T_1 > 100\,\mu$s, to avoid the complications of temporally correlated error in RB we have focused on benchmarking mid-circuit measurement with the ancilla in the ground state.

One possible way to overcome this issue would be to randomize either the initial ancilla state preparation, or randomize state re-initialization after each mid-circuit measurement. This can be done by randomly inserting identity or $\hat{X}$-gates at the start of the circuit, or after each mid-circuit measurement. A further extension would be to consider the full Pauli-twirl of the mid-circuit measurement, so that the action on the control qubits was guaranteed to be a stochastic channel [49]. In aggregate, the control qubit will then experience the average error channel induced by ancilla measurement, unconditioned on the ancilla state.

Similarly, we could randomize the initialization of the control qubit, which should not impact `mcm-rb` or `delay-rb`, but could potentially change the result of `mcm-rep`. This would be the case, for example, if the measurement induced an amplitude damping error on the control qubit. We leave exploration of these extensions of the mcm-rb suite, and how their measured EPCs connect to experimentally relevant quantities, for future study.

---

[4] Note that while the celebrated quantum Zeno effect plays an important role in determining exactly how long a sequence needs to be before the ancilla has likely decayed, only for measurements repeated infinitely quickly would the ancilla by pinned in its excited state in perpetuity.
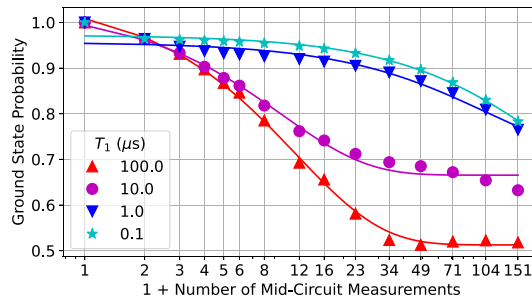
Figure 7. Non-Markovian Control Error Induced by the Ancilla. Control-qubit decay of the `mcm-rb` protocol, i.e. ϵ_{rb}^{c}, for the ZZ-coupling error model of equation (6) combined with ancilla relaxation decay for varying values of the relaxation timescale T_1. For all curves, the ancilla-qubit was initialized in the excited state, $\nu = 50$ kHz, and t_m is as before. Note that the x-axis is shifted by 1 to accommodate a value of 0 on the log-scale.

7. Conclusion

In this work we have presented a RB suite for mid-circuit measurements, whose central protocol interleaves mid-circuit measurements on an ancilla qubit between Clifford gates on a control qubit. The remaining two protocols of the suite replace either the mid-circuit measurement or the Clifford gates with idle delays of equal duration, and serve as reference experiments to enable error quantization through an interleaved RB procedure. As we have demonstrated on an IBM Quantum Falcon device, our benchmarking suite can be trivially extended to an entire multi-qubit chip, benchmarking multiple control and ancilla qubits simultaneously.

