Robust error accumulation suppression for quantum circuits

Tatsuki Odake,[1] Philip Taranto,[2,1] Nobuyuki Yoshioka,[3] Toshinari Itoko,[4] Kunal Sharma,[5] Antonio Mezzacapo,[5] and Mio Murao[1,*]

[1]*Department of Physics, Graduate School of Science, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan*
[2]*Department of Physics & Astronomy, University of Manchester, Manchester M13 9PL, United Kingdom*
[3]*Department of Applied Physics, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[4]*IBM Quantum, IBM Research–Tokyo, Tokyo, Japan*
[5]*IBM Quantum, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA*

(Received 8 October 2024; accepted 3 June 2025; published 8 July 2025)

We present a *robust error accumulation suppression* (REAS) technique to manage errors in quantum computers. Our method reduces the accumulation of errors in any quantum circuit composed of single- or two-qubit gates expressed as $e^{-i\sigma\theta}$ for Pauli operators σ and $\theta \in [0, \pi)$, which forms a universal gate set. For coherent errors, which include gate overrotation and crosstalk, we demonstrate a reduction of the error scaling in an L-depth circuit from $O(L)$ to $O(\sqrt{L})$. This asymptotic error suppression behavior can be proven in a regime where all gates, including those constituting the error-suppressing protocol itself, are noisy. Going beyond coherent errors, we derive the general form of decoherence noise that can be suppressed by REAS. Lastly, we experimentally demonstrate the effectiveness of our approach regarding realistic errors using 100-qubit circuits with up to 64 two-qubit gate layers on IBM Quantum processors.

DOI: 10.1103/yx9t-nctk

I. INTRODUCTION

A central problem in quantum information processing is to counteract noise arising from uncontrolled interactions between the principal system and its environment. Such noise causes errors to build up rapidly throughout quantum computations, *a priori* limiting their applicability to logarithmic-depth circuits [1]. Concurrently, the demonstration of quantum algorithms that are exponentially faster than their classical counterparts, e.g., Shor's algorithm [2,3], has stimulated enormous global efforts towards quantum computing to underpin next-generation technologies [4,5].

Developments in quantum error correction provide a promising route to scalable, fault-tolerant quantum devices [6–11]. Nonetheless, the errors must remain below the "error correction threshold" to provide any practical benefit. For moderate logical gate error rates (on the order of 10^{-8}), the errors accumulate at a rate that is insurmountable by correction procedures alone; thus, one must incorporate other methods of dealing with errors beyond correction.

To this end, two classes of complementary approaches are widely considered, namely, *error mitigation* and *error suppression*. Quantum error mitigation aims to reduce the bias in estimating expectation values of physical observables at the cost of increased variance [12–19]. Incorporating such methods has recently led to quantum simulations that are competitive with state-of-the-art classical ones [20–23]. Nevertheless, a crucial drawback of error mitigation is the exponential increase in the number of measurements required [24–26].

Quantum error suppression takes a different approach: instead of shifting the burden to measurement overheads, here one modifies the noisy circuit to design a quantum operation that approximates the target one. A paradigmatic example is the well-known dynamical decoupling protocol, which rectifies decoherence effects on idle qubits by periodically repeating fast random pulses that delete any spuriously imprinted information [27]. However, dynamical decoupling only serves to keep the state of the system from decohering, and it is *a priori* unclear how to extend it to situations where noisy gates are being applied to a physical system for the purpose of a computation.

Fortunately, in such scenarios, coherent errors in circuits composed of Clifford gates, e.g., accidental gate overrotation, can be suppressed via related techniques that beneficially harness randomness, e.g., probabilistic compiling [28–30] and Pauli twirling [31–33]. More importantly, the procedure of randomized compiling applies beyond such limitations and can even suppress the growth of errors in universal (i.e., non-Clifford) circuits [34–36]. Nonetheless, the class of noise that can be counteracted via such methods remains somewhat restricted and may be surpassed in realistic near-term settings, such as partially fault-tolerant architectures with analog rotation gates [37]. Second, it is not clear that known techniques

*Contact author: murao@phys.s.u-tokyo.ac.jp

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

are robust against errors in their own application (as opposed to those occurring in the computation gates). Lastly, the extent to which error suppression techniques can counteract decoherence gate-level errors (while keeping the desired gate applied) is yet to be fully determined in general, although experimental evidence suggests that such errors can be somewhat dealt with [22,23,38–42].

Here, to overcome these problems, we present an advanced procedure for managing errors in quantum computers: *robust error accumulation suppression* (REAS). Our method reduces the accumulation of errors in any circuit composed of single- or two-qubit gates expressed as $e^{-i\sigma\theta}$ for Pauli operators σ and $\theta \in [0, \pi)$, which forms a universal gate set. Improving upon standard error suppression methods, our technique reduces the accumulation of first-order error terms while keeping the variance unchanged, leveraging a nested and correlated structure of the inserted Pauli gates. In the case of coherent errors, which includes spectator crosstalk [43], we demonstrate a reduction of the error scaling in an L-depth circuit from $O(L)$ to $O(\sqrt{L})$. Crucially, REAS makes no assumption on the cleanness of the error-suppressing protocol itself and is, therefore, truly robust, applying to situations in which the newly inserted gates themselves are noisy. Furthermore, we detail the class of decoherence noise that can be suppressed via REAS, namely, that which is well approximated by weak system-environment coupling, and demonstrate numerical evidence for the quadratic suppression of error scaling.

This paper is outlined as follows. In Sec. II, we formalize our framework and analyze a motivating example. In Sec. III, we present our method for robustly suppressing the accumulation of errors and prove our main theoretical results. To highlight the applicability and effectiveness of our approach, we supplement this analysis with numerical simulations in Sec. IV, as well as provide an experimental demonstration using IBM Quantum processors in Sec. V.

II. ROBUST ERROR ACCUMULATION SUPPRESSION

A. Layered circuits

We begin by presenting the framework that describes circuit-based quantum computation with errors. Suppose that one wishes to implement a layered circuit consisting of single- or two-qubit gates of the form $e^{-i\sigma\theta}$ for rotation angle $\theta \in [0, \pi)$ (which includes all gates of the form $e^{-i\sigma\theta'}$ for $\theta' \in \mathbb{R}$ up to a global phase). Consider such a circuit of L layers of gates applied to an n-qubit system; in each independent layer ℓ, the particular qubits that are acted on nontrivially are denoted by Q_ℓ, which consists of disjoint subsets of $\{1, \ldots, n\}$. Grouping the layers into B blocks (which can now comprise sequentially implemented gates on certain qubits) leads to

$$\prod_{b=1}^{B} \prod_{\ell_b} \left(\bigotimes_{\vec{q} \in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\,\theta_{\ell_b,\vec{q}}} \right), \tag{1}$$

where ℓ_b labels the layer *within* each block b. For a block with L_B layers, the total number of layers is $BL_B = L$; throughout, we consider $B \propto L$, so the scaling behavior in terms of layers or blocks is identical up to constant factors. We refer to the Pauli operators $\sigma_{\ell_b,\vec{q}}$ inside the exponential in the circuit as "computation gates" to distinguish them from other types of Pauli gates that will be inserted at will in order to suppress potential errors in the computation, which we will term "inserted gates" (see Fig. 1).

