Robust Error Accumulation Suppression

Tatsuki Odake,[1] Philip Taranto,[1] Nobuyuki Yoshioka,[2] Toshinari Itoko,[3]
Kunal Sharma,[4] Antonio Mezzacapo,[4] 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 Applied Physics, Graduate School of Engineering, The
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
[3]*IBM Quantum, IBM Research – Tokyo, Tokyo, Japan*
[4]*IBM Quantum, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598*
(Dated: 31st January 2024)

We present an advanced quantum error suppression technique, which we dub *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 $\theta \in [0, \pi)$; since such gates form a universal gate set, our results apply to a strictly larger class of circuits than those comprising only Clifford gates, thereby generalizing previous results. In the case of coherent errors—which include crosstalk—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 have non-negligible coherent noise. Furthermore, we show that REAS can also suppress certain types of decoherence noise by transforming some gates to be robust against such noise, which is verified by the demonstration of the quadratic suppression of error scaling in the numerical simulation. Our results, therefore, present an advanced, robust method of error suppression that can be used in conjunction with error correction as a viable path toward fault-tolerant quantum computation.

I. INTRODUCTION

A central problem in quantum information processing is to counteract the noise effect that arises from uncontrolled interactions between the system of interest and its external environment. Such errors build up rapidly throughout quantum computations, typically limiting their practical applicability to logarithmic-depth circuits [1]. At the same time, the discovery of quantum algorithms that can run exponentially more efficiently than their classical counterparts, e.g., Shor's algorithm [2, 3], has stimulated global efforts to developing quantum technologies [4, 5].

Key developments in quantum error correction provide a promising route to scaleable, fault-tolerant quantum devices [6–10]; however, the errors must be well below the so-called "error correction threshold" to provide any practical benefit. In other words, 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, and thus, one must incorporate other methodologies that do not rely on correcting the errors via measurement and feedback.

To this end, two classes of such complementary approaches are widely considered, namely *error mitigation* and *error suppression*. Quantum error mitigation is a technique that aims to reduce the bias in estimating expectation values of physical observables at the cost of increased estimation variance [11–18]. Incorporating such error mitigation methods has recently led to demonstrations of quantum simulations that are even competitive with state-of-the-art classical counterparts [19–21]. Nevertheless, a crucial drawback of error mitigation is the exponential increase in the number of measurements required to improve estimation accuracy [22–24].

Quantum error suppression takes a different approach: instead of shifting the burden to measurement overheads, it aims to modify parts of the noisy circuit in order to construct an implementable quantum operation that best approximates the target one. A paradigmatic example of this is the well-known dynamical decoupling protocol, which cancels out decoherence effects on idling qubits by periodically repeating fast random pulses that delete any spuriously imprinted information [25]. However, dynamical decoupling only serves to keep the state from decohering, and it is a priori unclear how to extend it to situations where noisy gates are being applied for computation. Fortunately, in such scenarios, coherent errors in circuits composed of Clifford gates, such as those caused by accidental gate overrotation, can be suppressed via related techniques that similarly beneficially harness randomness, e.g., probabilistic compiling [26–28], Pauli twirling [29–31], and randomized compiling [32, 33].

Nonetheless, the class of noise that can be counteracted via such methods is rather limited and often surpassed in realistic settings. Although coherent gate errors in Clifford circuits can be dealt with by previous techniques, a number of practical challenges remain that must be surmounted to unleash the full potential of error suppression. Firstly, most known methods break down when considering non-Clifford circuits. Secondly, it is not clear that the known techniques 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 determined. To summarize, it remains an open question whether one can reliably suppress a

wide class of errors that concurrently include both coherent and decoherent effects—as is the case in reality—even for non-Clifford circuits.

In this work, we present an advanced quantum error suppression technique, which we dub *robust error accumulation suppression* (**REAS**), to address these problems. 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 $\theta \in [0, \pi)$; since such gates form a universal gate set, our results apply to a strictly larger class of circuits than those comprising only Clifford gates, thereby generalizing previous results. In the case of coherent errors—which includes spectator error as discussed in Ref. [34]—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 have non-negligible coherent noise. Furthermore, we show that REAS can also suppress certain types of decoherence noise, namely those that are well-approximated by a weak system-environment coupling, by demonstrating numerical evidence for the quadratic suppression of error scaling. Our results, therefore, present an advanced, robust method of error suppression that can be used in conjunction with error correction as a viable path toward fault-tolerant quantum computation.

II. ROBUST ERROR ACCUMULATION SUPPRESSION

A. Layered Circuits

We begin by presenting the framework of 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 can cover all gates of the form $e^{-i\sigma\theta'}$ for $\theta' \in \mathbb{R}$ up to a global phase). Such a circuit consisting of L layers of gates applied to an n-qubit system can be expressed as

$$\prod_{\ell=1}^{L} \left(\bigotimes_{\vec{q} \in Q_l} e^{-i\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}} \right). \tag{1}$$

Here, ℓ labels the layers of the circuit, Q_ℓ consists of disjoint subsets of $\{1, \ldots, n\}$ of at most length two, with $\vec{q} \in Q_\ell$ specifying the particular qubits that acted on non-trivially in layer ℓ (with identity operators implied on the rest). One can furthermore modify the above description by grouping layers into B blocks (which can now comprise sequentially implemented gates on certain qubits), leading to the form

$$\prod_{b=1}^{B} \prod_{\ell_b} \left(\bigotimes_{\vec{q} \in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}} \right), \tag{2}$$

where now ℓ_b labels the layer *within* each block b. We refer to the Pauli operators $\sigma_{\ell_b,\vec{q}}$ inside the exponential in the desired circuit as "computation gates" in order to distinguish them from another type of Pauli gates that will be introduced in the following section.

B. Motivating Example

In order 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. (2). Intuitively, the operation applied to each layer is shifted slightly by some additional unitary rotation, leading to the actual implementation of

$$\left(\bigotimes_{\vec{q} \in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}} \right) [I + i\epsilon C_{\ell_b}], \tag{3}$$

where $\|C_{\ell_b}\|_{\mathrm{op}} \leq 1$ 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 Pauli twirling method, 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. Since, for tensor products of qubit Pauli gates, any such overall operation either commutes or anti-commutes with any Pauli gates $e^{-i\sigma\theta}$ applied in the subsequent block, one can define

$$s := \begin{cases} 1 & \text{if } \sigma \text{ and } \sigma_{\vec{v}} \text{ commute} \\ -1 & \text{if } \sigma \text{ and } \sigma_{\vec{v}} \text{ anti-commute} \end{cases} \tag{4}$$

in order to express the transformation between blocks succinctly as

$$e^{-i\sigma\theta} \mapsto \sigma_{\vec{v}} e^{-is\sigma\theta} \sigma_{\vec{v}}. \tag{5}$$

Note that the technique of transforming Pauli rotation gates to the form of Eq. (5) itself is not new (see, e.g., Ref. [35]). However, in previous works, the error on the inserted Pauli gate $\sigma_{\vec{v}}$, as well as the possible dependence of such errors on the rotation direction s, was not explicitly considered. Here, we demonstrate explicitly that such errors can also be dealt with, therefore showing that our method of error suppression is robust.

In the absence of errors, both the original gate $e^{-i\sigma\theta}$ and the average transformed one $(1/4)^n \sum_{\vec{v}} \sigma_{\vec{v}} e^{-is\sigma\theta} \sigma_{\vec{v}}$ lead to the same transformation. However, whenever the computation gate $e^{is\sigma_{\vec{v}}\theta}$ is 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

$$\left(\frac{1}{4}\right)^n \sum_{\vec{v}} \sigma_{\vec{v}} \left[e^{-is\sigma\theta}(I + i\epsilon C_s) \right] \sigma_{\vec{v}} = \left(\frac{1}{4}\right)^n \sum_{\vec{v}} e^{-i\sigma\theta}[I + i\epsilon\sigma_{\vec{v}}C_s\sigma_{\vec{v}}] = e^{-i\sigma\theta}[I + i\epsilon(\alpha I + \beta\sigma)] = e^{i\epsilon\alpha}e^{-i(\theta-\epsilon\beta)\sigma} + O(\epsilon^2), \tag{6}$$

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

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

and

$$\left(\frac{1}{4^n}\right) \sum_{\vec{v}\in\{0,1,2,3\}^n} \sigma_{\vec{v}}s\sigma'\sigma_{\vec{v}} = \begin{cases} \sigma & \text{for } \sigma = \sigma' \\ 0 & \text{for } \sigma \neq \sigma' \end{cases}. \tag{8}$$

Since $e^{i\epsilon\alpha}$ 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 expression for how the overall layered circuit with coherent errors transforms on average upon inserting random Pauli gates between blocks is

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

$$= \prod_{b=1}^{B} \prod_{\ell_b} \left(\bigotimes_{\vec{q}\in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\,(\theta_{\ell_b,\vec{q}}+\epsilon\beta_{\ell_b,\vec{q}})} \right) + O(\epsilon + B\epsilon^2). \tag{9}$$

Here, we have introduced the notation $\vec{u}_{j:k}$ for two vectors that satisfy $\sigma_{\vec{u}_{j:k}} \propto \sigma_{\vec{u}_j}\sigma_{\vec{u}_k}$ and implicitly understand the value of s to depend upon whether or not each inserted Pauli gate string commutes or anti-commutes with each layer of computational gates in the neighboring block. We refer to the random Pauli operators $\sigma_{\vec{v}_1}, \sigma_{\vec{v}_{2:1}} \dots, \sigma_{\vec{v}_B}$ as the "inserted Pauli gates" to distinguish them from the computational Pauli gates.

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 a level of robustness.

C. 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 A. 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 B. Details are provided in App. A.

Calibration. The calibration step is performed by considering each layer of the circuit individually. For each chosen layer ℓ, one repeats the computational gates of that layer K times and intersperses it with random Pauli gate strings, which leads to[1]

$$\left(\frac{1}{4^n}\right)^K \sum_{\vec{v}_1,\ldots,\vec{v}_K} \sigma_{\vec{v}_K} \left(\bigotimes_{\vec{q} \in Q_\ell} e^{-is\,\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}\right) \prod_{k=1}^{K} \left[\sigma_{\vec{v}_{k+1:k}} \left(\bigotimes_{\vec{q} \in Q_\ell} e^{-is\,\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}\right)\right] \sigma_{\vec{v}_1}. \tag{10}$$

Since the undesired shift in the rotation angle accumulates by repeating the same layer K times, the value of $\epsilon\beta_{\ell,\vec{q}}$ can be estimated via robust phase estimation [36]. Since each Q_ℓ only contains disjoint subsets, this procedure can be applied for all ℓ simultaneously. 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 $O(\epsilon + B\epsilon^2)$ 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 gate strings into the calibrated layered circuit to specifically suppress the errors in each block. Explicitly, 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.