The mcm-rb suite classifies errors based on their error signature, which is the relationship between the RB-decay curves from the three protocols in the suite for both the control and ancilla qubits. We discussed the three major error signatures: non-QND measurement error, control-qubit error, two-qubit error; and highlighted examples of these error signatures from our deployment of the mcm-rb suite on an IBM Quantum Falcon device. Each error signature can be the result of many different physical error models, and we explored several in numerical simulation. By comparing to the average infidelity of our simulation models, we demonstrated that the mcm-rb suite can function as an IRB procedure and quantify the error added by mid-circuit measurement.
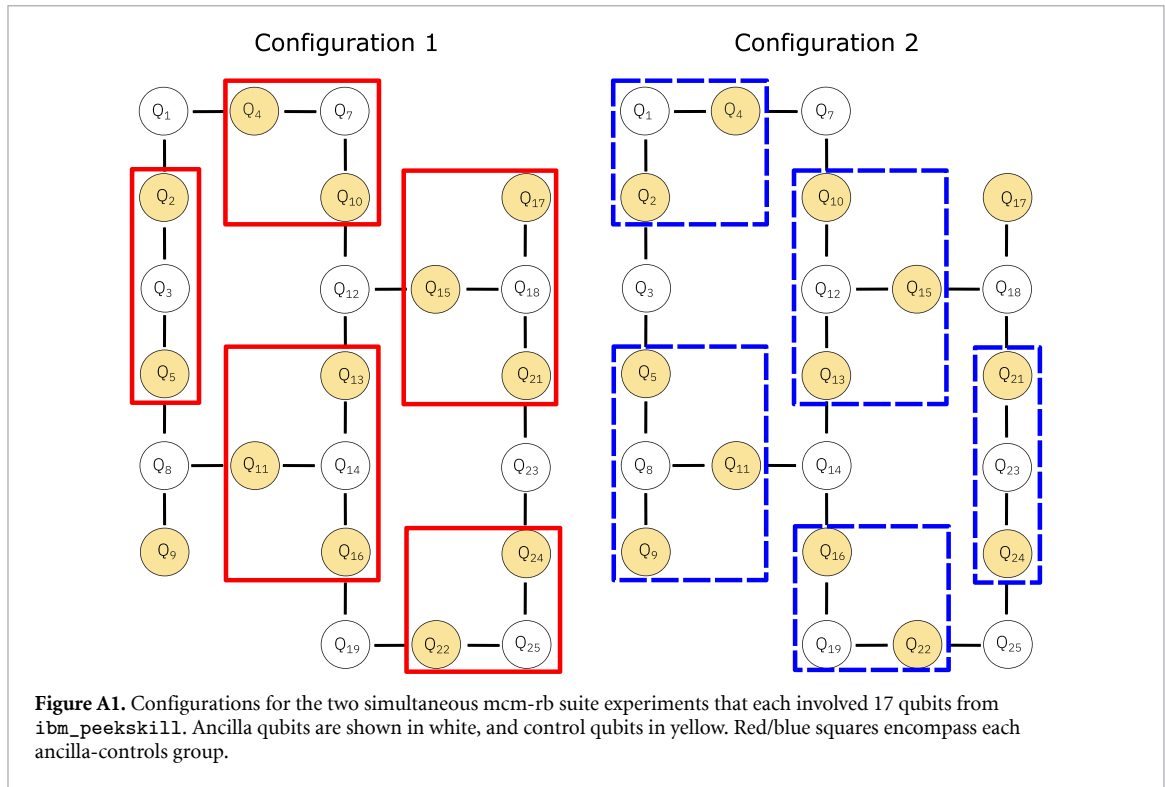
Our benchmarking suite can be readily adapted to other quantum–classical operations beyond mid-circuit measurement. These include a larger part of or even the full circuit for a stabilizer check, and real-time operations such as measurement and feed-forward [50]. While we have motivated mid-circuit measurement by its necessity in fault-tolerant quantum computing, many proposed near-term algorithms would benefit from this capability or the real-time operations it enables [51–54]. Thus, we expect the mcm-rb suite and developments upon it to also have immediate impact in characterizing devices for near-term applications.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://zenodo.org/record/6815663.

Acknowledgment

Figure A1. Configurations for the two simultaneous mcm-rb suite experiments that each involved 17 qubits from `ibm_peekskill`. Ancilla qubits are shown in white, and control qubits in yellow. Red/blue squares encompass each ancilla-controls group.

Appendix A. Experimental details

The mode of operation we employ to perform the mcm-rb sutie simultaneously across `ibm_peekskill` is to parallelize both over control and ancilla qubits. We break the full set of qubits into groups that each consist of one ancilla qubit and two or three control qubits. As shown in figure A1, with just two configurations of 17 qubits each we can cover 23 of the 27 qubits on `ibm_peekskill`. For each protocol that involved Clifford RB, we used 40 random sequences for each of the 15 sequence lengths, and took 1024 shots for each sequence. Each configuration therefore consisted of 1800 circuits total, which was broken into five jobs that were run sequentially.