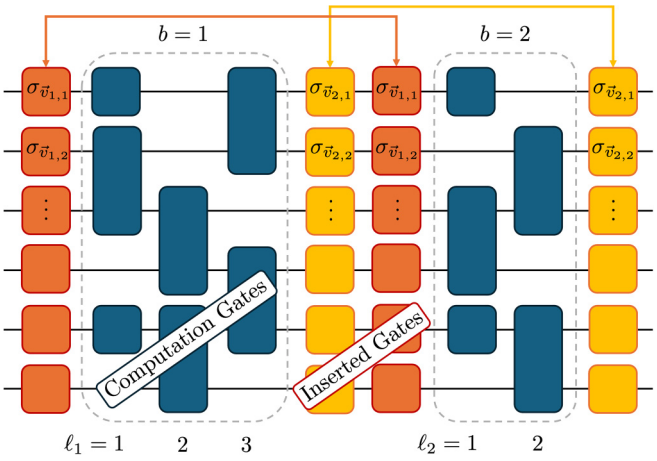


FIG. 1. REAS schematic. An $n = 6$ qubit quantum circuit, consisting of $B = 2$ blocks (gray dashed outline) of computation gates (blue), with the first block comprising three layers and the second comprising two layers. Each such computation gate acts on at most two qubits and is of the form $e^{-i\sigma\theta}$ for $\theta \in [0, \pi)$. Between each block b, a random single-qubit Pauli gate $e^{-i\sigma_{\vec{v}_{b,q}}}$ is applied to qubit q (orange, yellow). The nested, correlated structure of these inserted gates (depicted by the arrows at the top) leads to a quadratic suppression of the accumulation of errors (in terms of circuit length). This error suppression works even in the case where the inserted gates themselves are noisy since every gate (including the inserted random Pauli gates) is itself sandwiched by correlated random Pauli gates (e.g., the leftmost yellow Pauli gates are sandwiched by the orange Pauli gates).

B. Motivating example

To motivate the problem at hand, we now analyze the impact of (small) coherent gate errors on block-layered quantum circuits of the form in Eq. (1). Intuitively, the actual operation applied to each layer is shifted slightly with respect to the desired one by some additional unitary rotation, leading to the implementation of

$$\left(\bigotimes_{\vec{q} \in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\,\theta_{\ell_b,\vec{q}}} \right) \left[I + i\epsilon C_{\ell_b} \right], \tag{2}$$

where $\|C_{\ell_b}\|_{\text{op}} \leqslant 1$ for all b and ℓ_b, and $\epsilon > 0$ is small such that $I + i\epsilon C_{\ell_b}$ is close to the identity, encoding the assumption that the error in each layer is not too large. In any such layered circuit, there are $L = \sum_b \ell_b$ layers where $O(\epsilon)$ errors can occur, and thus the total error can be $O(L\epsilon)$ in the worst case.

Inspired by the method of Pauli twirling, we now consider the effect of inserting random Pauli operators on all qubits between each block. Let $\sigma_{\vec{v}_b} := \bigotimes_{k=1}^{n} \sigma_{v_b}^{(k)}$ for $v_b \in \{0, 1, 2, 3\}$ be a uniformly randomly sampled n-qubit Pauli operator applied before block b. Consider an arbitrary block b preceded by some string of inserted Pauli gates $\sigma_{\vec{v}_b}$. Now, for tensor products of (inserted) qubit Pauli gates, any such overall

operation either commutes or anticommutes with any (one- or two-qubit) Pauli gates $e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}$ applied in the subsequent block. This permits us to define

$$s := \begin{cases} 1 & \text{if } \sigma_{\ell_b,\vec{q}} \text{ and } \sigma_{\vec{v}_b} \text{ commute} \\ -1 & \text{if } \sigma_{\ell_b,\vec{q}} \text{ and } \sigma_{\vec{v}_b} \text{ anti-commute.} \end{cases} \tag{3}$$

Although we do not explicitly label it for ease of notation, s inherently depends upon ℓ_b, $\vec{q}$, and $\vec{v}_b$: each computational gate $\sigma_{\ell_b,\vec{q}}$ in the bock may have a different sign factor s depending on its commutation relation with the specific inserted Pauli string $\sigma_{\vec{v}_b}$ (so strictly speaking, $s = s_{\ell_b,\vec{q},\vec{v}_b}$).

With this, we can express the circuit transformation between blocks succinctly as follows. When a random Pauli string $\sigma_{\vec{v}_b}$ is inserted before block b, the computational gates within the block are sandwiched by said Pauli string:

$$\prod_{\ell_b} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}} \mapsto \sigma_{\vec{v}_b}\left(\prod_{\ell_b} e^{-is\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}\right)\sigma_{\vec{v}_b}. \tag{4}$$

Here, the arrow $\mapsto$ indicates that we rewrite the original computational circuit into this equivalent form by appropriately conjugating with the inserted Pauli gates. The factor $s = \pm 1$ accounts for whether the inserted Pauli string $\sigma_{\vec{v}_b}$ commutes $(+1)$ or anticommutes (-1) with its neighboring computational Pauli gates $\sigma_{\ell_b,\vec{q}}$. Note that the technique of transforming computational Pauli gates to the form of Eq. (4) itself is not new; however, in previous works, the error on the inserted Pauli gate $\sigma_{\vec{v}_b}$, as well as the possible dependence of such errors on the rotation direction through s, was not explicitly considered. Here, we demonstrate that such errors can also be dealt with, highlighting that our method is robust against coherent noise *in its own implementation*.

Consider now for the sake of simplicity single-layer blocks, in which case we simply write $\sigma_{\ell_b,\vec{q}} =: \sigma$ and $\theta_{\ell_b,\vec{q}} =: \theta$ for clarity. In the absence of any errors, both the original gate $e^{-i\sigma\theta}$ and the average transformed one $(1/4)^n \sum_{\vec{v}_b} \sigma_{\vec{v}_b} e^{-is\sigma\theta}\sigma_{\vec{v}_b}$ lead to the same overall gate applied. However, whenever the computation gates $e^{is\sigma\theta}$ are affected by a coherent error (i.e., corresponding to a unitary rotation of the form $I + i\epsilon C_s$), the original gate suffers from an $O(\epsilon)$ error, whereas the transformed gate undergoes the change

$$\begin{aligned} &\left(\frac{1}{4}\right)^n \sum_{\vec{v}_b} \sigma_{\vec{v}_b}[e^{-is\sigma\theta}(I + i\epsilon C_s)]\sigma_{\vec{v}_b} \\ &= \left(\frac{1}{4}\right)^n \sum_{\vec{v}_b} e^{-i\sigma\theta}\left[I + i\epsilon\sigma_{\vec{v}_b}C_s\sigma_{\vec{v}_b}\right] \\ &= e^{-i\sigma\theta}[I + i\epsilon(\alpha I + \beta\sigma)] = e^{i\epsilon\alpha}e^{-i(\theta-\epsilon\beta)\sigma} + O(\epsilon^2), \end{aligned} \tag{5}$$

where α and β are the coefficients of the term in C_s proportional to I and $s\sigma$, respectively. This can be shown from

$$\left(\frac{1}{4^n}\right) \sum_{\vec{v}_b \in \{0,1,2,3\}^n} \sigma_{\vec{v}_b}\sigma\sigma_{\vec{v}_b} = \begin{cases} I & \text{for } \sigma = I, \\ 0 & \text{for } \sigma \neq I \end{cases} \tag{6}$$

and

$$\left(\frac{1}{4^n}\right) \sum_{\vec{v}_b \in \{0,1,2,3\}^n} \sigma_{\vec{v}_b} s\sigma'\sigma_{\vec{v}_b} = \begin{cases} \sigma & \text{for } \sigma = \sigma', \\ 0 & \text{for } \sigma \neq \sigma'. \end{cases} \tag{7}$$

Since $e^{i\epsilon\alpha}$ in Eq. (5) only contributes to a global phase and can thus be ignored, the additional $O(\epsilon)$ shift $-\epsilon\beta$ in the rotation angle is the only error that must be dealt with in order to erase all $O(\epsilon)$ error within each layer.