III. MAIN RESULTS

A. Suppressing the Accumulation of Coherent Errors

The REAS scheme suppresses *any* coherent error, both on the computational Pauli gates and on the inserted ones, in such a way that the error on the overall circuit $\prod_{\ell=1}^{L} \left(\bigotimes_{\vec{q} \in Q_\ell} e^{-i\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}\right)$ decreases from $O(\epsilon L)$ to $O(\epsilon + \epsilon^2 L)$ on average.[2]

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 related to the difference between the ideal and any individual gate, in App. B, we show that the former implies that the root mean square (**RMS**) of the latter difference 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 $\prod_{\ell=1}^{L} \left(\bigotimes_{\vec{q} \in Q_\ell} e^{-i\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}\right)$ and any single-shot unitary operator implemented after calibration in a given run with the random Pauli gate string inserted scales as $O(\sqrt{\epsilon + \epsilon^2 L})$ which approaches $O(\epsilon\sqrt{L})$ in the regime $L \gg 1/\epsilon$.

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

$$\left(\frac{1}{4^n}\right)^B \sum_{\vec{v}_1,\ldots,\vec{v}_B} \left[\sigma_{\vec{v}_B} \prod_{\ell_B} \left(\bigotimes_{\vec{q} \in Q_{\ell_B}} e^{-is\,\sigma_{\ell_B,\vec{q}}\theta_{\ell_B,\vec{q}}}\right)\right] \prod_{b=1}^{B-1} \left[\sigma_{\vec{v}_{b+1:b}} \prod_{\ell_b} \left(\bigotimes_{\vec{q} \in Q_{\ell_b}} e^{-is\,\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}\right)\right] \sigma_{\vec{v}_1}, \tag{11}$$

---

[1] To reiterate, the value of s in each term here is not the same, as it depends explicitly upon the commutation relation with the Pauli gate string inserted in its neighboring position.

[2] It is straightforward to see that if one takes the maximum size of any block to be upper bounded by a constant, this result generalizes to reduce the error on the overall circuit $\prod_{b=1}^{B} \prod_{\ell_b} \left(\bigotimes_{\vec{q} \in Q_{\ell_b}} e^{-i\sigma_{\ell_b,\vec{q}}\theta_{\ell_b,\vec{q}}}\right)$ from $O(\epsilon B)$ to $O(\epsilon + \epsilon^2 B)$.

so that the overall gate sequence is transformed into an expression of form

$$\left(\frac{1}{4^n}\right)^B \sum_{\vec{v}_1,\ldots,\vec{v}_B}$$

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

$$(12)$$

This can be rewritten to

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

$$= [\cdots]\left(\frac{1}{4^n}\right)\sum_{\vec{v}_{b'}}\sigma_{\vec{v}_{b'}}\left(\frac{1}{4^n}\sum_{\vec{v}_{b'+1}}i\epsilon C_{\vec{v}_{b';b'+1}}\right)\sigma_{\vec{v}_{b'}}[\cdots] = O(\epsilon^2),$$

$$(13)$$

where $[\cdots]$ represents the terms that are not important for the following result. Since $\sum_{\vec{v}_{b'+1}}C_{\vec{v}_{b'}+\vec{v}_{b'+1}}$ is independent of $\vec{v}_{b'}$ and the term in $i\epsilon C_{\vec{v}_{b';b'+1}}$ that is proportional to the identity can be taken to be $O(\epsilon^2)$ without loss of generality, as shown in App. A.

B. Suppressing the Accumulation of Decoherence Errors (with Single Pauli Transformation)

The original procedure of REAS comprising calibration and then inserting random Pauli gates cannot 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 App. C for details.

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

$$e^{isYIt} \to e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}$$

$$e^{isZIt} \to e^{-i\frac{\pi}{4}XI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{i\frac{\pi}{4}XI}$$

$$e^{isXIt} \to e^{i\frac{\pi}{4}ZI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{-i\frac{\pi}{4}ZI}.$$

$$(14)$$

This transformation makes the single-qubit Pauli rotation gate robust against decoherence errors, thereby enabling the tolerance of the REAS procedure against such noise. As we will discuss below, this mapping allows us to suppress decoherence errors that are up to first order in the operator norm of the interaction Hamiltonian.[3]

Upon invoking this 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 original Pauli rotation gates and the inserted random Pauli operators in such a way that the error on the overall circuit $\prod_{\ell=1}^{L}\left(\bigotimes_{\vec{q}\in Q_\ell}e^{-i\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}\right)$ changes from $O(\epsilon L)$ to $O(\epsilon + \epsilon^2 L)$ in the average case; similarly, the RMS of the diamond distance between the error-free unitary operation implemented by $\prod_{\ell=1}^{L}\left(\bigotimes_{\vec{q}\in Q_\ell}e^{-i\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}\right)$ and any single-shot unitary operator implemented after calibration in a given run with random Pauli gate string inserted scales as $O(\sqrt{\epsilon + \epsilon^2 L})$.

We model the decoherence error via introducing an interaction Hamiltonian $H'_{\ell,\vec{q},s_{\vec{q}}}(t) \in \mathcal{H}\otimes\mathcal{H}_{\text{env}}$ of order ϵ that leads to the pulse Hamiltonian that implements the desired $e^{-is_{\vec{q}}\sigma_{\ell,\vec{q}}\theta_{\ell,\vec{q}}}$. Note that, since $H'_{\ell,\vec{q},s_{\vec{q}}}(t)$ can contain arbitrary terms that lead to both coherence and incoherence errors, our statements here are a direct generalization of those made in the previous section regarding coherent errors.

---

[3] The ZZ rotation common to all terms can be changed to any other two-qubit Pauli rotation (e.g., ZX) as long as the s-dependence only appears non-trivially at the two-qubit level.

We now move on to analyze the class of decoherence errors that can be dealt with via REAS. We assume that the decoherence noise applied to the layer $\left(\bigotimes_{\vec{q} \in Q_\ell} e^{-is_{\vec{q}} \sigma_{\ell,\vec{q}} \theta_{\ell,\vec{q}}} \right)$ can be expressed as

$$
\mathcal{T} \exp \left[-i \int_0^T \mathrm{d}t' \left(\sum_{\vec{q} \in Q_\ell} H_{\ell,\vec{q},s_{\vec{q}}}(t') \right) + H_{\mathrm{env}} + \sum_{\vec{q} \in Q_\ell} H'_{\ell,\vec{q},s_{\vec{q}}}(t') \right], \tag{15}
$$

where $H_{\ell,\vec{q},s_{\vec{q}}}(t') \in \mathcal{H}$ is the time-dependent pulse Hamiltonian to implement the desired gate $e^{-is_{\vec{q}} \sigma_{\ell,\vec{q}} \theta_{\ell,\vec{q}}}$, $H'_{\ell,\vec{q},s_{\vec{q}}}(t') \in \mathcal{H} \otimes \mathcal{H}_{\mathrm{env}}$ is the undesired Hamiltonian that is induced when applying $H_{\ell,\vec{q},s_{\vec{q}}}$ in the presence of the environment, $H_{\mathrm{env}} \in \mathcal{H}_{\mathrm{env}}$ is the background Hamiltonian on 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 $H'_{\ell,\vec{q},s_{\vec{q}}}(t')$ does not contain interaction terms involving more than two systems (either qubits in $\mathcal{H}$ or subsystem in $\mathcal{H}_{\mathrm{env}}$); since most Hamiltonians in nature involve at most two-body interactions, this assumption seems natural.

Many relevant decoherence mechanisms, such as depolarizing, amplitude damping, and dephasing, can be expressed by this interaction Hamiltonian models. Indeed, for single-qubit decoherence channels

$$
\mathcal{E}_p^{\mathrm{dpl}}(\rho) := (1-p)\rho + p\frac{I}{2}, \qquad \mathcal{E}_\gamma^{\mathrm{ad}}(\rho) := E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger, \qquad \mathcal{E}_p^{\mathrm{dph}}(\rho) := (1-p)\rho + pZ\rho Z, \tag{16}
$$

where

$$
E_0 := \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{\gamma} \end{pmatrix} \quad \text{and} \quad E_1 := \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \tag{17}
$$

their action can be modeled as

$$
\mathcal{E}_{\sin^2(2\delta t)}^{\mathrm{dpl}}(\rho) = \mathrm{tr}_{\mathrm{env}} \left[e^{-i\delta t(X \otimes X_{\mathrm{env}} + Y \otimes Y_{\mathrm{env}} + Z \otimes Z_{\mathrm{env}})} \left(\rho \otimes \frac{I_{\mathrm{env}}}{2} \right) e^{i\delta t(X \otimes X_{\mathrm{env}} + Y \otimes Y_{\mathrm{env}} + Z \otimes Z_{\mathrm{env}})} \right]
$$

$$
\mathcal{E}_{\sin^2(\delta t)}^{\mathrm{ad}}(\rho) = \mathrm{tr}_{\mathrm{env}} \left[e^{-i\frac{\delta t}{2}(X \otimes X_{\mathrm{env}} + Y \otimes Y_{\mathrm{env}})} (\rho \otimes |0\rangle \langle 0|_{\mathrm{env}}) e^{i\frac{\delta t}{2}(X \otimes X_{\mathrm{env}} + Y \otimes Y_{\mathrm{env}})} \right]
$$

$$
\mathcal{E}_{\sin^2(\delta t)}^{\mathrm{dph}}(\rho) = \mathrm{tr}_{\mathrm{env}} \left[e^{-i\delta t(Z \otimes Z_{\mathrm{env}})} \left(\rho \otimes \frac{I_{\mathrm{env}}}{2} \right) e^{i\delta t(Z \otimes Z_{\mathrm{env}})} \right]. \tag{18}
$$

As is the case regarding dynamical decoupling, note that different ways of modeling decoherence in this interaction picture can lead to the different performances of REAS. For instance, in a similar discussion to the one shown in [37], any decoherence error modeled by quantum channels on the system without referring to its environment cannot be erased by REAS (as can be easily seen by considering the case where depolarizing channels act on all gates).

IV. NUMERICAL RESULTS

We now present numerical simulations that showcase the efficacy of our results. In all situations, we consider the setting where $\mathcal{H}$ (qubits: S_1, S_2) and $\mathcal{H}_{\mathrm{env}}$ (qubits: E_1, E_2) are two-qubit systems. All Pauli rotation gates and inserted Pauli gates are affected by decoherence noise simulated by applying the global unitary $e^{-i\gamma H} \in \mathcal{H} \otimes \mathcal{H}_{\mathrm{env}}$, where H is normalized as $\|H\|_2 = 1$ and the parameter γ specifies the scale of noise, after each gate. 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 $S_1 E_1$ and $S_2 E_2$ are randomly chosen for each gate (unless specified otherwise). For simplicity, we assume that the rotation angle can only take values in $\pm \frac{\pi}{2}$, $\pm \frac{\pi}{4}$, and $\pm \frac{\pi}{8}$ and do not consider parallel implementation of the computational Pauli gates.