In section 4.2 we consider a model for non-QND measurement error where after each mid-circuit measurement the state of an ancilla qubit has probability p of flipping to the orthogonal state. As we start in the ground-state the probability we end in the ground state is the probability that there have been an even number of state flips in the sequence. This probability is the sum of the probabilities of all possible even numbers of flips, which for N mid-circuit measurements is given by

$$P_{\mathrm{GS}} = \sum_{k}^{\lfloor N/2 \rfloor} p^{2k} (1-p)^{N-2k}. \tag{A.1}$$

The closed form expressions for N even and N odd are

$$P_{\mathrm{GS}} = \begin{cases} \frac{(1-p)^{N+2} - p^{N+2}}{1-2p} & N \text{ even} \\ \frac{(1-p)\left((1-p)^{N+1} - p^{N+1}\right)}{1-2p} & N \text{ odd} \end{cases}, \tag{A.2}$$

which can be unified into the single expression given in equation (1).

Appendix B. Numerical simulation details

Jupyter notebooks implementing our simulations can be found at [39]. The Qiskit Aer simulator we employ is a gate-based simulator of a quantum circuit, with non-unitary noise channels implemented in Kraus form. To simulate the action of mid-circuit measurement, we apply a completely dephasing channel to the ancilla

Table B1. Error channels used for each of our simulations. $\mathcal{E}^c_{T_1,T_2}$ applies an identity channel to the ancilla, and a phase and amplitude damping channel to the control qubit equivalent to relaxation and dephasing with $T_1 = 345\,\mu s$, $T_2 = 280\,\mu s$, and a duration $t_m = 0.71\,\mu s$, which is meant to be representative of our experimental device. $\mathcal{E}^{a,c}_{T_1,T_2}$ applies the same channel as $\mathcal{E}^c_{T_1,T_2}$ to the control qubit, and a phase and amplitude damping channel to the ancilla qubit with varying T_1 (see figure 7) and $T_2 = T_1/3$. $\mathcal{E}^a_{\mathrm{dep}}$ implements identity on the control and the depolarizing channel of equation (3) on the ancilla. $\hat{U}^c_{\mathrm{Stark}}$ and $\mathcal{E}^c_{p_m}$ apply identity on the ancilla, while on the control applying the Stark-shift unitray Z-phase error and the cross-measurement error channel of equation (4), respectively. $\hat{U}_{\mathrm{Col}}$ implements the two-qubit unitary error of equation (5).

Simulation	$\mathcal{E}_{\mathrm{pre}}$	$\mathcal{E}_{\mathrm{post}}$
non-QND measurement	$\mathcal{E}^c_{T_1,T_2}$	$\mathcal{E}^a_{\mathrm{dep}}$
Stark shift	$\hat{U}^c_{\mathrm{Stark}}$	$\mathcal{E}^c_{T_1,T_2}$
Cross-measurement	$\mathcal{E}^c_{p_m}$	$\mathcal{E}^c_{T_1,T_2}$
Frequency collision	$\hat{U}_{\mathrm{Col}}$	$\mathcal{E}^c_{T_1,T_2}$
ZZ-coupling	$\hat{U}_{ZZ}$	$\mathcal{E}^{a,c}_{T_1,T_2}$

qubit, $\mathcal{E}^a_{\mathrm{m}}$, which is described by the Kraus operators $K_0 = |0\rangle\langle 0|$ and $K_1 = |1\rangle\langle 1|$. As we discard the outcomes of our mid-circuit measurements in experiment, $\mathcal{E}^a_{\mathrm{m}}$ is equivalent to the action of ideal mid-circuit measurement in the ensemble average.

To add errors to the mid-circuit measurement, we sandwich each application of $\mathcal{E}^a_{\mathrm{m}}$ with pre-measurement and post-measurement two-qubit noise channels, which we label $\mathcal{E}_{\mathrm{pre}}$ and $\mathcal{E}_{\mathrm{post}}$. Table B1 shows the choice of $\mathcal{E}_{\mathrm{pre}}$ and $\mathcal{E}_{\mathrm{post}}$ for each of our simulations. In addition to the errors before and after measurement, all single-qubit gates in our simulation have a depolarizing error channel ($\eta = 10^{-3}$) applied after the action of the gate.