The general (multilayer block) expression for how the overall layered circuit with coherent errors transforms on average upon inserting random Pauli gates between blocks is

$$\begin{aligned} &\left(\frac{1}{4^n}\right)^B \sum_{\vec{v}_1,\ldots,\vec{v}_B} \left[\sigma_{\vec{v}_B} \prod_{\ell_B} \left(\bigotimes_{\vec{q}\in Q_{\ell_B}} e^{-is\,\sigma_{\ell_B,\vec{q}}\,\theta_{\ell_B,\vec{q}}}\right)\right] \\ &\quad \times \prod_{b=1}^{B-1} \left[\sigma_{\vec{v}_{b+1:b}} \prod_{\ell_b} \left(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-is\,\sigma_{\ell_b,\vec{q}}\,\theta_{\ell_b,\vec{q}}}\right)\right]\sigma_{\vec{v}_1} \\ &= \prod_{b=1}^{B}\prod_{\ell_b} \left(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\,(\theta_{\ell_b,\vec{q}}+\epsilon\beta_{\ell_b,\vec{q}})}\right) + O(\epsilon + B\epsilon^2). \end{aligned} \tag{8}$$

Here, we have introduced the notation $\vec{v}_{j:k}$ for two vectors that satisfy $\sigma_{\vec{v}_{j:k}} \propto \sigma_{\vec{v}_k}\sigma_{\vec{v}_j}$ and implicitly understand the value of s to depend upon whether or not each inserted Pauli gate string commutes or anticommutes with each layer of computational gates in the neighboring block. We refer to the random Pauli operators $\sigma_{\vec{v}_1}, \sigma_{\vec{v}_{2:1}} \ldots, \sigma_{\vec{v}_B}$ as the "inserted gates" to distinguish them from the computational Pauli gates. Note the *nested and correlated* structure of the inserted random Pauli gates: each inserted gate $\sigma_{\vec{v}_j}$ is sandwiched by either $\sigma_{\vec{v}_{j-1}}$ or $\sigma_{\vec{v}_{j+1}}$ for $1 < j < B$, which is what provides our scheme with robustness against noise in the inserted gates themselves.

In the following, we will demonstrate a method that cancels the $O(\epsilon)$ errors caused by the shift in rotation angle. This leaves an overall error that scales as $O(\epsilon + B\epsilon^2)$, which approaches $O(B\epsilon^2)$ for circuits with high gate depth. Importantly, our procedure also works when the inserted operations themselves are noisy (which we have not explicitly included above but will below), thereby demonstrating robustness of the method.

III. MAIN RESULTS

A. Robust error accumulation suppression

We now present our method of *robust error accumulation suppression* (REAS). This procedure comprises of first performing a *calibration* process by estimating the rotation error in each layer, and then *inserting random Pauli gates* in order to counteract this in a tailored manner, removing all $O(\epsilon)$ errors as we will show in Sec. III B. An optional third step involves transforming all single-qubit computational gates into two-qubit ones in such a way that certain types of decoherence errors can also be suppressed, as we will discuss in Sec. III C. Details are provided in Supplemental Material (SM) Sec. S1 [44].

Calibration. The calibration step is performed by considering each layer of the circuit individually. For each chosen layer ℓ_b, one repeats the computational gates of that layer K times and intersperses it with random Pauli gate strings, which

leads to the transformation[1]

$$\left(\frac{1}{4^n}\right)^K \sum_{\vec{v}_1,\ldots,\vec{v}_K} \sigma_{\vec{v}_K}\left(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-is\,\sigma_{\ell_b,\vec{q}}\,\theta_{\ell_b,\vec{q}}}\right) \times \prod_{k=1}^{K}\left[\sigma_{\vec{v}_{k+1:k}}\left(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-is\,\sigma_{\ell_b,\vec{q}}\,\theta_{\ell_b,\vec{q}}}\right)\right]\sigma_{\vec{v}_1}. \tag{9}$$

Since the undesired shift in the rotation angle accumulates by repeating the same layer K times, the value of $\epsilon\beta_{\ell_b,\vec{q}}$ is a systematic (i.e., additive) error and can therefore be estimated via robust phase estimation [45].

This technique uses a robust variant of the well-known phase estimation procedure [46] for determining such systematic errors in a universal single-qubit gate set to achieve optimal efficiency, i.e., Heisenberg scaling, without requiring entanglement or auxiliary systems. Phase estimation is a metrological technique that works by performing sequences of measurements to obtain observables from which systematic errors such as gate overrotation can be extracted, thereby providing a foundation for gate calibration. The power of this method comes from the fact that small errors in gates coherently accumulate into large observables when applied in sequence many times, allowing them to be efficiently estimated. In the standard version, state preparations and measurements must be performed perfectly, and gates must be of an assumed form. However, the robust variant relaxes these conditions and permits imperfections; nonetheless, Heisenberg scaling is still achievable. See SM Sec. S2 [44].

In order to obtain an estimate with at most $O(\epsilon^2)$ error, K should be taken to be $O(1/\epsilon^2)$. Note finally that the error term of $O(\epsilon + B\epsilon^2)$ in Eq. (8) can still be small enough to effectively perform robust phase estimation by taking B proportional to $1/\epsilon^2$.

Inserting random Pauli gates. Using the value of the rotation error shifts obtained during the calibration procedure, one can then insert random Pauli gates into the calibrated layered circuit to specifically suppress the errors in each block (see Fig. 1). This is achieved by shifting the rotation angle of all computational Pauli gates in the original circuit by angles of the same magnitude as that estimated in the calibration but of the opposite sign to cancel out the error.

It is straightforward to see that our method works to suppress errors in non-Clifford circuits since the values of the rotation angles above are not necessarily from a discrete set. Moreover, as we will demonstrate in the coming section, additional coherent errors on the inserted Pauli gates can also be dealt with, thereby endowing our method with robustness that is crucial in realistic settings.

B. Suppressing the accumulation of coherent errors

The REAS scheme suppresses any coherent error, both on the computational Pauli gates and on the inserted ones such that the error on the overall circuit $\prod_{b=1}^{B}\prod_{\ell_b}(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}})$ decreases from $O(\epsilon B)$ to $O(\epsilon+\epsilon^2 B)$ (assuming that the maximum size of any block is upper bounded by a constant, as we will do throughout).

Up to now, we have analyzed REAS in the setting where one averages over the inserted random Pauli gates. Although it might seem at first glance that the average case being close to the ideal one is not necessarily related to the error on any individual gate, in SM Sec. S3 [44], we show that the former condition implies that the *root mean square* (RMS) of the latter is small, thereby bounding "worst case" scenarios. In particular, we show that the RMS error (using the diamond distance) between the error-free unitary implemented by $\prod_{b=1}^{B}\prod_{\ell_b}(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}})$ and any single-shot unitary operator implemented after calibration in a given run with the random Pauli gate string inserted scales as $O(\sqrt{\epsilon+\epsilon^2 B})$, which approaches $O(\epsilon\sqrt{B})$ in the regime $B \gg 1/\epsilon$.

Now, we briefly explain how the REAS suppresses any coherent error occurring on the implementation of the inserted random Pauli operators themselves. Suppose that a single error term $i\epsilon C_{\vec{v}_{b'+1;b'}}$ occurs on a Pauli operator $\sigma_{\vec{v}_{b'+1;b'}}$ of

$$\left(\frac{1}{4^n}\right)^B \sum_{\vec{v}_1,\ldots,\vec{v}_B}\left[\sigma_{\vec{v}_B}\prod_{\ell_B}\left(\bigotimes_{\vec{q}\in Q_{\ell_B}} e^{-is\,\sigma_{\ell_B,\vec{q}}\,\theta_{\ell_B,\vec{q}}}\right)\right] \times \prod_{b=1}^{B-1}\left[\sigma_{\vec{v}_{b+1:b}}\prod_{\ell_b}\left(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-is\,\sigma_{\ell_b,\vec{q}}\,\theta_{\ell_b,\vec{q}}}\right)\right]\sigma_{\vec{v}_1}, \tag{10}$$

so that the overall gate sequence implemented is

$$\left(\frac{1}{4^n}\right)^B \sum_{\vec{v}_1,\ldots,\vec{v}_B}\sigma_{\vec{v}_B}[\cdots]\sigma_{\vec{v}_{b'+2;b'+1}}\prod_{\ell_{b'+1}}\left(\bigotimes_{\vec{q}\in Q_{\ell_{b'+1}}} e^{-is\sigma_{\ell_{b'+1},\vec{q}}\theta_{\ell_{b'+1},\vec{q}}}\right) \times \left(\sigma_{\vec{v}_{b'+1;b'}} i\epsilon C_{\vec{v}_{b'+1;b'}}\right)\prod_{\ell_{b'}}\left(\bigotimes_{\vec{q}\in Q_{\ell_{b'}}} e^{-is\sigma_{\ell_{b'},\vec{q}}\theta_{\ell_{b'},\vec{q}}}\right) \times \sigma_{\vec{v}_{b';b'-1}}[\cdots]\sigma_{\vec{v}_1}. \tag{11}$$