For the deep circuit in Sec. IV A, we compare our REAS calibration method to the theoretical prediction, as well as with conventional calibration techniques (i.e., performing phase estimation without inserting random Pauli gates). For the simulations shown in Secs. IV B and IV C, we chose the noise Hamiltonians in such a way that the first-order γ term in the shift in rotation angle becomes 0 to avoid the calibration step, which turned out to require too long computation time to give precise results. Finally, the initial state of the circuit is $|0\rangle^{\otimes 4}$ in all simulations.

A. Effectiveness of REAS in Deep Circuit

Here, we demonstrate the effectiveness of our REAS scheme in a deep circuit. We consider a circuit block of depth 3 in Fig. 1 (a) containing only two-qubit Pauli gates. The string "YZ8" refers to the Pauli rotation $e^{-i\frac{\pi}{8} ZY} \otimes I_{\mathrm{env}}$ (note
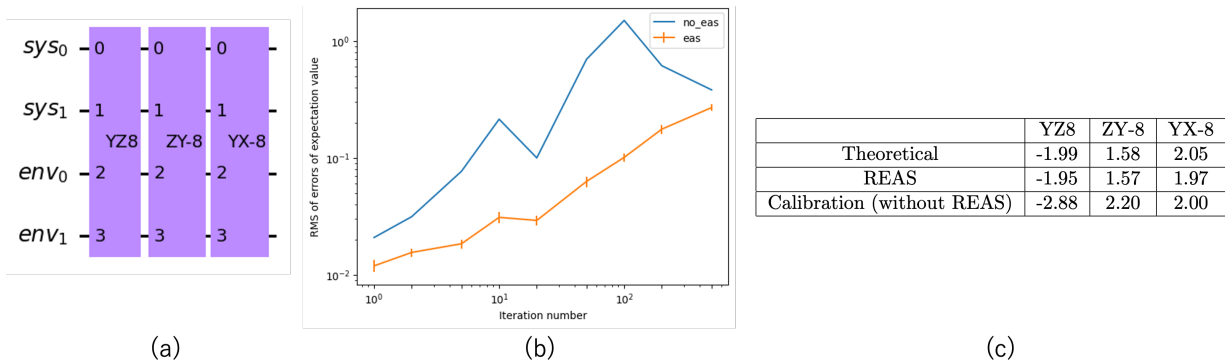
FIG. 1. (a) *Repeated circuit.*—We consider a deep circuit composed of repeated blocks $YZ8$, ZY-8, YZ-8, where, e.g., $YZ8$ corresponds to $e^{-i\frac{\pi}{8}YZ} \otimes I_{\text{env}}$ and similar for the others. (b) *REAS performance in a deep circuit.*—The blocks are iterated a number of times, leading to overall circuits of growing depth, and measure the expectation value of $IZII$. (c) *Shift in rotation angle* $(10^{-3}$ rad).—Here we calculate the shift in rotation angle, i.e., the coherent error, in theory as well as with REAS and with conventional calibration techniques, highlighting the performance of the former.

that the order of Y and Z in the string is reversed, such that Z acts on S_0 and Y acts on S_1, to make it consistent with the qiskit notation). Other strings are interpreted in the same way. In Fig. 1 (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 $O(\sqrt{L})$.

We first perform the REAS calibration to obtain an estimate of the shift in the rotation angle of each gate in $YZ8$, ZY-8, and YX-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 term proportional to the computational Pauli operator is likely to have a large coefficient, in order to clearly demonstrate the effect of calibration.

We then apply the inverse shift of the rotation angles applied according to the estimates obtained in the REAS calibration, before repeating the circuit block 1, 2, 5, 10, 20, 50, 100, 200, and 500 times (so that the total circuit depth can reach 1500 at maximum), and measuring the expectation value of the observable $IZII$ (i.e., Z on S_1). Finally, we compute the RMS (for sampling number 50) between the ideal expectation value and that obtained in the noisy circuit, both with and without REAS.

The graph in Fig. 1 (b) shows that the scaling of the RMS of the difference in the expectation value is nearly of order (Iteration number)$^{1/2}$. Although the theory only predicts that the RMS is upper bounded by a function scaling like (Iteration number)$^{1/2}$, 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 ≈ 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. 1 (c) shows that the value of rotation shift caused by the coherent error is precisely estimated by the REAS calibration with error less than 0.1×10^{-3} rad. In contrast, the estimate of this rotation shift deviates further from the theoretical prediction when using calibration without REAS. This result implies that the usual calibration based on the phase estimation without inserting random Pauli cannot precisely remove the effect of rotation shift and thus REAS calibration is required.

B. Suppression of the First-order Error

In Fig. 2, 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. We randomly generate a circuit with a depth of 1000, which only contains two-qubit Pauli rotations. For $\gamma \in \{0.0001, 0.0003, 0.001, 0.003\}$, we compute the average trace distance between the state obtained after running the error-free circuit and that obtained after running a noisy circuit without/with REAS (sampling number: 4000). The simulation in Fig. 2 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.
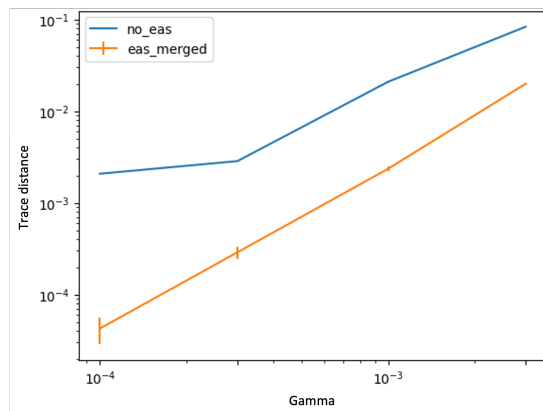
FIG. 2. *Suppression of First-order Error.*—Here we demonstrate that in the limit $\gamma \to 0$, the first order error term vanishes when REAS is applied, in contrast to the case where it is not, as predicted by theory.

C. Effectiveness of Single Pauli Transformation

In Fig. 3, we demonstrate how incorporating the single Pauli transformation helps to reduce decoherence noise. We randomly generated a circuit with a depth of 1000, which contains both single-qubit and two-qubit Pauli gates. For $\gamma \in \{0.0002, 0.0005, 0.001, 0.005\}$, we then computed the average trace distance between the state obtained after running the ideal circuit and that obtained after running the noisy circuit first without REAS, and then with REAS, both with and without the single Pauli transformation (sampling number: 4000). The simulation shown in Fig. 3 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.
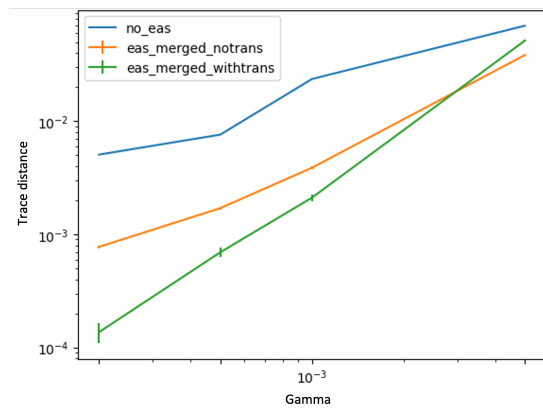

FIG. 3. *Effectiveness of Single Pauli Transformation.*—Here we demonstrate the suppression of decoherence errors by comparing three situations: first, without REAS (blue); second, with REAS but without incorporating the single Pauli transformation procedure (orange); and third, with REAS and single Pauli transformation (green). The behavior improves accordingly.

V. SUMMARY

To summarize, 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. 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/qubit loss will be an important generalization. Secondly, an improved technique of dealing with errors across many qubits, e.g., non-local crosstalk, is required since the performance of our technique suffers in terms of qubit number. 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 decreasing, making non-trivial temporal correlations more prevalent [38–44].

NOTE

Upon completion of this work, we became aware of a related method that generalizes Pauli twirling techniques to non-Clifford circuits [45].

ACKNOWLEDGMENTS

[1] D. Aharonov, M. Ben-Or, R. Impagliazzo, and N. Nisan, Limitations of noisy reversible computation (1996), arXiv:quant-ph/9611028.

[2] P. W. Shor, Algorithms for quantum computation: discrete logarithms and factoring, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (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), arXiv:quant-ph/9508027.

[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), arXiv:1712.03773.

[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, P. G. Kwiat, T. Ladd, J. Lowell, D. Maslov, N. Mason, A. Y. Matsuura, R. McDermott, R. van Meter, A. Miller, J. Orcutt, M. Saffman, M. Schleier-Smith, M. K. Singh, P. Smith, M. Suchara, F. Toudeh-Fallah, M. Turlington, B. Woods, and T. Zhong, *A Roadmap for Quantum Interconnects*, Tech. Rep. (Argonne National Laboratory, Argonne, IL United States, 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), arXiv:quant-ph/9512032.

[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), arXiv:quant-ph/9602019.

[10] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015), arXiv:1302.3428.

[11] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017), arXiv:1612.02058.

[12] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017), arXiv:1611.09301.

[13] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018), arXiv:1712.09271.

[14] B. Koczor, Exponential error suppression for near-term quantum devices, Phys. Rev. X **11**, 031057 (2021), arXiv:2011.05942.

[15] 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), arXiv:2011.07064.

[16] N. Yoshioka, H. Hakoshima, Y. Matsuzaki, Y. Tokunaga, Y. Suzuki, and S. Endo, Generalized quantum subspace expansion, Phys. Rev. Lett. **129**, 020502 (2022), arXiv:2107.02611.

[17] H. Hakoshima, S. Endo, K. Yamamoto, Y. Matsuzaki, and N. Yoshioka, Localized virtual purification (2023), arXiv:2308.13500.

[18] 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), arXiv:2210.00921.

[19] 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 **549**, 242 (2017), arXiv:1704.05018.

[20] 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 **567**, 491 (2019), arXiv:1805.04492.

[21] 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 **618**, 500 (2023).

[22] 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), arXiv:2208.09385.

[23] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, Phys. Rev. Lett. **131**, 210602 (2023), arXiv:2208.09178.

[24] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation (2022), arXiv:2210.11505.

[25] L. Viola, E. Knill, and S. Lloyd, Dynamical decoupling of open quantum systems, Phys. Rev. Lett. **82**, 2417 (1999), arXiv:quant-ph/9809071.