For simulations of `mcm-rb` we use 60 random Clifford sequences at each sequence length, and we sweep the error parameter of each model to generate the data points shown in the upper panels of the figures in section 5. The calculation of the exact average gate infidelity, $1 - \mathcal{F}$, for the two single-qubit control error channels can be done analytically [55], and the expressions are

$$1 - \mathcal{F}_{\mathrm{Stark}} = \frac{1}{3}\left(1 - \cos\left(2\phi\right)\right), \tag{B.1}$$

$$1 - \mathcal{F}_{\mathrm{CM}} = \frac{p_m}{3}. \tag{B.2}$$

To calculate the average gate infidelity on the control qubit for the two-qubit collision error, we must first calculate the effective single-qubit channel this error induces on the control qubit. To do so, following the approach of [56], we first construct the Choi state of the two-qubit channel

$$\sigma_{\mathcal{U}_{\mathrm{col}}} = \frac{1}{4}\sum_{j,k}|j\rangle\langle k| \otimes \mathcal{U}_{\mathrm{col}}\left(|j\rangle\langle k|\right), \tag{B.3}$$

where $\{|j\rangle\}$ is an orthonormal basis for the two-qubit Hilbert space and $\mathcal{U}_{\mathrm{col}}(\rho) = \hat{U}_{\mathrm{col}}\rho\hat{U}^\dagger_{\mathrm{col}}$ is the quantum channel representation of the unitary $\hat{U}_{\mathrm{col}}$.

For a quantum channel $\mathcal{E}$ that acts on the linear operator space of a Hilbert space $\mathcal{H}$ the Choi state is constructed by acting with the channel $\mathcal{I} \otimes \mathcal{E}$ on a maximally entangled state of the Hilbert space $\mathcal{H} \otimes \mathcal{H}$, where $\mathcal{I}$ is the identity channel. In our case, $\mathcal{H} \otimes \mathcal{H} = \mathcal{H}^a \otimes \mathcal{H}^c \otimes \mathcal{H}^a \otimes \mathcal{H}^c$, where $\mathcal{H}^{c/a}$ is the Hilbert space for the control/ancilla qubit. To calculate the effective channel on the control qubit alone, we perform a partial trace of $\sigma_{\mathcal{U}_{\mathrm{col}}}$ over the two copies of the ancilla Hilbert space

$$\sigma^c_{\mathcal{U}_{\mathrm{col}}} = \frac{1}{4}\sum_{j,k}\mathrm{Tr}_a\left[|j\rangle\langle k|\right] \otimes \mathrm{Tr}_a\left[\mathcal{U}_{\mathrm{col}}\left(|j\rangle\langle k|\right)\right]. \tag{B.4}$$