This expression can be rewritten as

$$\left(\frac{1}{4^n}\right)^2 \sum_{\vec{v}_{b'},\vec{v}_{b'+1}}\prod_{b=b'+1}^{B}\prod_{\ell_b}\left(\bigotimes_{\vec{q}\in Q_\ell} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}\right) \times \sigma_{\vec{v}_{b'}} i\epsilon C_{\vec{v}_{b';b'+1}}\sigma_{\vec{v}_{b'}}\prod_{b=1}^{b'}\prod_{\ell_b}\left(\bigotimes_{\vec{q}\in Q_\ell} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}\right) = [\cdots]\left(\frac{1}{4^n}\right)\sum_{\vec{v}_{b'}}\sigma_{\vec{v}_{b'}}\left(\frac{1}{4^n}\sum_{\vec{v}_{b'+1}} i\epsilon C_{\vec{v}_{b';b'+1}}\right) \times \sigma_{\vec{v}_{b'}}[\cdots] = O(\epsilon^2), \tag{12}$$

where $[\cdots]$ represents the terms that are irrelevant. As shown in SM Sec. S1 [44], this overall expression is $O(\epsilon^2)$ since

[1]Note that the value of s in each term here differs, as it depends explicitly upon the commutation relation with the Pauli gates inserted at its neighboring position.

$\sum_{\vec{v}_{b'+1}} C_{\vec{v}_{b':b'+1}}$ is independent of $\vec{v}_{b'}$ and the term in $i\epsilon C_{\vec{v}_{b':b'+1}}$ proportional to the identity can be taken to be $O(\epsilon^2)$ without loss of generality.

Lastly, in SM Sec. S4 [44], we analyze the relation between REAS and *randomized compiling* (RC) [34]. Although both methods make use of inserting random gates to suppress errors, they do so in different ways, leading to distinct scaling behaviors. In REAS, the inserted gates depend upon their commutation relation with their neighboring gates, in contrast to those in RC which automatically cancel coherent errors by averaging over Pauli operators. Thus, the REAS approach requires additional necessity to deal with s_k-dependent errors on gates of the form $e^{\pm i\sigma_k\theta_k}$; the aforementioned calibration and single Pauli transformation subroutines can deal with such errors (but are based upon a particular model of decoherence error; see below and SM Sec. S5 [44]). Of course, such dependence, together with the fact that REAS can increase the depth of the circuit (albeit only to at most twice as long), can be thought of as a disadvantage of REAS compared to RC. On the other hand, these differences makes REAS more robust against gate-dependent errors on the twirling gates.

C. Suppressing the accumulation of decoherence errors (with single Pauli transformation)

The original procedure of REAS comprising calibration and then inserting random Pauli gates cannot immediately suppress general error due to decoherence. However, we now incorporate an additional step, which we dub "single Pauli transformation," that allows our method to deal with certain classes of decoherence errors, in particular those modeled by weak system-environment interactions. See SM Sec. S5 for details [44].

Single Pauli transformation. Here we modify all single-qubit computation gates $e^{-is\sigma_{l_b,\vec{q}}\theta_{l_b,\vec{q}}}$ (i.e., those except for the inserted random Pauli operators) as follows:

$$\begin{aligned} e^{isYIt} &\to e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ},\\ e^{isZIt} &\to e^{-i\frac{\pi}{4}XI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{i\frac{\pi}{4}XI},\\ e^{isXIt} &\to e^{i\frac{\pi}{4}ZI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{-i\frac{\pi}{4}ZI}. \end{aligned} \tag{13}$$

This transformation makes the single-qubit Pauli rotation gate robust against decoherence errors, thereby enabling the tolerance of the REAS procedure against such noise. The trading of such decoherence errors for gate errors comes at the cost of introducing longer sequences of two-qubit gates; however, since these sequences increased by a constant factor, in certain situations, the scaling behavior will become favorable. As we will show below, this mapping leads to the suppression of decoherence errors that are up to first order in the operator norm of the interaction Hamiltonian.

Upon invoking the single Pauli transformation, the error-suppressing properties of REAS against decoherence errors become similar to those explained previously. In particular, this procedure suppresses decoherence error on both the computational gates and the inserted random Pauli operators in such a way that the error on the overall circuit $\prod_{\ell_b=1}^{B}(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}})$ changes from $O(\epsilon B)$ to $O(\epsilon+\epsilon^2 B)$ in the average case; similarly, the RMS of the diamond distance between the error-free unitary operation implemented by $\prod_{\ell_b=1}^{B}(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}})$ and any single-shot unitary operator implemented after calibration in a given run with random Pauli gates inserted scales as $O(\sqrt{\epsilon+\epsilon^2 B})$.

We now move on to analyze the class of decoherence errors that can be dealt with via REAS. We model the decoherence error via introducing an interaction Hamiltonian $H'_{\ell_b,\vec{q},s_{\vec{q}}}(t)\in\mathcal{H}\otimes\mathcal{H}_{\text{env}}$ of order ϵ that leads to the pulse Hamiltonian that implements the desired $e^{-is_{\vec{q}}\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}$. Note that, since $H'_{\ell_b,\vec{q},s_{\vec{q}}}(t)$ can contain arbitrary terms that lead to both coherent and incoherent errors, our statements here are a direct generalization of those made in the previous section regarding coherent errors. Note, however, that this approach requires engineered interactions with auxiliary qubits in order to suppress the decoherence error; in other words, the decoherence process must be engineered or at least accurately modeled rather than stemming from truly spurious noise.

We begin by assuming that the decoherence noise applied to the layer $(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-is_{\vec{q}}\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}})$ can be expressed as

$$\mathcal{T}\exp\left[-i\int_0^T \mathrm{d}t'\left(\sum_{\vec{q}\in Q_{\ell_b}} H_{\ell_b,\vec{q},s_{\vec{q}}}(t')\right)+H_{\text{env}} + \sum_{\vec{q}\in Q_{\ell_b}} H'_{\ell_b,\vec{q},s_{\vec{q}}}(t')\right], \tag{14}$$

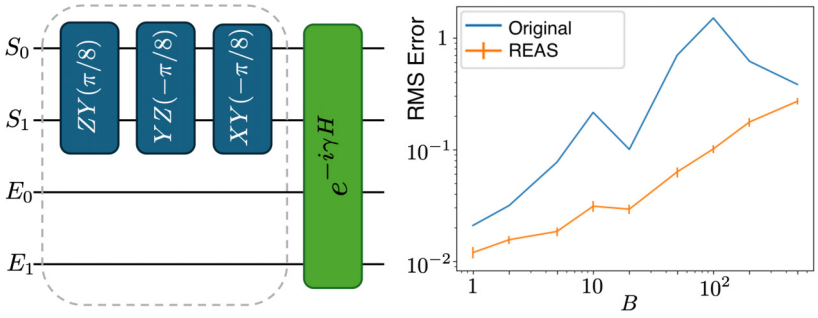
where $H_{\ell_b,\vec{q},s_{\vec{q}}}(t')\in\mathcal{H}$ is the time-dependent pulse Hamiltonian to implement the desired gate $e^{-is_{\vec{q}}\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}$, $H'_{\ell_b,\vec{q},s_{\vec{q}}}(t')\in\mathcal{H}\otimes\mathcal{H}_{\text{env}}$ is the undesired Hamiltonian that is induced when applying $H_{\ell_b,\vec{q},s_{\vec{q}}}$ in the presence of noise, $H_{\text{env}}\in\mathcal{H}_{\text{env}}$ is the background Hamiltonian of the -environment, and $\mathcal{T}$ is the time-ordering operator. Again, identity operators on subspaces where the dynamics acts trivially are implied. We also assume that the undesired Hamiltonian $H'_{\ell_b,\vec{q},s_{\vec{q}}}(t')$ does not contain interaction terms involving more than two systems (either qubits in $\mathcal{H}$ or systems in $\mathcal{H}_{\text{env}}$); since most Hamiltonians in nature involve at most two-body interactions, this assumption is natural.