[26] M. B. Hastings, Turning gate synthesis errors into incoherent errors, Quantum Info. Comput. **17**, 488 (2017), arXiv:1612.01011.

[27] E. Campbell, Shorter gate sequences for quantum computing by mixing unitaries, Phys. Rev. A **95**, 042306 (2017), arXiv:1612.02689.

[28] S. Akibue, G. Kato, and S. Tani, Probabilistic state synthesis based on optimal convex approximation, npj Quantum Inf. **10**, 3 (2024), arXiv:2303.10860.

[29] 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), arXiv:quant-ph/0507134.

[30] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, Phys. Rev. Lett. **106**, 180504 (2011), arXiv:1009.3639.

[31] M. R. Geller and Z. Zhou, Efficient error models for fault-tolerant architectures and the Pauli twirling approximation, Phys. Rev. A **88**, 012314 (2013), arXiv:1305.2021.

[32] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016), arXiv:1512.01098.

[33] 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), arXiv:2207.07272.

[34] 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), arXiv:2007.02925.

[35] D. Layden, G. Mazzola, R. V. Mishmash, M. Motta, P. Wocjan, J.-S. Kim, and S. Sheldon, Quantum-enhanced markov chain Monte Carlo, Nature **619**, 282 (2023), arXiv:2203.12497.

[36] 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), arXiv:1502.02677.

[37] C. Arenz, R. Hillier, M. Fraas, and D. Burgarth, Distinguishing decoherence from alternative quantum theories by dynamical decoupling, Phys. Rev. A **92**, 022102 (2015), arXiv:1405.7644.

[38] 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), arXiv:2004.14018.

[39] 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), arXiv:2003.14045.

[40] 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), arXiv:2102.01327.

[41] 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), arXiv:2107.05403.

[42] P. Figueroa-Romero, K. Modi, and M.-H. Hsieh, Towards a general framework of Randomized Benchmarking incorporating non-Markovian Noise, Quantum **6**, 868 (2022), arXiv:2202.11338.

[43] G. White, F. Pollock, L. Hollenberg, K. Modi, and C. Hill, Non-Markovian Quantum Process Tomography, PRX Quantum **3**, 020344 (2022), arXiv:2106.11722.

[44] 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), arXiv:2210.15333.

[45] R. U. Jader dos Santos, Ben Bar, Deeper quantum circuits via pseudo-twirling coherent errors mitigation in non-Clifford gates (2024), arXiv:2401.09040.

[46] T. Odake, H. Kristjánsson, A. Soeda, and M. Murao, Higher-order quantum transformations of Hamiltonian dynamics (2023), arXiv:2303.09788.

Appendix A: Suppression of Coherent Errors by REAS

Here, we demonstrate how the REAS protocol can deal with coherent errors. For simplicity, suppose that the gate sequence that one wishes to implement in the computation does not involve parallel application of Pauli rotation gates, i.e., it can be written as

$$\prod_{j=1}^{L} e^{-i\sigma_j\theta_j} \in \mathcal{L}(\mathcal{H}) \tag{A1}$$

using single- or two-qubit Pauli operators $\sigma_1, \ldots, \sigma_L$ and rotation angles $\theta_1, \ldots, \theta_N$. Again, we abbreviate the tensor product with the identity on the systems that undergo trivial dynamics. Suppose that coherent errors affect each Pauli rotation $e^{-is\sigma_j\theta_j}$ ($s \in \{+1, -1\}$) so that the actual noisy gate implemented is expressed as $e^{-is\sigma_j\theta_j}[I + i\epsilon C_{s,j}]$ for some unitary $I + i\epsilon C_{s,j}$ where $\epsilon > 0$ and $\|C_{s,j}\|_{\mathrm{op}} \leq 1$ for all (s, j). Here, we assume that the same error occurs on the same type of gate at any time, namely, $C_{s,j} = C_{s,k}$ if $(\sigma_j, \theta_j) = (\sigma_k, \theta_k)$, which is rather natural and required to enable calibration.

Since each $C_{s,j}$ is defined on $\mathcal{H}$, it can be decomposed in terms of a linear combination of n-qubit Pauli operators as

$$C_{s,j} =: \sum_{\vec{v} \in \{0,1,2,3\}^n} \alpha_{\vec{v},j}(s)\sigma_{\vec{v}} \tag{A2}$$

using complex numbers $\alpha_{\vec{v}}(s) \in \mathbb{C}$. Importantly, note that $\alpha_{\vec{v},j}(s)$ can be further decomposed into a sum of s-dependent part and s-independent part as $\alpha_{\vec{v},j}(s) =: \overline{\alpha}_{\vec{v},j} + s\Delta\alpha_{\vec{v},j}$, where $\overline{\alpha}_{\vec{v},j} := \frac{1}{2}[\alpha_{\vec{v},j}(1) + \alpha_{\vec{v},j}(-1)]$ and $\Delta\alpha_{\vec{v},j} := \frac{1}{2}[\alpha_{\vec{v},j}(1) - \alpha_{\vec{v},j}(-1)]$. We will frequently make use of this technique throughout this analysis.

Since the operator $[I + i\epsilon C_{s,j}]$ is unitary, it follows that

$$\left[I + i\epsilon \sum_{\vec{v}} \alpha_{\vec{v},j}(s)\sigma_{\vec{v}}\right]\left[I - i\epsilon \sum_{\vec{v}} \alpha^*_{\vec{v},j}(s)\sigma_{\vec{v}}\right] = I + i\epsilon\left(\sum_{\vec{v}}(\alpha_{\vec{v},j}(s) - \alpha^*_{\vec{v},j}(s))\right) + O(\epsilon^2) = I, \tag{A3}$$

which implies that the 0th order ϵ term $\alpha^{(0)}_{\vec{v},j}$ of $\alpha_{\vec{v},j}$ is real. Moreover, we can take the 0th order ϵ term $\alpha_{\vec{0},j}$, namely $\alpha^{(0)}_{\vec{0},j}$, the dominant term of the coefficient of I, to be 0 without loss of generality. This is because it can be canceled via an appropriate global phase, as follows

$$\frac{1 - i\epsilon\alpha^{(0)}_{\vec{0},j}}{\sqrt{1 + (\epsilon\alpha^{(0)}_{\vec{0},j})^2}}\left[I + i\epsilon\sum_{\vec{v}}(\alpha^{(0)}_{\vec{v},j}(s) + O(\epsilon))\sigma_{\vec{v}}\right] = \frac{1}{\sqrt{1 + (\epsilon\alpha^{(0)}_{\vec{0},j})^2}}\left[I + i\epsilon\left\{O(\epsilon)I + \sum_{\vec{v}\neq\vec{0}}\left(\alpha^{(0)}_{\vec{v},j}(s) + O(\epsilon)\right)\sigma_{\vec{v}}\right\}\right]$$

$$= (1 - O(\epsilon^2))\left[I + i\epsilon\left\{O(\epsilon)I + \sum_{\vec{v}\neq\vec{0}}\left(\alpha^{(0)}_{\vec{v},j}(s) + O(\epsilon)\right)\sigma_{\vec{v}}\right\}\right]$$

$$=: I + i\epsilon\sum_{\vec{v}\neq\vec{0}}\alpha^{(0)}_{\vec{v},j}(s)\sigma_{\vec{v}} + O(\epsilon^2). \tag{A4}$$

Thus, we assume that $C_{s,j}$ is defined such that $\alpha^{(0)}_{\vec{0},j}(s) = 0$ for all (s, j).

Now, we move on to the error analysis of the gate sequence obtained by inserting random Pauli gates via REAS into Eq. (A1). This leads to

$$\sigma_{\vec{v}_L}e^{-is_L\sigma_L\theta_L}\prod_{j=1}^{L-1}\left[\sigma_{\vec{v}_{j+1:j}}e^{-is_j\sigma_j\theta_j}\right]\sigma_{\vec{v}_1}, \tag{A5}$$

where $\vec{v}_1, \ldots, \vec{v}_L$ are uniformly randomly chosen vectors. For simplicity, here we have assumed that a random Pauli gate is inserted between all single Pauli rotation gates. Recall that $s_j \in \{+1, -1\}$ is defined as

$$s_j := \begin{cases} 1 & (\sigma_j \text{ and } \sigma_{\vec{v}_j} \text{ commutes}) \\ -1 & (\sigma_j \text{ and } \sigma_{\vec{v}_j} \text{ anti-commutes}) \end{cases}. \tag{A6}$$

When such a gate sequence is applied on an arbitrary input state $\rho \in \mathcal{L}(\mathcal{H})$, the result is

$$\left(\frac{1}{4^n}\right)^L \sum_{\vec{v}_1,\ldots,\vec{v}_L} \left[\sigma_{\vec{v}_L} e^{-is_L \sigma_L \theta_L} \overleftarrow{\prod_{j=1}^{L-1}} \left[\sigma_{\vec{v}_{j+1:j}} e^{-is_j \sigma_j \theta_j}\right] \sigma_{\vec{v}_1}\right] \rho \left[\sigma_{\vec{v}_1} \overrightarrow{\prod_{j=1}^{L-1}} \left[e^{is_j \sigma_j \theta_j} \sigma_{\vec{v}_{j+1:j}}\right] e^{is_L \sigma_L \theta_L} \sigma_{\vec{v}_L}\right]$$

$$= \left(\frac{1}{4^n}\right)^L \sum_{\vec{v}_1,\ldots,\vec{v}_L} \left[\sigma_{\vec{v}_L} e^{-is_L \sigma_L \theta_L} \overleftarrow{\prod_{j=1}^{L-1}} \left[\sigma_{\vec{v}_{j+1}} \sigma_{\vec{v}_j} e^{-is_j \sigma_j \theta_j}\right] \sigma_{\vec{v}_1}\right] \rho \left[\sigma_{\vec{v}_1} \overrightarrow{\prod_{j=1}^{L-1}} \left[e^{is_j \sigma_j \theta_j} \sigma_{\vec{v}_j} \sigma_{\vec{v}_{j+1}}\right] e^{is_L \sigma_L \theta_L} \sigma_{\vec{v}_L}\right]. \tag{A7}$$

Here, we use notation $\overleftarrow{\prod}$ and $\overrightarrow{\prod}$ to distinguish the order of applying the gate sequences, defined such that

$$\overleftarrow{\prod_{j=1}^{L}} M_j = M_L \cdots M_1 \qquad \text{and} \qquad \overrightarrow{\prod_{j=1}^{L}} M_j = M_1 \cdots M_L \tag{A8}$$

for linear operators $M_1,\ldots,M_L$. It is straightforward to verify that when no error terms appears, Eq. (A7) reduces to the ideal state transformation

$$\left(\overleftarrow{\prod_{j=1}^{L}} e^{-i\sigma_j \theta_j}\right) \rho \left(\overrightarrow{\prod_{j=1}^{L}} e^{i\sigma_j \theta_j}\right), \tag{A9}$$

which can be seen by moving the inner Pauli operators $\sigma_{\vec{v}_j}$ outward, making them pass through $e^{\pm is_j \sigma_j \theta_j}$ and thereby changing $e^{\pm is_j \sigma_j \theta_j}$ to $e^{\pm i\sigma_j \theta_j}$, and subsequently canceling with the corresponding outer $\sigma_{\vec{v}_j}$ operator.