For each value of Δ/J in figure 6 we perform this partial trace numerically to calculate the effective control qubit error channel Choi state, from which we can extract the average gate infidelity.
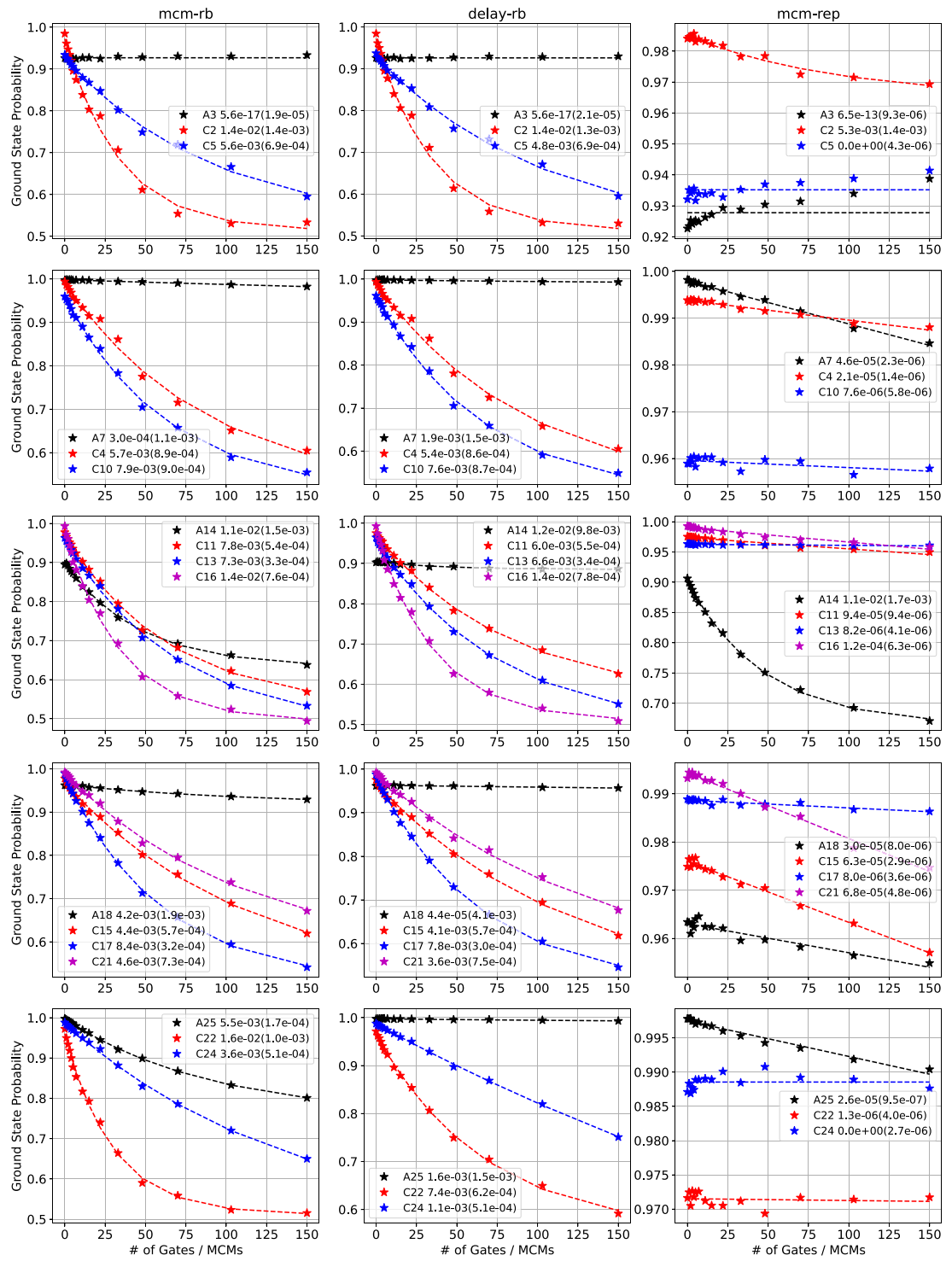
Figure C1. The mcm-rb suite decay curves for the qubits studied in configuration 1 of figure A1 on `ibm_peekskill`. Each row corresponds to a distinct ancilla-controls group, and each column is one protocol from the mcm-rb suite.