Note that many relevant decoherence mechanisms, such as depolarizing, amplitude damping, and dephasing, can be expressed by this interaction model. Indeed, for single-qubit decoherence channels

$$\begin{aligned} \mathcal{E}_p^{\text{dpl}}(\rho) &:= (1-p)\rho + p\frac{I}{2},\\ \mathcal{E}_\gamma^{\text{ad}}(\rho) &:= E_0\rho E_0^\dagger + E_1\rho E_1^\dagger,\\ \mathcal{E}_p^{\text{dph}}(\rho) &:= (1-p)\rho + pZ\rho Z, \end{aligned} \tag{15}$$

where

$$E_0 := \begin{pmatrix}1 & 0\\ 0 & \sqrt{\gamma}\end{pmatrix} \quad\text{and}\quad E_1 := \begin{pmatrix}0 & \sqrt{\gamma}\\ 0 & 0\end{pmatrix}, \tag{16}$$



(a) (b)

	$ZY(\pi/8)$	$YZ(-\pi/8)$	$XY(-\pi/8)$
Theory	−1.99	1.58	2.05
REAS	−1.95	1.57	1.97
Without REAS	−2.88	2.20	2.00

(c)

FIG. 2. (a) Deep circuit. We consider a deep circuit composed of repeated blocks of computational gates (blue) $ZY(\pi/8)$, $YZ(-\pi/8)$, $XY(-\pi/8)$, where, e.g., $ZY(\pi/8)$ corresponds to $e^{-i\frac{\pi}{8}ZY}\otimes I_{\rm env}$ and similar for the others. Following this, decoherence noise is simulated by applying $e^{-i\gamma H}$ (green) on all qubits (including the environment). (b) REAS performance in a deep circuit. The blocks are iterated a number B times, leading to overall circuits of growing depth, and we measure the RMS error with respect to the expectation value of $IZII$ [see Eq. (19)]. When REAS is applied (orange), the RMS error scales like $B^{0.501}$, which is close to the theoretical upper bound on the scaling of $B^{1/2}$. (c) Shift in rotation angle (10^{-3} rad). Here we calculate the shift in rotation angle $\Delta(\theta_i)$, i.e., that induced by the coherent noise, in theory as well as with REAS and with conventional calibration techniques, highlighting the performance of the former.

their action can be modeled explicitly as

$$\mathcal{E}^{\rm dpl}_{\sin^2(2\delta t)}(\rho)=\mathrm{tr}_{\rm env}\left[e^{-i\delta t(X\otimes X_{\rm env}+Y\otimes Y_{\rm env}+Z\otimes Z_{\rm env})}\left(\rho\otimes\frac{I_{\rm env}}{2}\right)e^{i\delta t(X\otimes X_{\rm env}+Y\otimes Y_{\rm env}+Z\otimes Z_{\rm env})}\right],$$
$$\mathcal{E}^{\rm ad}_{\sin^2(\delta t)}(\rho)=\mathrm{tr}_{\rm env}\left[e^{-i\frac{\delta t}{2}(X\otimes X_{\rm env}+Y\otimes Y_{\rm env})}(\rho\otimes|0\rangle\langle 0|_{\rm env})e^{i\frac{\delta t}{2}(X\otimes X_{\rm env}+Y\otimes Y_{\rm env})}\right],$$
$$\mathcal{E}^{\rm dph}_{\sin^2(\delta t)}(\rho)=\mathrm{tr}_{\rm env}\left[e^{-i\delta t(Z\otimes Z_{\rm env})}\left(\rho\otimes\frac{I_{\rm env}}{2}\right)e^{i\delta t(Z\otimes Z_{\rm env})}\right]. \tag{17}$$

As is the case regarding dynamical decoupling, note that different ways of modeling decoherence in this interaction picture can lead to varying performance of REAS. For instance, in a similar discussion to that of Ref. [47], any decoherence error that is modeled by quantum channels on the system without explicitly referring to its environment cannot be erased by REAS (as can be seen by considering the case where the depolarizing channel acts on all gates).

IV. NUMERICAL RESULTS

We now present numerical simulations that showcase the efficacy of our methods. We consider the setting where $\mathcal{H}$ (qubits: S_0, S_1) and $\mathcal{H}_{\rm env}$ (qubits: E_0, E_1) are two-qubit systems (see Fig. 2). The initial (global) state of the circuit is $|0\rangle^{\otimes 4}$ in all simulations. All computational gates and inserted Pauli gates are affected by decoherence noise simulated by applying the following global unitary after each gate $e^{-i\gamma H}\in\mathcal{H}\otimes\mathcal{H}_{\rm env}$, where H is normalized as $\|H\|_2=1$ and the parameter γ specifies the scale of noise. The value of γ is common across all gates, but the operator H consisting of single-qubit Pauli terms on all systems and two-qubit Pauli terms on S_0E_0 and S_1E_1 that are randomly chosen for each gate (unless specified otherwise).

In Sec. IV A, we compare our REAS calibration method to the theoretical prediction for a deep circuit, as well as with conventional calibration techniques (i.e., performing phase estimation without inserting random Pauli gates). In Secs. IV B and IV C, we choose the noise Hamiltonians such that the first-order γ term in the shift in rotation angle vanishes to avoid the calibration step, which required too long computation time to give precise results.

A. Effectiveness of REAS in deep quantum circuits

Here, we demonstrate the effectiveness of our REAS scheme in a deep quantum circuit. We consider a circuit block of depth 3 in Fig. 2(a) containing only two-qubit Pauli computation gates. The string $ZY(\pi/8)$ indicates the Pauli rotation $e^{-i\frac{\pi}{8}ZY}\otimes I_{\rm env}$; other strings have a similar meaning. In Fig. 2(b), we verify numerically that the theoretical prediction of the RMS between the ideal circuit and a single-shot unitary operator implemented after calibration in a given run of the circuit with random Pauli gates inserted scales as $O(\sqrt{B})$. (Since each block contains three gates, $B=3L$.)

We first perform the REAS calibration to obtain an estimate of the shift in the rotation angle of each gate in $ZY(\pi/8)$, $YZ(-\pi/8)$, and $XY(-\pi/8)$. Throughout the simulation, we fix the noise parameter $\gamma=0.01$ and introduce a bias in choosing H of the coherent noise $e^{-i\gamma H}$; this is done so that the terms proportional to the appropriate computational Pauli operators are likely to have a large coefficient, in order to clearly demonstrate the effect of calibration. The coherent noise $e^{-i\gamma H}$ invokes modification of rotation angles for $ZY(\theta_1)$, $YZ(\theta_2)$, and $XY(\theta_3)$, i.e., induces the shift $(\theta_1,\theta_2,\theta_3)\to(\theta_1+\Delta\theta_1,\theta_2+\Delta\theta_2,\theta_3+\Delta\theta_3)$.

We then apply the inverse shift of the rotation angles applied according to the estimates obtained in the

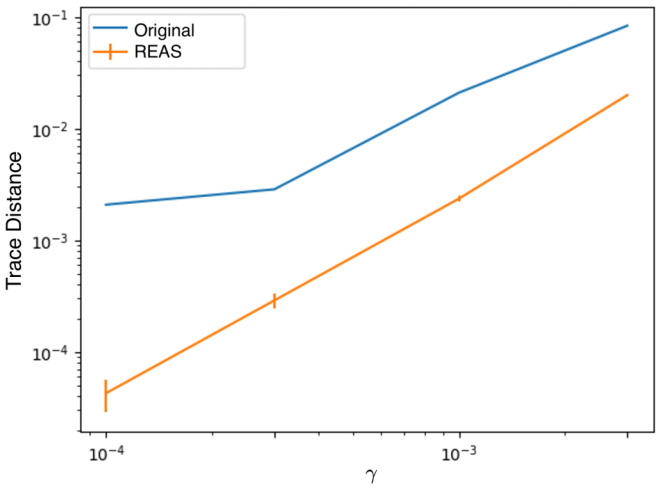


FIG. 3. Suppression of first-order error. Decoherence noise is simulated by applying $e^{-i\gamma H} \in \mathcal{H} \otimes \mathcal{H}_{\text{env}}$ after each gate, where H is normalized as $\|H\|_2 = 1$ and γ specifies the scale of noise [see Fig. 2(a)]. In the limit $\gamma \to 0$, the first-order error term vanishes when REAS is applied (orange, REAS), in contrast to the case where it is not (blue, original), as predicted by theory.