We now prove that, as claimed in the main text, coherent errors lead to Eq. (A7) being transformed as

$$\left(\overleftarrow{\prod_{j=1}^{L}} e^{-i\sigma_j (\theta_j - \epsilon\Delta\theta_j)}\right) \rho \left(\overrightarrow{\prod_{j=1}^{L}} e^{i\sigma_j (\theta_j - \epsilon\Delta\theta_j)}\right) + O(\epsilon + \epsilon^2 L), \tag{A10}$$

where $\epsilon\Delta\theta_j := \tan^{-1}(\epsilon\Delta\alpha^{(0)}_{\vec{w}_j,j})$ where $\sigma_{\vec{w}_j} := \sigma_j$. Our technique of erasing errors is based on the following equations regarding averaging over an n-qubit Pauli operator σ

$$\left(\frac{1}{4^n}\right) \sum_{\vec{v}\in\{0,1,2,3\}^n} \sigma_{\vec{v}} \sigma \sigma_{\vec{v}} = \begin{cases} I & \text{for } \sigma = I \\ 0 & \text{for } \sigma \neq I \end{cases} \tag{A11}$$

and

$$\left(\frac{1}{4^n}\right) \sum_{\vec{v}\in\{0,1,2,3\}^n} \sigma_{\vec{v}} s(\sigma_{\vec{v}},\sigma') \sigma \sigma_{\vec{v}} = \begin{cases} \sigma' & \text{for } \sigma = \sigma' \\ 0 & \text{for } \sigma \neq \sigma' \end{cases} \tag{A12}$$

where $s(\sigma_{\vec{v}},\sigma')$ is defined as

$$s(\sigma_{\vec{v}},\sigma') := \begin{cases} 1 & (\sigma_{\vec{v}} \text{ and } \sigma' \text{ commutes}) \\ -1 & (\sigma_{\vec{v}} \text{ and } \sigma' \text{ ant-commutes}) \end{cases}. \tag{A13}$$

With this, the error term $i\epsilon C_{s_{j'},j'}$ transforms as

$$\frac{1}{4^n} \sum_{\vec{v}_{j'}} \sigma_{\vec{v}_{j'}} i\epsilon C_{s_{j'},j'} \sigma_{\vec{v}_{j'}} = \frac{1}{4^n} \sum_{\vec{v}_{j'}} \sigma_{\vec{v}_{j'}} \left[\sum_{\vec{u}} i\epsilon(\overline{\alpha}_{\vec{u},j'} + s_{j'}\Delta\alpha_{\vec{u},j'})\sigma_{\vec{u}}\right] \sigma_{\vec{v}_{j'}} = O(\epsilon^2) I + i\epsilon\Delta\alpha_{\vec{w}_{j'},j'}\sigma_{j'} \tag{A14}$$

where $\sigma_{\vec{w}_{j'}} = \sigma_{j'}$ by definition. This expression can be simplified by absorbing the first-order ϵ part in $i\epsilon s_{j'}\Delta\alpha_{\vec{w}_{j'},j'}\sigma_{j'}$

to a shift in the rotation angle of $e^{-is_{j'}\sigma_{j'}\theta_{j'}}$ as follows

$$e^{-is_{j'}\sigma_{j'}\theta_{j'}}[I + i\epsilon C_{s_{j'},j'}] =: g\left(e^{-is_{j'}\sigma_{j'}\theta_{j'}}\right)$$

$$= e^{-is_{j'}\sigma_{j'}\theta_{j'}}e^{i\epsilon s_{j'}\Delta\theta_{j'}\sigma_{j'}}\frac{1 - is_{j'}\epsilon\Delta\alpha^{(0)}_{\vec{w}_{j'},j'}}{\sqrt{1 + (\epsilon\Delta\alpha^{(0)}_{\vec{w}_{j'},j'})^2}}\left[I + i\epsilon\sum_{\vec{u}\neq\vec{0}}(\overline{\alpha}_{\vec{u},j'} + s_{j'}\Delta\alpha_{\vec{u},j'})\sigma_{\vec{u}} + O(\epsilon^2)\right]$$

$$= e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}\left[1 - O(\epsilon^2)\right]\left[I + i\epsilon\sum_{\vec{u}\notin\{\vec{0},\vec{w}_{j'}\}}(\overline{\alpha}_{\vec{u},j'} + s_{j'}\Delta\alpha_{\vec{u},j'})\sigma_{\vec{u}} + i\epsilon\overline{\alpha}_{\vec{w}_{j'},j'}\sigma_{j'} + O(\epsilon^2)\right]$$

$$= e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}\left[I + i\epsilon\sum_{\vec{u}\notin\{\vec{0},\vec{w}_{j'}\}}(\overline{\alpha}_{\vec{u},j'} + s_{j'}\Delta\alpha_{\vec{u},j'})\sigma_{\vec{u}} + i\epsilon\overline{\alpha}_{\vec{w}_{j'},j'}\sigma_{j'} + O(\epsilon^2)\right]$$

$$=: e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}[I + i\epsilon D_{s_{j'},j'}] =: g'\left(e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}\right). \tag{A15}$$

Thus, the error term $i\epsilon D_{s_{j'},j'}$ no longer contains the first-order ϵ term proportional to $s_{j'}\sigma_{j'}$. Using this result, Eq. (A7) with coherent errors can be rewritten as

$$\left(\frac{1}{4^n}\right)^L\sum_{\vec{v}_1,\ldots,\vec{v}_L}\left[g(\sigma_{\vec{v}_L})g(e^{-is_L\sigma_L\theta_L})\overset{\overleftarrow{L-1}}{\underset{j=1}{\prod}}\left[g(\sigma_{\vec{v}_{j+1:j}})g(e^{-is_j\sigma_j\theta_j})\right]g(\sigma_{\vec{v}_1})\right]\rho[\cdots]$$

$$= \left(\frac{1}{4^n}\right)^L\sum_{\vec{v}_1,\ldots,\vec{v}_L}\left[g(\sigma_{\vec{v}_L})g'(e^{-is_L\sigma_L(\theta_L - \epsilon\Delta\theta_L)})\overset{\overleftarrow{L-1}}{\underset{j=1}{\prod}}\left[g(\sigma_{\vec{v}_{j+1:j}})g'(e^{-is_j\sigma_j(\theta_j - \Delta\theta_j)})\right]g(\sigma_{\vec{v}_1})\right]\rho[\cdots] \tag{A16}$$

where the error function g on inserted random Pauli operators $\sigma_{\vec{v}}$ is defined as $g(\sigma_{\vec{v}}) =: \sigma_{\vec{v}}[i + i\epsilon\widetilde{C}_{\vec{v}}]$ such that $i\epsilon\widetilde{C}_{\vec{v}}$ has no $O(\epsilon)$ term proportional to I. Here, the notation $[\ldots]$ on the right of ρ represents the Hermitian conjugate of the corresponding product on the left; we will continue to use similar notation below.

We move to analyze the suppression of errors. Consider a single error term on a particular Pauli rotation gate $e^{-is_{j'}\sigma_{j'}\theta_{j'}}$ in the computation, which leads to an expression of the form[4]

$$\left(\frac{1}{4^n}\right)^L\sum_{\vec{v}_1,\ldots,\vec{v}_L}\left[\sigma_{\vec{v}_L}[\cdots]\sigma_{\vec{v}_{j'+1},j'}(e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}i\epsilon D_{s_{j'},j'})\sigma_{\vec{v}_{j'-1:j'}}[\cdots]\sigma_{\vec{v}_1}\right]\rho[\sigma_{\vec{v}_1}[\cdots]\sigma_{\vec{v}_L}]$$

$$= \left(\frac{1}{4^n}\right)^L\sum_{\vec{v}_1,\ldots,\vec{v}_L}\left[\sigma_{\vec{v}_L}[\cdots]\sigma_{\vec{v}_{j'+1}}\sigma_{\vec{v}_{j'}}(e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}i\epsilon D_{s_{j'},j'})\sigma_{\vec{v}_{j'}}\sigma_{\vec{v}_{j'-1}}[\cdots]\sigma_{\vec{v}_1}\right]\rho[\sigma_{\vec{v}_1}[\cdots]\sigma_{\vec{v}_L}]. \tag{A17}$$

All pairs of inserted Pauli operators $\sigma_{\vec{v}_j}$ other than $\sigma_{\vec{v}_{j'}}$ that sandwich $i\epsilon D_{s_{j'},j'}$ can pass through the operator $e^{\pm is_j\sigma_j(\theta_j - \epsilon\Delta\theta_j)}$ and cancel with each other. Thus, Eq. (A17) can be rewritten as

$$\frac{1}{4}\sum_{\vec{v}_{j'}}\left[\left(\overset{\overleftarrow{L}}{\underset{j=j'+1}{\prod}}e^{-i\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right)\sigma_{\vec{v}_{j'}}e^{-is_{j'}\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}i\epsilon D_{s_{j'},j'}\sigma_{\vec{v}_{j'}}\left(\overset{\overleftarrow{j'-1}}{\underset{j=1}{\prod}}e^{-i\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right)\right]\rho\left[\overset{\overrightarrow{L}}{\underset{j=1}{\prod}}e^{is_j\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right]$$

$$= \left[\left(\overset{\overleftarrow{L}}{\underset{j=j'+1}{\prod}}e^{-i\sigma_j\theta_j}\right)e^{-i\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}\left(\frac{1}{4^n}\sum_{\vec{v}_{j'}}\sigma_{\vec{v}_{j'}}i\epsilon D_{s_{j'},j'}\sigma_{\vec{v}_{j'}}\right)\left(\overset{\overleftarrow{j'-1}}{\underset{j=1}{\prod}}e^{-i\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right)\right]\rho\left[\overset{\overrightarrow{L}}{\underset{j=1}{\prod}}e^{is_j\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right]$$

$$= \left[\left(\overset{\overleftarrow{L}}{\underset{j=j'+1}{\prod}}e^{-i\sigma_j\theta_j}\right)e^{-i\sigma_{j'}(\theta_{j'} - \epsilon\Delta\theta_{j'})}\left(O(\epsilon^2)I + O(\epsilon^2)\sigma_{\vec{w}_{j'}}\right)\left(\overset{\overleftarrow{j'-1}}{\underset{j=1}{\prod}}e^{-i\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right)\right]\rho\left[\overset{\overrightarrow{L}}{\underset{j=1}{\prod}}e^{is_j\sigma_j(\theta_j - \epsilon\Delta\theta_j)}\right]$$

$$= O(\epsilon^2). \tag{A18}$$

---

[4] We will only explicitly show the calculation for the error term $i\epsilon D_{s_{j'},j'}$ in question appearing on the left of ρ; since the term on the right is simply its Hermitian conjugate, the analysis there is be similar.