Appendix C. Full mcm-rb suite data

The full set of mcm-rb suite decay curves for configuration 1 and configuration 2 on `ibm_peekskill` are shown in figures C1 and C2 respectively. For these plots, each row corresponds to a distinct ancilla-controls group shown in figure A1. Each column is one of the three protocols in the mcm-rb suite.
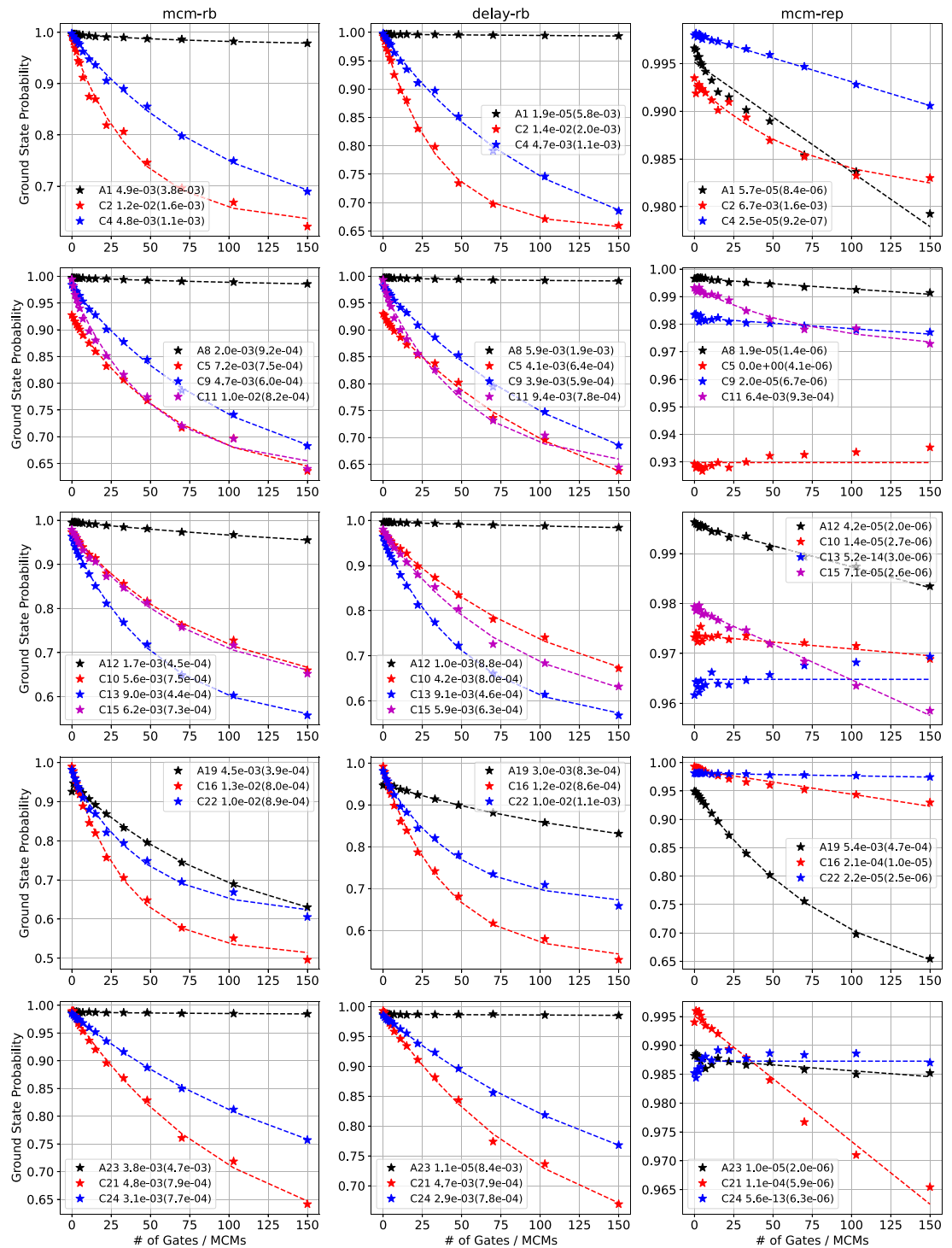
Figure C2. The mcm-rb suite decay curves for the qubits studied in configuration 2 of figure A1 on `ibm_peekskill`. Each row corresponds to a distinct ancilla-controls group, and each column is one protocol from the mcm-rb suite.