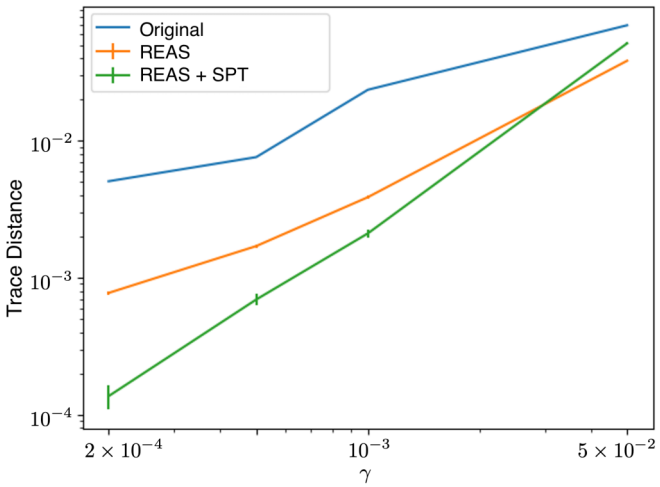


FIG. 4. Effectiveness of single Pauli transformation. We compare the suppression of decoherence errors in three situations: first, without REAS (blue, original); second, with REAS but without incorporating the single Pauli transformation procedure (orange, REAS); and third, with REAS and single Pauli transformation (green, REAS + SPT). The behavior improves accordingly in the limit $\gamma \to 0$ (although note that for larger errors, no strict hierarchy applies, as is evident by the crossing of the orange and green lines).

REAS calibration, before repeating the circuit block $B \in \{1, 2, 5, 10, 20, 50, 100, 200, 500\}$ times, and measuring the expectation value of the observable $IZII$ (i.e., Z on S_1). Finally, we compute the RMS (sampling number: 50) between the ideal expectation value and that obtained in the noisy circuit, both with and without REAS. Figure 2(b) shows that the scaling of the RMS of the difference in the expectation value is nearly of order $B^{1/2}$. Although the theory only predicts that the RMS is upper bounded by a function scaling like $B^{1/2}$, which does not imply that a change in the value of RMS always follows this scaling (e.g., oscillations of the RMS value are possible), the value of the power calculated here is approximately 0.501, corresponding closely to the upper bound. In contrast, the corresponding scaling for the case without REAS grows faster (approximately linearly) until it reaches the saturation point. The table in Fig. 2(c) shows that the value of rotation shift $(\theta_1, \theta_2, \theta_3) \to (\theta_1 + \Delta\theta_1, \theta_2 + \Delta\theta_2, \theta_3 + \Delta\theta_3)$ caused by the coherent noise $e^{-i\gamma H}$ is precisely estimated via REAS calibration with error less than $\times 10^{-4}$ rad. In contrast, the estimate of this rotation shift deviates further from the theoretical prediction when using calibration without inserting random Pauli gates. This result implies that the standard calibration technique based on the phase estimation without inserting random Pauli gates cannot precisely remove the effect of rotation shift and thus the full REAS calibration is required.

B. Suppression of the first-order error

In Fig. 3, we demonstrate that the first-order term of the error vanishes in the regime where γ is near zero (i.e., the weak-coupling regime), as predicted by theory. To do so, we randomly generate a circuit with a depth of 1000 that only contains two-qubit Pauli rotations as the computational gates. For simplicity, we assume that the rotation angles in these random circuits take values in $\pm\frac{\pi}{2}, \pm\frac{\pi}{4}$, and $\pm\frac{\pi}{8}$. For the noise parameter $\gamma \in \{0.0001, 0.0003, 0.001, 0.003\}$, we compute the average trace distance between the state obtained after running the error-free circuit and that obtained after running a noisy circuit both with and without the REAS procedure applied (sampling number: 4000). The simulation in Fig. 3 shows a near quadratic dependence of the trace distance on γ when using REAS, whereas the scaling for the case without REAS is near linear for small γ.

C. Effectiveness of the single Pauli transformation

In Fig. 4, we demonstrate how incorporating the single Pauli transformation helps to reduce decoherence noise. We randomly generate a circuit with a depth of 1000 that contains both single-qubit and two-qubit Pauli gates. For $\gamma \in \{0.0002, 0.0005, 0.001, 0.005\}$, we then compute the average trace distance between the state obtained after running the ideal circuit and that obtained after running the noisy circuit, both with and without REAS applied and both with and without the single Pauli transformation (sampling number: 4000). The simulation shown in Fig. 4 shows that the circuit implemented with the single Pauli transformation included displays a smaller trace distance in the small γ regime compared to the case where Pauli rotation is implemented directly without said transformation procedure.

V. EXPERIMENTAL DEMONSTRATION

To complement the numerical results above, we finally demonstrate the performance of a special case of our REAS method on available quantum computing systems. For this experiment, we make the following assumptions which seem well justified in modern quantum computers with superconducting qubits: (i) two-qubit gates have much larger errors

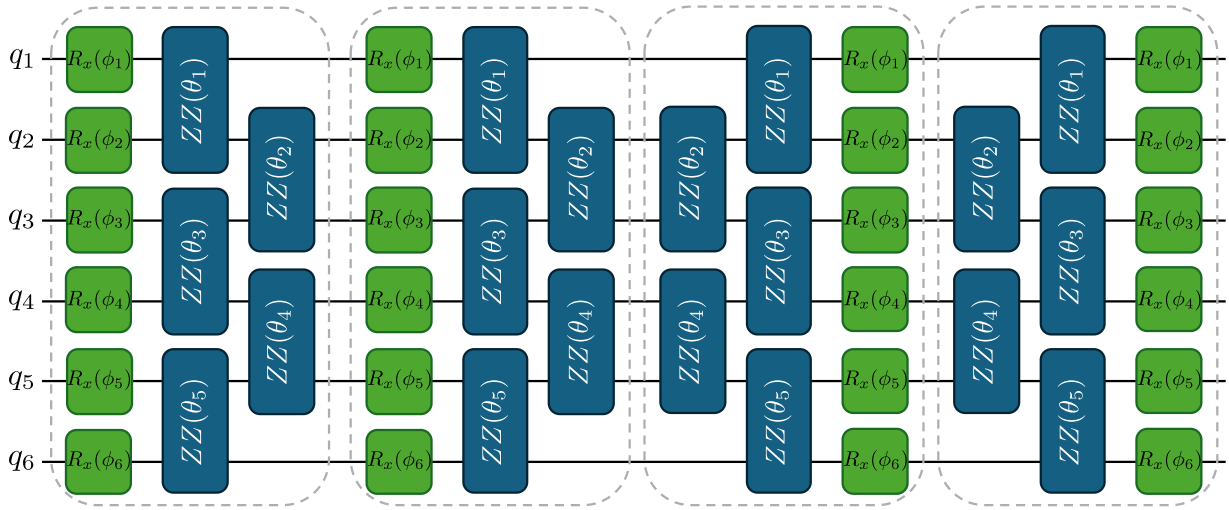


FIG. 5. Example circuit. A compute-uncompute Trotterized circuit of simulating the time evolution of a one-dimensional transversal-field Ising model on six qubits using two Trotter steps. For the experimental demonstration, we employ circuits with the same structure on 100 qubits with 1 to 16 Trotter steps.

than single-qubit ones, the latter of which are negligible; and (ii) standard calibration (i.e., procedures used in IBM Quantum systems) of rotation angles for two-qubit Pauli rotation gates is highly accurate. Based on these assumptions, we tested a variant of REAS using the following setup: (i) the single Pauli transformation step was not applied, meaning only coherent errors were addressed; (ii) ordinary calibration (without twirling) was performed for all gates; (iii) consecutive single-qubit gates applied to the same qubit were merged as much as possible (i.e., in Fig. 1, the inserted gates were merged with the computation gates). As such, the experimental demonstration does not unleash the full power of REAS and instead performs a variant of randomized compiling [34,36,48]. Nonetheless, our experiment shows that such a method works well for suppressing the accumulation of errors in quantum circuits with many qubits, which, to the best of our knowledge, has not yet been demonstrated on presently available platforms.