Next, we consider the case where a single error term in Eq. (A16) occurs on one of the inserted Pauli operators $\sigma_{\vec{v}_{j'+1:j'}}$. This leads to Eq. (A16) being expressed as

$$
\left(\frac{1}{4^n}\right)^L \sum_{\vec{v}_1,\ldots,\vec{v}_L} \left[\sigma_{\vec{v}_L}[\cdots]\sigma_{\vec{v}_{j'+2,j'+1}}e^{-is_{j'+1}\sigma_{j'+1}(\theta_{j'+1}-\epsilon\Delta\theta_{j'+1})}(\sigma_{\vec{v}_{j'+1:j'}}i\epsilon\widetilde{C}_{\vec{v}_{j'+1:j'}})e^{-is_{j'}\sigma_{j'}(\theta_{j'}-\epsilon\Delta\theta_{j'})}\sigma_{\vec{v}_{j':j'-1}}[\cdots]\sigma_{\vec{v}_1}\right]\rho[\cdots]
$$

$$
=\left(\frac{1}{4^n}\right)^L \sum_{\vec{v}_1,\ldots,\vec{v}_L} \left[\sigma_{\vec{v}_L}[\cdots]\sigma_{\vec{v}_{j'+2}}\sigma_{\vec{v}_{j'+1}}e^{-is_{j'+1}\sigma_{j'+1}(\theta_{j'+1}-\epsilon\Delta\theta_{j'+1})}(\sigma_{\vec{v}_{j'+1}}\sigma_{\vec{v}_{j'}}i\epsilon\widetilde{C}_{\vec{v}_{j'+1:j'}})e^{-is_{j'}\sigma_{j'}(\theta_{j'}-\epsilon\Delta\theta_{j'})}\sigma_{\vec{v}_{j'}}\sigma_{\vec{v}_{j'-1}}[\cdots]\sigma_{\vec{v}_1}\right]\rho[\cdots]
$$

$$
=\left(\frac{1}{4^n}\right)^2 \sum_{\vec{v}_{j'+1},\vec{v}_{j'}} \left(\prod_{j=j'+1}^{L} e^{-i\sigma_j(\theta_j-\epsilon\Delta\theta_j)}\right)\sigma_{\vec{v}_{j'}}i\epsilon\widetilde{C}_{\vec{v}_{j'+1:j'}}\sigma_{\vec{v}_{j'}}\left(\prod_{j=1}^{j'} e^{-i\sigma_j(\theta_j-\epsilon\Delta\theta_j)}\right)\rho[\cdots]
$$

$$
=\left(\prod_{j=j'+1}^{L} e^{-i\sigma_j(\theta_j-\epsilon\Delta\theta_j)}\right)\left\{\frac{1}{4^n}\sum_{\vec{v}_{j'}}\sigma_{\vec{v}_{j'}}\left(\frac{1}{4^n}\sum_{\vec{v}_{j'+1}}\widetilde{C}_{\vec{v}_{j'+1:j'}}\right)\sigma_{\vec{v}_{j'}}\right\}\left(\prod_{j=1}^{j'} e^{-i\sigma_j(\theta_j-\epsilon\Delta\theta_j)}\right)\rho[\cdots], \tag{A19}
$$

which is also $O(\epsilon^2)$ since $\sum_{\vec{v}_{j'+1}}\widetilde{C}_{\vec{v}_{j'+1:j'}} = \sum_{\vec{u}}\widetilde{C}_{\vec{u}}$ is independent of $\vec{v}_{j'}$ and has only $O(\epsilon^2)$ term proportional to I.

Finally, let us consider how many first- and second-order ϵ terms can possibly appear in Eq. (A16). The first-order ϵ terms can only appear in the expansion of Eq. (A16) when a single error term is chosen at either $\sigma_{\vec{v}_1}$ or $\sigma_{\vec{v}_L}$. Thus, the first-order ϵ terms in Eq. (A16) do not scale in terms of L and behave like $O(\epsilon)$. The second-order ϵ terms can originate from any term in the expansion of Eq. (A16) where a single error term occurs at a single Pauli rotation in the computation $e^{\pm is_j\sigma_j(\theta_j-\epsilon\Delta\theta_j)}$ or at an inserted Pauli operator $\sigma_{\vec{v}_{j:j+1}}$; the sum of such terms is $O(\epsilon^2 L)$. Technically speaking, a second-order ϵ term can also possibly appear when two error terms occur in Eq. (A16), but for such cases to have a second-order ϵ term, the errors must appear in close locations; this means that such cases are largely irrelevant in practice. For example, when the error occurs at $e^{\pm is_j\sigma_j(\theta_j-\epsilon\Delta\theta_j)}$ and $\sigma_{\vec{v}_{k:k+1}}$ with either $j-k\geq 2$ or $k-j\geq 1$, it can be easily shown that the error terms are averaged over random Pauli operators and therefore no more second-order ϵ^2 terms remains. Thus, the total sum of these terms in the expansion of Eq. (A16) will also be $O(\epsilon^2 L)$.

In general, ϵ^n terms in Eq. (A16) can be shown to be $O(\epsilon^n L^{\lfloor\frac{n}{2}\rfloor})$, where $\lfloor a\rfloor$ denotes the largest integer not exceeding a. This follows from the fact that the first-order ϵ term in the error can remain only if there exists another error term in a nearby gate. Using this fact, the claim made in Eq. (A10) is shown. $\qquad\square$

Appendix B: Root Mean Squared Error Bounds

To obtain an upper bound of the RMS of the error of expectation values in terms of an arbitrary observable M, the following statement can be used.

Lemma 1. *For an arbitrary unitary operation defined by $\mathcal{U}(\rho) := U\rho U^\dagger$ with a unitary operator U and a density operator ρ on a Hilbert space $\mathcal{H}$, if a set of deterministic quantum operations (completely-positive trace-preserving maps) $\mathcal{F}_j : \mathcal{L}(\mathcal{H})\to\mathcal{L}(\mathcal{H})$ and a probability distribution $\{p_j\}$ satisfies*

$$
\left\|\mathcal{U}-\sum_j p_j\mathcal{F}_j\right\|_\diamond \leq \Delta
$$

for some $\Delta > 0$, the variance of the averaged operation $\sum_j p_j\mathcal{F}_j$ is upper bounded by

$$
\sup_{\substack{\dim(\mathcal{H}') \\ |\psi\rangle;\||\psi\rangle\|=1}} \sum_j p_j\|\mathcal{U}\otimes\mathcal{I}(|\psi\rangle\langle\psi|)-\mathcal{F}_j\otimes\mathcal{I}(|\psi\rangle\langle\psi|)\|_1^2 \leq 2\Delta, \tag{B1}
$$

where $\mathcal{I}$ is the identity operation on a Hilbert space $\mathcal{H}'$ and $|\psi\rangle$ is a pure state on $\mathcal{H}\otimes\mathcal{H}'$.

The proof is given in App. B of Ref. [46]. For our case, U is the error-free unitary that one desires to implement, $j := (\vec{v}_1,\ldots,\vec{v}_N)$ is the tuple of all random vectors chosen in the REAS scheme, and p_j and $\mathcal{F}_j$ are the probability that j and the quantum operation on $\mathcal{H}$ which is applied on the state (including the final partial trace of $\mathcal{H}_{\mathrm{env}}$) when j is chosen, respectively.

Consider $\mathcal{E}_{\mathrm{shifted}}$ correspond to a gate implemented with a shift in the rotation angle, i.e., the r.h.s. of Eq. (9). Since the result regarding such coherent errors implies that $\|\mathcal{E}_{\mathrm{shifted}}-\mathcal{E}'\|_\diamond = C(\epsilon+N\epsilon^2)$ for constant $C > 0$ and $\mathcal{E}' = \sum_j p_j\mathcal{F}_j$, where $p_j := (1/4)^{nN}$ and $\mathcal{F}_j(\rho) := \mathrm{tr}_{\mathrm{env}}[\mathcal{E}'_j(\rho\otimes\rho_{\mathrm{env}})]$, Lem. 1 leads to the following.

Corollary 1. *For an arbitrary observable A on $\mathcal{H}$ such that $\|A\|_{\mathrm{op}} \leq 1$, the following equation holds:*

$$\sup_{|\psi\rangle; \||\psi\rangle\|=1} \sum_j p_j [\mathrm{tr}\{A\mathcal{E}_{\mathrm{shifted}}(|\psi\rangle\langle\psi|)\} - \mathrm{tr}\{A\mathcal{F}_j(|\psi\rangle\langle\psi|)\}]^2 \leq 2C(\epsilon + N\epsilon^2) \tag{B2}$$

where $j := (\vec{v}_1, \ldots, \vec{v}_N)$, $p_j = (1/4)^{nN}$, and $\mathcal{F}_j(\rho) := \mathrm{tr}_{\mathrm{env}}[\mathcal{E}'_j(\rho \otimes \rho_{\mathrm{env}})]$.

This means that the RMS of the difference between the error of expectation value is bounded above by $\sqrt{2C(\epsilon + N\epsilon^2)}$ which approaches to $\sqrt{2C}\epsilon\sqrt{N}$ in the regime $N \gg 1/\epsilon$.

Appendix C: Class of Decoherence Errors that can be Suppressed

We now move to demonstrate the more general class of decoherence errors that can be suppressed when our REAS method is supplemented by the single Pauli transformation introduced in Sec. III B. Specifically, we show that this protocol can suppress errors that are first-order in ϵ for a particular class of open system-environment dynamics, which includes depolarizing, amplitude-damping, and decoherence models. We assume that the interaction Hamiltonian involves at most two-body interactions.