ORCID iDs

L C G Govia ● https://orcid.org/0000-0002-3263-648X
P Jurcevic ● https://orcid.org/0000-0003-1234-6386
C J Wood ● https://orcid.org/0000-0001-7606-7349

References

[1] Egan L *et al* 2021 *Nature* **598** 281–6
[2] Chen Z and Google Quantum AI 2021 *Nature* **595** 383–7

[3] Postler L *et al* 2022 *Nature* **605** 675–80

[4] Satzinger K J and Google Quantum AI 2021 *Science* **374** 1237–41

[5] Ryan-Anderson C *et al* 2021 *Phys. Rev. X* **11** 041058

[6] Abobeih M H, Wang Y, Randall J, Loenen S J H, Bradley C E, Markham M, Twitchen D J, Terhal B M and Taminiau T H 2022 *Nature* **606** 884–9

[7] Chen E H, Yoder T J, Kim Y, Sundaresan N, Srinivasan S, Li M, Córcoles A D, Cross A W and Takita M 2022 *Phys. Rev. Lett.* **128** 110504

[8] Krinner S *et al* 2022 *Nature* **605** 669–74

[9] Zhao Y *et al* 2022 *Phys. Rev. Lett.* **129** 030501

[10] Sundaresan N, Yoder T J, Kim Y, Li M, Chen E H, Harper G, Thorbeck T, Cross A W, Córcoles A D and Takita M 2023 *Nat. Commun.* **14** 2852

[11] Ristè D, Dukalski M, Watson C A, de Lange G, Tiggelman M J, Blanter Y M, Lehnert K W, Schouten R N and DiCarlo L 2013 *Nature* **502** 350–4

[12] Livingston W P, Blok M S, Flurin E, Dressel J, Jordan A N and Siddiqi I 2022 *Nat. Commun.* **13** 2307

[13] Luis A and Sánchez-Soto L L 1999 *Phys. Rev. Lett.* **83** 3573–6

[14] Fiurášek J 2001 *Phys. Rev. A* **64** 024102

[15] D'Ariano G M, Maccone L and Presti P L 2004 *Phys. Rev. Lett.* **93** 250407

[16] Lundeen J S, Feito A, Coldenstrodt-Ronge H, Pregnell K L, Silberhorn C, Ralph T C, Eisert J, Plenio M B and Walmsley I A 2009 *Nat. Phys.* **5** 27–30

[17] Chen Y, Farahzad M, Yoo S and Wei T-C 2019 *Phys. Rev. A* **100** 052315

[18] Rudinger K, Ribeill G J, Govia L C G, Ware M, Nielsen E, Young K, Ohki T A, Blume-Kohout R and Proctor T 2022 *Phys. Rev. Appl.* **17** 014014

[19] Pereira L, García-Ripoll J J and Ramos T 2022 *Phys. Rev. Lett.* **129** 010402

[20] Pereira L, García-Ripoll J J and Ramos T 2022 *Npj Quantum Inf.* **9** 22

[21] Helsen J, Roth I, Onorati E, Werner A and Eisert J 2022 *PRX Quantum* **3** 020357

[22] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R B, Jost J D, Langer C, Ozeri R, Seidelin S and Wineland D J 2008 *Phys. Rev. A* **77** 012307

[23] Magesan E, Gambetta J M and Emerson J 2011 *Phys. Rev. Lett.* **106** 180504

[24] Magesan E, Gambetta J M and Emerson J 2012 *Phys. Rev. A* **85** 042311

[25] Magesan E *et al* 2012 *Phys. Rev. Lett.* **109** 080505

[26] Gambetta J M *et al* 2012 *Phys. Rev. Lett.* **109** 240504

[27] Wallman J, Granade C, Harper R and Flammia S T 2015 *New J. Phys.* **17** 113020

[28] Wallman J J, Barnhill M and Emerson J 2016 *New J. Phys.* **18** 043021

[29] Cross A W, Magesan E, Bishop L S, Smolin J A and Gambetta J M 2016 *npj Quantum Inf.* **2** 16012