We use Trotterized circuits to simulate the time evolution of a one-dimensional transverse-field Ising model and analyze how the error scales with respect to the number of Trotter steps, which is proportional to the number of two-qubit gate layers. The Hamiltonian of the model is

$$H = -\sum_{i=1}^{N-1} g_i Z_i Z_{i+1} - \sum_{i=1}^{N} h_i X_i, \tag{18}$$

where g_i represents the interaction strength between spins i and $i+1$, h_i represents the external magnetic field strength at site i, and N is the number of spins (i.e., qubits) of the system.

In order to calculate the ideal output of the circuit, we use the so-called compute-uncompute procedure, which consists of a first stage where the original circuit U is applied, and a second stage where it is reversed via application of $U^\dagger$. The ideal output state of such circuits applied to an initial blank register $|0\rangle^{\otimes N}$ will be identical to the input state, which has trivial expectation value (i.e., either 0 or 1) for any Pauli observable. This allows us to compute how far the experimentally obtained expectation values are from the ideal ones.

In Fig. 5, we show an example circuit in the case of $N = 6$ with two Trotter steps. As shown in the figure, the ith ZZ term in the Hamiltonian of Eq. (18) generates a $ZZ(\theta_i)$ rotation gate in each Trotter step, where θ_i is an angle proportional to g_i and the time step. In our experiment, we choose the value θ_i by sampling the pulse amplitude that implements the $ZZ(\theta_i)$ rotation gate (see SM Sec. S6 for details [44]). As for the X_i term in the Hamiltonian, this evolution is implemented by an $R_x(\phi_i)$ gate, where ϕ_i is an angle proportional to h_i and the time step. We use the uniform value $\phi_i = 0.1\pi$ for all $i = 1 \ldots N$ in our experiment.

We evaluate the amount of errors by the RMS error of the expectation value of a Pauli observable P, i.e.,

$$\sqrt{\sum_k p_k \| \langle P \rangle_{\text{ideal}} - \langle P \rangle_{\exp(k)} \|^2}, \tag{19}$$

where p_k is the probability of drawing the kth set of random Paulis in the REAS procedure and $\langle P \rangle_{\exp(k)}$ represents the expectation value of P experimentally observed in the kth circuit realization.[2]

For the experiment, we made use of 127-qubit IBM Quantum processors [49], namely, `ibm_brisbane`, `ibm_osaka`, and `ibm_cusco`, comparing the average error in expectation values both with and without the REAS procedure applied. In both cases, we considered Trotter steps: 1, 2, 3, 4, 5, 6, 8, 10, 13, 16. Here, e.g., 16 Trotter steps corresponds to 64 two-qubit gate layers since each step consists of two layers and the steps are doubled by uncomputing. For the cases where REAS is applied, we sampled 25 circuits for each number of Trotter

[2]Note that we are interested in the average of errors of expectation values (i.e., the effect of error suppression), not in measuring the error on the averaged expectation value (i.e., the effect of error mitigation).

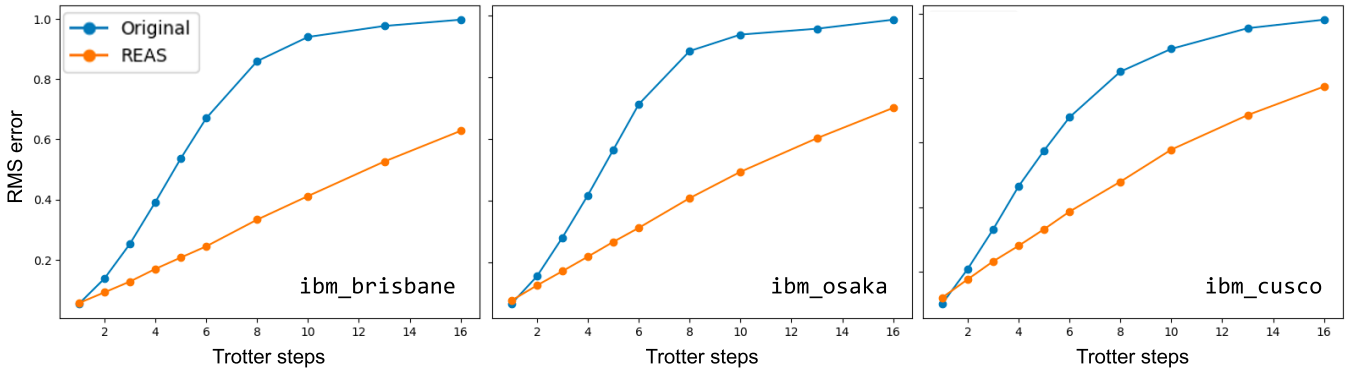


FIG. 6. Experimental demonstration. Average RMS error over 100 weight-1 Z observables by Trotter steps with REAS (orange, REAS) and without REAS (blue, original) using three IBM Quantum processors: `ibm_brisbane`, `ibm_osaka`, and `ibm_cusco`. These results highlight the efficacy and robustness of our method across distinct quantum devices.

steps, meaning that we ran 10 circuits for the situation without REAS and 250 circuits for that with it. In order to balance the total number of shots, we set 10 000 shots for the former case and 400 shots for the latter. For each circuit, we measured all weight-1 Pauli observables for 100 qubits and computed the RMS error [see Eq. (19)] for each observable. To accurately estimate the expectation values of the observables, we applied a readout error mitigation technique called twirled readout error eXtinction (TREX) [50]. Figure 6 demonstrates how the RMS error (averaged over all 100 weight-1 Z observables) scales with Trotter steps. Clearly, the suppression of errors via REAS improves as the number of Trotter steps grows for all experiments on the three processors.

VI. DISCUSSION

We have presented an advanced, robust method of quantum error suppression. The ability to deal with non-Clifford circuits and incorporate both coherent and decoherence errors, both in the original circuit and in its modification, stems from combining robust phase estimation techniques with a calibrated Pauli twirling-type procedure, as described in Secs. II and III. To highlight the immediate applicability of our results, we present numerical evidence for the quadratic reduction in error growth presented in Sec. IV and furthermore present an experimental demonstration in Sec. V. Since the types of errors that our method is tailor made to deal with go far beyond the current state of the art, we expect our methods to be widely adopted across a variety of platforms.

Looking forward, extending our method to deal with leakage or qubit loss will be an important generalization. Second, an improved technique of dealing with correlated errors across many qubits, e.g., nonlocal crosstalk. Lastly, actively suppressing errors in complex non-Markovian processes remains of utmost importance, especially as the timescales over which one can access quantum systems are ever shortening, making nontrivial temporal correlations more prevalent [51–57].

ACKNOWLEDGMENTS

We thank J. Emerson for interesting discussions. We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. P.T. acknowledges funding from the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowships for Research in Japan and the IBM-UTokyo Laboratory. N.Y. is supported by PRESTO, JST, Grant No. JPMJPR2119, COI-NEXT program Grant No. JPMJPF2221, JST ERATO Grant No. JPMJER2302, and JST CREST Grant No. JPMJCR23I4, Japan and IBM Quantum. M.M. was supported by MEXT Quantum Leap Flagship Program (MEXT QLEAP) Grants No. JPMXS0118069605 and No. JPMXS0120351339, JSPS KAKENHI Grants No. 21H03394 and No. 23K21643, and IBM Quantum.

DATA AVAILABILITY

The data are available upon reasonable request from the authors.

[1] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation, arXiv:quant-ph/9611028.

[2] P. W. Shor, Algorithms for quantum computation: Discrete logarithms and factoring, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Washington, DC, 1994), p. 124.

[3] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[4] A. Acín, I. Bloch, H. Buhrman, T. Calarco, C. Eichler, J. Eisert, D. Esteve, N. Gisin, S. J. Glaser, F. Jelezko, S. Kuhr, M. Lewenstein, M. F. Riedel, P. O. Schmidt, R. Thew, A. Wallraff, I. Walmsley, and F. K. Wilhelm, The quantum

technologies roadmap: A European community view, New J. Phys. **20**, 080201 (2018).