In attempting to extend the error-suppressing method for the case of coherent errors to the open setting of decoherence, one immediately runs into problem that some errors terms on the computational Pauli gates $e^{-is_j\sigma_j\theta_j}$ cannot be dealt with. In particular, any term of the form $s_j\sigma_j \otimes M_{\mathrm{env}}$, where $M_{\mathrm{env}} \neq I_{\mathrm{env}} \in \mathcal{H}_{\mathrm{env}}$, cannot be absorbed into the shift of the rotation angle θ_j, since σ_j is coupled with M_{env}, and therefore cannot be erased by averaging over random inserted Pauli operators as in Eqs. (A11) and (A12). This inability to suppress such terms can introduce unwanted $O(\epsilon L)$ errors in Eq. (A10), since such first-order ϵ terms can appear in all Pauli computation gates throughout the circuit.

We circumvent this issue by modifying the computational gate sequence implementing $e^{-is_j\sigma_j\theta_j}$ in such a way that the error terms proportional to $s_j\sigma_j \otimes M_{\mathrm{env}}$ for $M_{\mathrm{env}} \neq I$ is at most $O(\epsilon^2)$. This *single Pauli transformation* consists of decomposing any $e^{-is\sigma\theta}$ into (a) a two-qubit Pauli gate dependent on the rotation direction $s \in \{+1, -1\}$ and (b) another Pauli gate (which can be either a one- or two-qubit Pauli gate) that is independent of s. Intuitively, the trick works to suppress error terms that are first-order in ϵ since these only appear when an error occurs on a single gate in the computation $e^{-is_j\sigma_j\theta_j}$ and: (a) the s-dependent two-qubit Pauli gate does not produce harmful first-order ϵ terms; and (b) the other s-independent Pauli gate does not produce any error term that depends upon s.

First, we show that any gate $e^{-is\sigma_{\vec{v}_0}\theta}$ corresponding to a two-qubit Pauli operator involves only $O(\epsilon^2)$ terms proportional to $s\sigma_{\vec{v}_0} \otimes M_{\mathrm{env}}$ for $M_{\mathrm{env}} \neq I$. Note that we always take the Hamiltonian to be traceless without loss of generality (since any trace part only contributes to a global phase).

Lemma 2. *Consider a time interval $t \in [0, T]$, any two-qubit Pauli operator $\sigma_{\vec{v}_0} \in \mathcal{L}(\mathcal{H})$, a time-dependent interaction Hamiltonian $\epsilon H'(t) \in \mathcal{L}(\mathcal{H} \otimes \mathcal{H}_{\mathrm{env}})$ which contains at most two-body terms and has operator norm $O(\epsilon)$, a background Hamiltonian H_{env} of the environment, and a function $\Omega(t)$ such that $\int_0^T \mathrm{d}\tau\Omega(\tau) = \theta$. In this setting, the total Hamiltonian $s\Omega(t)\sigma_{\vec{v}_0} \otimes I_{\mathrm{env}} + I \otimes H_{\mathrm{env}} + \epsilon H'(t)$ gives rise to the unitary operator*

$$\mathcal{T}\exp\left[-i\int_0^t \mathrm{d}\tau(s\Omega(t)\sigma_{\vec{v}_0} \otimes I_{\mathrm{env}} + I \otimes H_{\mathrm{env}} + \epsilon H'(t))\right] =: \left(e^{-is\sigma_{\vec{v}_0}\theta} \otimes e^{-iH_{\mathrm{env}}T}\right)[I + i\epsilon C_{\vec{v}_0}]$$

$$=: \left(e^{-is\sigma_{\vec{v}_0}\theta} \otimes e^{-iH_{\mathrm{env}}T}\right)\left[I + i\epsilon \sum_{\vec{w} \in \{0,1,2,3\}^n} \sigma_{\vec{w}} \otimes D_{\vec{v}_0;\vec{w}}(s)\right], \tag{C1}$$

where $\epsilon D_{\vec{v}_0;\vec{v}_0}(s) = \epsilon\alpha I + O(\epsilon^2)$ for some α.

Note that the assumption that the operator norm of the interaction Hamiltonian is $O(\epsilon)$ is consistent with the assumption that the error term $C_{\vec{v}_0}$ has operator norm of the same order.

Proof: For general time-dependent Hamiltonians $H_1(t)$ and $H_2(t)$, the Hamiltonian dynamics generated by $H_1 + H_2$ can be written as

$$\mathcal{T}\exp\left[-i\int_0^t \mathrm{d}\tau\{H_1(\tau) + H_2(\tau)\}\right] = U_1(t)\mathcal{T}\exp\left[-i\int_0^t \mathrm{d}\tau U_1(\tau)^\dagger H_2(\tau)U_1(\tau)\right], \tag{C2}$$

where

$$U_1(t) := \mathcal{T}\exp\left[-i\int_0^t \mathrm{d}\tau H_1(\tau)\right]. \tag{C3}$$

Substituting $H_1 \leftarrow s\Omega\sigma_{\vec{v}_0} + H_{\text{env}}$ and $H_2 \leftarrow \epsilon H'$ into the above expression and taking the upper limit of the integration up to time T leads to

$$\mathcal{T}\exp\left[-i\int_0^T \mathrm{d}\tau(s\Omega(\tau)\sigma_{\vec{v}_0} \otimes I_{\text{env}} + I \otimes H_{\text{env}} + \epsilon H'(\tau))\right] = \left(e^{-is\theta\sigma_{\vec{v}_0}} \otimes e^{-iH_{\text{env}}T}\right)\mathcal{T}\exp\left[-i\epsilon\int_0^T \mathrm{d}\tau U(\tau)^\dagger H'(\tau)U(\tau)\right]$$

$$= \left(e^{-is\theta\sigma_{\vec{v}_0}} \otimes e^{-iH_{\text{env}}T}\right)\left[I - i\epsilon\int_0^T \mathrm{d}\tau U(\tau)^\dagger H'(\tau)U(\tau) + O(\epsilon^2)\right], \tag{C4}$$

where we have defined

$$U(t) := e^{-is\omega(t)\sigma_{\vec{v}_0}} \otimes e^{-iH_{\text{env}}t} \qquad \omega(t) := \int_0^t \mathrm{d}\tau\,\Omega(\tau). \tag{C5}$$

Now, by decomposing the interaction Hamiltonian $H'(t)$ as

$$H'(t) := \sum_{\vec{w}} \sigma_{\vec{w}} \otimes H'_{\vec{w}}(t), \tag{C6}$$

Eq. (C4) can be rewritten as

$$\left(e^{-is\theta\sigma_{\vec{v}_0}} \otimes e^{-iH_{\text{env}}T}\right)\left[I - i\epsilon\int_0^T \mathrm{d}\tau U(\tau)^\dagger\left(\sum_{\vec{w}}\sigma_{\vec{w}} \otimes H'_{\vec{w}}(t)\right)U(\tau) + O(\epsilon^2)\right]$$

$$= \left(e^{-is\theta\sigma_{\vec{v}_0}} \otimes e^{-iH_{\text{env}}T}\right)\left[I - i\epsilon\int_0^T \mathrm{d}\tau\left(\sum_{\vec{w}}(e^{is\omega(t)\sigma_{\vec{v}_0}}\sigma_{\vec{w}}e^{-is\omega(t)\sigma_{\vec{v}_0}}) \otimes (e^{iH_{\text{env}}t}H'_{\vec{w}}(t)e^{-iH_{\text{env}}t})\right) + O(\epsilon^2)\right]. \tag{C7}$$

Lastly, note that $e^{is\omega(t)\sigma_{\vec{v}_0}}\sigma_{\vec{w}}e^{-is\omega(t)\sigma_{\vec{v}_0}}$ contains a $\sigma_{\vec{v}_0}$ component if and only if $\vec{w} = \vec{v}_0$, which can be verified using

$$\mathrm{tr}\left[\sigma_{\vec{v}_0}(e^{is\omega(t)\sigma_{\vec{v}_0}}\sigma_{\vec{w}}e^{-is\omega(t)\sigma_{\vec{v}_0}})\right] = \mathrm{tr}\left[\sigma_{\vec{v}_0}\sigma_{\vec{w}}\right]. \tag{C8}$$

This finally leads us to be able to write $D_{\vec{v}_0;\vec{v}_0}$ [see Eq. (C1)] as

$$D_{\vec{v}_0;\vec{v}_0} = -\int_0^T \mathrm{d}\tau(e^{iH_{\text{env}}t}H'_{\vec{v}_0}(t)e^{-iH_{\text{env}}t}) + O(\epsilon^2). \tag{C9}$$

Here, $H'_{\vec{v}_0}(t)$ is proportional to the identity only in the case where $\sigma_{\vec{v}_0}$ is a two-qubit Pauli gate. This follows by assumption, since the total term $\sigma_{\vec{v}_0} \otimes H'_{\vec{v}_0}(t)$ being of weight at most two implies that $H'_{\vec{v}_0}(t)$ must be of weight zero; this logic breaks down for single-qubit $\sigma_{\vec{v}_0}$, as will discuss below. Thus $\epsilon D_{\vec{v}_0;\vec{v}_0} = \epsilon\alpha I_{\text{env}} + O(\epsilon^2)$ for some α, as required. $\qquad\square$

From this lemma, one can see that any two-qubit Pauli gate does not involve any first-order ϵ terms which cannot be dealt with by the REAS method. However, crucially, the same logic does not hold in the case where $\sigma_{\vec{v}_0}$ is a *single-qubit* Pauli operator, which can be seen by considering the following counterexample:

$$e^{-i(sX\otimes I_{\text{env}} + \epsilon sX\otimes X_{\text{env}})t} = (e^{-isXt} \otimes I_{\text{env}})e^{-i\epsilon s(X\otimes X_{\text{env}})t} = (e^{-isXt} \otimes I_{\text{env}})[I - i\epsilon st(X \otimes X_{\text{env}}) + O(\epsilon^2)] \tag{C10}$$

which has a first-order ϵ term proportional to $s(X \otimes X_{\text{env}})$. Therefore, we simply modify any single-qubit Pauli gates into two-qubit ones by making use of the single Pauli transformation defined in Eq. (14). In this way, we can remove any first-order ϵ terms, as formalized below.

Theorem 1. *For any interaction Hamiltonian with an operator norm $O(\epsilon)$ that contains at most two-body terms, the gate sequences*

$$e^{isYIt} \rightarrow e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}$$

$$e^{isZIt} \rightarrow e^{-i\frac{\pi}{4}XI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{i\frac{\pi}{4}XI}$$

$$e^{isXIt} \rightarrow e^{i\frac{\pi}{4}ZI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{-i\frac{\pi}{4}ZI}. \tag{C11}$$

implement single qubit Pauli gates $e^{-is\sigma_{\vec{v}_0}}$ without giving rise to any first-order ϵ error terms of form $s\sigma_{\vec{v}_0} \otimes M_{\mathrm{env}}$.
More precisely, any noisy gate $e^{-is\sigma_{\vec{v}_0}}$ leads to

$$(e^{-is\sigma_{\vec{v}_0}} \otimes U_{\mathrm{env}}) \left[I + i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes (\overline{D}'_{\vec{v}_0;\vec{w}} + s\Delta D'_{\vec{v}_0;\vec{w}}) \right] \tag{C12}$$

for some unitary $U_{\mathrm{env}} \in \mathcal{H}_{\mathrm{env}}$, for which $\epsilon\Delta D'_{\vec{v}_0;\vec{v}_0} = \epsilon\alpha I + O(\epsilon^2)$ for some α.