[30] Wood C J and Gambetta J M 2018 *Phys. Rev. A* **97** 032306

[31] McKay D C, Sheldon S, Smolin J A, Chow J M and Gambetta J M 2019 *Phys. Rev. Lett.* **122** 200502

[32] Helsen J, Xue X, Vandersypen L M K and Wehner S 2019 *npj Quantum Inf.* **5** 71

[33] Proctor T J, Carignan-Dugas A, Rudinger K, Nielsen E, Blume-Kohout R and Young K 2019 *Phys. Rev. Lett.* **123** 030503

[34] Erhard A, Wallman J J, Postler L, Meth M, Stricker R, Martinez E A, Schindler P, Monz T, Emerson J and Blatt R 2019 *Nat. Commun.* **10** 5347

[35] McKay D C, Cross A W, Wood C J and Gambetta J M 2020 Correlated randomized benchmarking (arXiv:2003.02354)

[36] Proctor T, Seritan S, Rudinger K, Nielsen E, Blume-Kohout R and Young K 2022 Scalable Randomized Benchmarking of Quantum Computers Using Mirror Circuits *Phys. Rev. Lett.* **129** 150502

[37] Morvan A, Ramasesh V V, Blok M S, Kreikebaum J M, O'Brien K, Chen L, Mitchell B K, Naik R K, Santiago D I and Siddiqi I 2021 *Phys. Rev. Lett.* **126** 210504

[38] Gaebler J P, Baldwin C H, Moses S A, Dreiling J M, Figgatt C, Foss-Feig M, Hayes D and Pino J M 2021 *Phys. Rev. A* **104** 062440

[39] (https://doi.org/10.5281/zenodo.6815663)

[40] Braginsky V B, Vorontsov Y I and Thorne K S 1980 *Science* **209** 547–57

[41] Gambetta J, Blais A, Schuster D I, Wallraff A, Frunzio L, Majer J, Devoret M H, Girvin S M and Schoelkopf R J 2006 *Phys. Rev. A* **74** 042318

[42] Govia L C G and Wilhelm F K 2015 *Phys. Rev. Appl.* **4** 054001

[43] Pommerening J C and DiVincenzo D P 2020 *Phys. Rev. A* **102** 032623

[44] Sank D *et al* 2016 *Phys. Rev. Lett.* **117** 190503

[45] Khezri M *et al* 2023 Measurement-induced state transitions in a superconducting qubit: Within the rotating-wave approximation *Phys. Rev. Applied* **20** 054008

[46] Rudinger K *et al* 2021 *PRX Quantum* **2** 040338

[47] 2021 Qiskit: an open-source framework for quantum computing

[48] Ball H, Stace T M, Flammia S T and Biercuk M J 2016 *Phys. Rev. A* **93** 022303

[49] Beale S J and Wallman J J 2023 Randomized compiling for subsystem measurements (arXiv:2304.06599)

[50] Ristè D, Govia L C G, Donovan B, Fallek S D, Kalfus W D, Brink M, Bronn N T and Ohki T A 2020 *npj Quantum Inf.* **6** 71

[51] Urbanek M, Nachman B and de Jong W A 2020 *Phys. Rev. A* **102** 022427

[52] Córcoles A D, Takita M, Inoue K, Lekuch S, Minev Z K, Chow J M and Gambetta J M 2021 *Phys. Rev. Lett.* **127** 100501

[53] Botelho L, Glos A, Kundu A, Miszczak J A, Salehi O and Zimborás Z 2022 *Phys. Rev. A* **105** 022441

[54] Piveteau C and Sutter D 2022 *IEEE Trans. Inf. Theory* (https://doi.org/10.1109/TIT.2023.3310797)

[55] Emerson J, Alicki R and Życzkowski K 2005 *J. Opt. B: Quantum Semiclass. Opt.* **7** S347–52

[56] Govia L C G, Ribeill G J, Ristè D, Ware M and Krovi H 2020 *Nat. Commun.* **11** 1084