[5] D. D. Awschalom, H. Bernien, R. Brown, A. Clerk, E. Chitambar, A. Dibos, J. Dionne, M. Eriksson, B. Fefferman, G. D. Fuchs, J. Gambetta, E. Goldschmidt, S. Guha, F. J. Heremans, K. D. Irwin, A. B. Jayich, L. Jiang, J. Karsch, M. Kasevich, S. Kolkowitz *et al.*, A roadmap for quantum interconnects, Technical Report ANL-22/83; 179439, Argonne National Laboratory, Argonne, IL, 2022.

[6] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[7] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, Phys. Rev. A **54**, 1098 (1996).

[8] A. M. Steane, Error correcting codes in quantum theory, Phys. Rev. Lett. **77**, 793 (1996).

[9] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Perfect quantum error correcting code, Phys. Rev. Lett. **77**, 198 (1996).

[10] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[11] R. Acharya *et al.* (Google Quantum AI and collaborators), Quantum error correction below the surface code threshold, Nature (London) **638**, 920 (2025).

[12] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[13] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[14] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018).

[15] B. Koczor, Exponential error suppression for near-term quantum devices, Phys. Rev. X **11**, 031057 (2021).

[16] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, Phys. Rev. X **11**, 041036 (2021).

[17] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, Phys. Rev. Lett. **129**, 020502 (2022).

[18] H. Hakoshima, S. Endo, K. Yamamoto, Y. Matsuzaki, and N. Yoshioka, Localized virtual purification, Phys. Rev. Lett. **133**, 080601 (2024).

[19] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95**, 045005 (2023).

[20] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[21] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature (London) **567**, 491 (2019).

[22] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, Scalable error mitigation for noisy quantum circuits produces competitive expectation values, Nat. Phys. **19**, 752 (2023).

[23] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).

[24] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Universal cost bound of quantum error mitigation based on quantum estimation theory, Phys. Rev. Lett. **131**, 210601 (2023).

[25] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, Phys. Rev. Lett. **131**, 210602 (2023).

[26] Y. Quek, D. Stilck França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, Nat. Phys. **20**, 1648 (2024).

[27] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999).

[28] M. B. Hastings, Turning gate synthesis errors into incoherent errors, Quantum Inf. Comput. **17**, 488 (2017).

[29] E. Campbell, Shorter gate sequences for quantum computing by mixing unitaries, Phys. Rev. A **95**, 042306 (2017).

[30] S. Akibue, G. Kato, and S. Tani, Probabilistic state synthesis based on optimal convex approximation, npj Quantum Inf. **10**, 3 (2024).

[31] W. Dür, M. Hein, J. I. Cirac, and H.-J. Briegel, Standard forms of noisy quantum operations via depolarization, Phys. Rev. A **72**, 052326 (2005).

[32] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011).

[33] M. R. Geller and Z. Zhou, Efficient error models for fault-tolerant architectures and the Pauli twirling approximation, Phys. Rev. A **88**, 012314 (2013).

[34] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[35] J. Hines, M. Lu, R. K. Naik, A. Hashim, J.-L. Ville, B. Mitchell, J. M. Kriekebaum, D. I. Santiago, S. Seritan, E. Nielsen, R. Blume-Kohout, K. Young, I. Siddiqi, B. Whaley, and T. Proctor, Demonstrating scalable randomized benchmarking of universal gate sets, Phys. Rev. X **13**, 041030 (2023).

[36] J. dos Santos, B. Bar, and R. Uzdin, Deeper quantum circuits via pseudo-twirling coherent errors mitigation in non-Clifford gates, arXiv:2401.09040.

[37] Y. Akahoshi, K. Maruyama, H. Oshima, S. Sato, and K. Fujii, Partially fault-tolerant quantum computing architecture with error-corrected Clifford gates and space-time efficient analog rotations, PRX Quantum **5**, 010337 (2024).

[38] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, Phys. Rev. X **11**, 041039 (2021).

[39] J.-L. Ville, A. Morvan, A. Hashim, R. K. Naik, M. Lu, B. Mitchell, J.-M. Kreikebaum, K. P. O'Brien, J. J. Wallman, I. Hincks, J. Emerson, E. Smith, E. Younis, C. Iancu, D. I. Santiago, and I. Siddiqi, Leveraging randomized compiling for the quantum imaginary-time-evolution algorithm, Phys. Rev. Res. **4**, 033140 (2022).

[40] S. Ferracin, A. Hashim, J.-L. Ville, R. Naik, A. Carignan-Dugas, H. Qassim, A. Morvan, D. I. Santiago, I. Siddiqi, and

J. J. Wallman, Efficiently improving the performance of noisy quantum computers, Quantum **8**, 1410 (2024).

[41] T. Kurita, H. Qassim, M. Ishii, H. Oshima, S. Sato, and J. Emerson, Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver, Quantum **7**, 1184 (2023).

[42] N. Goss, S. Ferracin, A. Hashim, A. Carignan-Dugas, J. M. Kreikebaum, R. K. Naik, D. I. Santiago, and I. Siddiqi, Extending the computational reach of a superconducting qutrit processor, arXiv:2305.16507.

[43] N. Sundaresan, I. Lauer, E. Pritchett, E. Magesan, P. Jurcevic, and J. M. Gambetta, Reducing unitary and spectator errors in cross resonance with optimized rotary echoes, PRX Quantum **1**, 020318 (2020).

[44] See Supplemental Material at http://link.aps.org/supplemental/10.1103/yx9t-nctk for technical details and proofs.

[45] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Phys. Rev. A **92**, 062315 (2015).

[46] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde, Demonstrating Heisenberg-limited unambiguous phase estimation without adaptive measurements, New J. Phys. **11**, 073023 (2009).

[47] C. Arenz, R. Hillier, M. Fraas, and D. Burgarth, Distinguishing decoherence from alternative quantum theories by dynamical decoupling, Phys. Rev. A **92**, 022102 (2015).

[48] D. Layden, G. Mazzola, R. V. Mishmash, M. Motta, P. Wocjan, J.-S. Kim, and S. Sheldon, Quantum-enhanced Markov chain Monte Carlo, Nature (London) **619**, 282 (2023).

[49] IBM Quantum, https://quantum.ibm.com.

[50] E. van den Berg, Z. K. Minev, and K. Temme, Model-free readout-error mitigation for quantum expectation values, Phys. Rev. A **105**, 032620 (2022).

[51] G. A. L. White, C. D. Hill, F. A. Pollock, L. C. L. Hollenberg, and K. Modi, Demonstration of non-Markovian process characterisation and control on a quantum processor, Nat. Commun. **11**, 6301 (2020).

[52] Y. Guo, P. Taranto, B.-H. Liu, X.-M. Hu, Y.-F. Huang, C.-F. Li, and G.-C. Guo, Experimental demonstration of instrument-specific quantum memory effects and non-Markovian process recovery for common-cause processes, Phys. Rev. Lett. **126**, 230401 (2021).

[53] K. Goswami, C. Giarmatzi, C. Monterola, S. Shrapnel, J. Romero, and F. Costa, Experimental characterization of a non-Markovian quantum process, Phys. Rev. A **104**, 022432 (2021).

[54] P. Figueroa-Romero, K. Modi, R. J. Harris, T. M. Stace, and M.-H. Hsieh, Randomized benchmarking for non-Markovian noise, PRX Quantum **2**, 040351 (2021).

[55] P. Figueroa-Romero, K. Modi, and M.-H. Hsieh, Towards a general framework of randomized benchmarking incorporating non-Markovian noise, Quantum **6**, 868 (2022).

[56] G. White, F. Pollock, L. Hollenberg, K. Modi, and C. Hill, non-Markovian quantum process tomography, PRX Quantum **3**, 020344 (2022).

[57] G. A. L. White, K. Modi, and C. D. Hill, Filtering crosstalk from bath non-Markovianity via spacetime classical shadows, Phys. Rev. Lett. **130**, 160401 (2023).