Note that in Eq. (C11), the only s-dependency on the r.h.s. occurs through the term e^{isZZt} in the middle. It is straightforward to see that the proof of Thm. 1 can be generalized to the case where any other two-qubit Pauli gate is used to incorporate the s-dependency, e.g., e^{isZXt}, as long as the overall gate sequences lead to the correct single Pauli rotation gates.

Proof: We explicitly show the proof for only the last equation of Eq. (C11); the logic can be easily generalized to demonstrate the other two. First, note that $e^{\pm i\frac{\pi}{4}\sigma}$ for all Pauli operators σ are Clifford operators acting as

$$e^{i\frac{\pi}{4}\sigma}\sigma' e^{-i\frac{\pi}{4}\sigma} = \begin{cases} \sigma' & \sigma \text{ and } \sigma' \text{ commutes} \\ \frac{i}{2}[\sigma,\sigma'] & \sigma \text{ and } \sigma' \text{ anti-commutes} \end{cases}. \tag{C13}$$

Combining this with the identity $Ue^{-iHt}U^\dagger = e^{-iUHU^\dagger t}$ for unitary U and Hamiltonian H, it can be easily checked that

$$e^{isXIt} = e^{i\frac{\pi}{4}ZI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{-i\frac{\pi}{4}ZI} \tag{C14}$$

in the error-free situation.

Now, suppose that $e^{is(ZZ\otimes I)t}$ (i.e., I is taken to act on the rest of $\mathcal{H}$) leads to an s-dependent error of the form

$$e^{is(ZZ\otimes I)t} \to (e^{is(ZZ\otimes I)t} \otimes U_{\mathrm{env}}) \left[I + i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes (\overline{D}_{\vec{w}} + s\Delta D_{\vec{w}}) \right] \tag{C15}$$

for some unitary U_{env}. For the s-independent gates, the errors induced are of the form

$$G_j \to (G_j \otimes U_{\mathrm{env},j}) \left[I + i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes C_{\vec{w},j} \right] \tag{C16}$$

where G_1, G_2, G_3, and G_4 refers to $e^{i\frac{\pi}{4}(ZI\otimes I)}$, $e^{i\frac{\pi}{4}(XZ\otimes I)}$, $e^{-i\frac{\pi}{4}(XZ\otimes I)}$, and $e^{-i\frac{\pi}{4}(ZI\otimes I)}$, respectively. Then, the 0th-order ϵ term in the noisy version of the sequence $e^{i\frac{\pi}{4}ZI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{-i\frac{\pi}{4}ZI}$ that implements the desired single Pauli rotation e^{isXIt} can be written as

$$e^{is(XI\otimes I)t} \otimes \widetilde{U}, \tag{C17}$$

where

$$\widetilde{U} := U_{\mathrm{env},1}U_{\mathrm{env},2}U_{\mathrm{env}}U_{\mathrm{env},3}U_{\mathrm{env},4}. \tag{C18}$$

We now move to analyze the first-order ϵ error term. Such a term in the noisy version of $e^{i\frac{\pi}{4}ZI}e^{i\frac{\pi}{4}XZ}e^{isZZt}e^{-i\frac{\pi}{4}XZ}e^{-i\frac{\pi}{4}ZI}$ appears when an error occurs at a single gate in the sequence. Suppose for now that said error occurs at $e^{is(ZZ\otimes I)t}$; then the term of interest is of the form

$$(G_1 \otimes U_{\mathrm{env},1})(G_2 \otimes U_{\mathrm{env},2})(e^{is(ZZ\otimes I)t} \otimes U_{\mathrm{env}}) \left[i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes (\overline{D}_{\vec{w}} + s\Delta D_{\vec{w}}) \right] (G_3 \otimes U_{\mathrm{env},3})(G_4 \otimes U_{\mathrm{env},4}). \tag{C19}$$

This expression can be rewritten as

$$i\epsilon \sum_{\vec{w}} \left(G_1 G_2 e^{is(ZZ\otimes I)t}G_3 G_4 \cdot G_4^\dagger G_3^\dagger \sigma_{\vec{w}} G_3 G_4 \right) \otimes (U_{\mathrm{env},1}U_{\mathrm{env},2}U_{\mathrm{env}}(\overline{D}_{\vec{w}} + s\Delta D_{\vec{w}})U_{\mathrm{env},3}U_{\mathrm{env},4})$$

$$= i\epsilon \sum_{\vec{w}} \left(e^{is(XI\otimes I)t} \cdot G_1 G_2 \sigma_{\vec{w}} G_3 G_4 \right) \otimes (U_{\mathrm{env},1}U_{\mathrm{env},2}U_{\mathrm{env}}(\overline{D}_{\vec{w}} + s\Delta D_{\vec{w}})U_{\mathrm{env},3}U_{\mathrm{env},4})$$

$$= (e^{is(XI\otimes I)t} \otimes \widetilde{U}) \left[i\epsilon \sum_{\vec{w}} (G_1 G_2 \sigma_{\vec{w}} G_3 G_4) \otimes U_{\mathrm{env},4}^\dagger U_{\mathrm{env},3}^\dagger (\overline{D}_{\vec{w}} + s\Delta D_{\vec{w}})U_{\mathrm{env},3}U_{\mathrm{env},4} \right], \tag{C20}$$

using the fact that $G_3^\dagger = G_2$ and $G_4^\dagger = G_1$. Since $G_1 G_2 \sigma_{\vec{w}_0} G_3 G_4 = XI \otimes I$ for $\sigma_{\vec{w}_0} = (ZZ \otimes I)$, the term with $s(XI \otimes I)$ on the system part in the error term can be expressed as

$$i\epsilon s(XI \otimes I) \otimes (U_{\text{env},4}^\dagger U_{\text{env},3}^\dagger \Delta D_{\vec{w}_0} U_{\text{env},3} U_{\text{env},4}) \tag{C21}$$

Since $\epsilon \Delta D_{\vec{w}_0} = \epsilon \alpha I + O(\epsilon^2)$ by Lem. 2, Eq. (C21) does not have the first-order ϵ term where the environment part is not proportional to the identity.

Next, we consider the case where an error term occurs at one of $G_1, \ldots, G_4$. We only explicitly show the proof for the case where error term is chosen at G_1; the other cases follow similarly. The corresponding term is expressed as

$$(G_1 \otimes U_{\text{env},1}) \left[i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes C_{\vec{w},1} \right] (G_2 \otimes U_{\text{env},2})(e^{is(ZZ \otimes I)t} \otimes U_{\text{env}})(G_3 \otimes U_{\text{env},3})(G_4 \otimes U_{\text{env},4})$$

$$= i\epsilon \sum_{\vec{w}} (G_1 \sigma_{\vec{w}} G_2 e^{is(ZZ \otimes I)t} G_3 G_4) \otimes (U_{\text{env},1} C_{\vec{w},1} U_{\text{env},2} U_{\text{env}} U_{\text{env},3} U_{\text{env},4})$$

$$= (e^{is(XI \otimes I)t} \otimes \widetilde{U}) \left[i\epsilon \sum_{\vec{w}} (G_2 e^{is(ZZ \otimes I)t} G_3 G_4)^\dagger \sigma_{\vec{w}} (G_2 e^{is(ZZ \otimes I)t} G_3 G_4) \ldots \right.$$

$$\left. \ldots \otimes (U_{\text{env},2} U_{\text{env}} U_{\text{env},3} U_{\text{env},4})^\dagger C_{\vec{w},1} (U_{\text{env},2} U_{\text{env}} U_{\text{env},3} U_{\text{env},4}) \right], \tag{C22}$$

which displays no s-dependence, and thus cannot correspond to any term with $s(XI \otimes I)$ on the system part.

Therefore, in total, the noisy version of the sequence $G_1 G_2 e^{isZZt} G_3 G_4$ is written as

$$(e^{is(XI \otimes I)t} \otimes \widetilde{U}) \left[I + i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes (\overline{D}_{\vec{w}}' + s\Delta D_{\vec{w}}') \right] \tag{C23}$$

where $\epsilon \Delta D_{\vec{w}_0}' = \epsilon \alpha I + O(\epsilon^2)$ for $\sigma_{\vec{w}_0} := XI \otimes I$ and some α, as required. $\qquad\square$

Finally, combining Lem. 2 and Thm. 1, we can show that for decoherence errors on all Pauli rotation gates $e^{-is\sigma_{\vec{v}_0}\theta}$ can be dealt with by implementing the gates in such a way that their noisy version is written as

$$e^{-is\sigma_{\vec{v}_0}\theta} \to (e^{-is\sigma_{\vec{v}_0}\theta} \otimes U) \left[I + i\epsilon \sum_{\vec{w}} \sigma_{\vec{w}} \otimes (\overline{D}_{\vec{w}}' + s\Delta D_{\vec{w}}') \right] \tag{C24}$$

for $(\vec{v}_0, \theta)$-dependent unitary U and linear operators $\overline{D}_{\vec{w}}'$ and $\Delta D_{\vec{w}}'$, where $\epsilon \Delta D_{\vec{v}_0}' = \epsilon \alpha I + O(\epsilon^2)$ for some α. This means that the first-order ϵ error term of $e^{-is\sigma_{\vec{v}_0}\theta}$ is classified into the cases

(a) $I \otimes M$

(b) $\sigma_{\vec{w}} \otimes M$ for $\vec{w} \neq \vec{0}$

(c) $s\sigma_{\vec{w}} \otimes M$ for $\vec{w} \neq \vec{v}_0$

(d) $\alpha s\sigma_{\vec{w}} \otimes I_{\text{env}}$,

where M is any linear operator and α is arbitrary. Terms of type (a) can be avoided by absorbing them into the definition of U (namely, change $U \to Ue^{i\epsilon M^{(0)}}$ where $M^{(0)}$ is the 0th-order ϵ term of M). Terms of types (b) and (c) can be erased by averaging over random Pauli operators in a similar way to coherent errors described in App. A. Finally, terms of the form (d) can be absorbed into a shift in the rotation angle. Therefore, we have demonstrated that our proposed REAS scheme can deal with decoherence errors appearing as an $O(\epsilon)$ interaction Hamiltonian with at most two-body interaction in a similar way to how they deal with coherent errors.