Zero-noise extrapolation for quantum-gate error mitigation with identity insertions

Andre He [1,*] Benjamin Nachman [1,†] Wibe A. de Jong [2,‡] and Christian W. Bauer [1,§]

[1]*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[2]*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

Quantum-gate errors are a significant challenge for achieving precision measurements on noisy intermediate-scale quantum (NISQ) computers. This paper focuses on zero-noise extrapolation (ZNE), a technique that can be implemented on existing hardware, studying it in detail and proposing modifications to existing approaches. In particular, we consider identity insertion methods for amplifying noise because they are hardware agnostic. We build a mathematical formalism for studying existing ZNE techniques and show how higher order polynomial extrapolations can be used to systematically reduce depolarizing errors. Furthermore, we introduce a method for amplifying noise that uses far fewer gates than traditional methods. This approach is compared with existing methods for simulated quantum circuits. Comparable or smaller errors are possible with fewer gates, which illustrates the potential for empowering an entirely new class of moderate-depth circuits on near term hardware.

I. INTRODUCTION

Gate-based quantum computers are composed of unitary operations that are inevitably noisy due to interactions with the environment. Among the sources of errors facing quantum computation (such as readout errors; see, e.g., Ref. [1]), these gate errors are significant and a key challenge for achieving precision measurements on noisy intermediate-scale quantum (NISQ) computers [2]. The long-term solution to these errors is active error correction via error-correcting codes [3–7]. However, these algorithms require significant qubit and quantum gate resources and are therefore prohibitively expensive on NISQ devices. Active error detection and correction has been demonstrated for simple quantum circuits [8–17], but complete error correction is infeasible for moderately deep circuits with current hardware.

An alternative strategy for mitigating quantum-gate errors is to perform a series of measurements with systematically amplified errors and then to use these measurements to extrapolate to zero error. These strategies are called zero-noise extrapolation (ZNE). One possible strategy for enlarging the errors is to slow down gate operations [18]. This provides a continuous handle for making circuits noisier, but it also requires control over the hardware operations beyond that of a typical quantum computer user. A variety of "software" alternatives have been proposed which only require modifying the original quantum circuit and are hardware agnostic [19]. Some of these approaches require knowledge of the quantum computer noise model in order to amplify the noise [20,21]. In contrast, a noise-model-agnostic strategy is to replace unitary operations U with $U(UU^{\dagger})$. The additional $UU^{\dagger}$ does not

effect the zero-noise result of the circuit as $UU^{\dagger}$ is the identity. However, this *identity insertion* will amplify the noise. One dominant source of error in current quantum computing hardware is the gate error, which arises during the application of an entangling gate such as the two-gate controlled NOT operation (CNOT). This paper will use CNOT gates as the prototypical example, but the method applies to any unitary operation. One can repeat the identity insertion method multiple times by replacing the ith CNOT in a given circuit with

$$r_i = 2n_i + 1 \tag{1}$$

CNOT gates, for integer $n_i \geqslant 0$. When $n_i = n$ for all i, this is the *fixed identity insertion method* (FIIM). Measurements are performed for various values of r and then a fit is performed in order to extrapolate to $r \to 0$. The application of FIIM was first proposed in Ref. [19] using a linear fit and exponential fits were studied in Ref. [22]. Linear superpositions of enlarged noise circuits were also studied in Ref. [21], which are similar to our results on higher order fits. One challenge with FIIM is that it requires a large number of quantum gates: For a fit with n parameters, one needs to add at least $(2n + 1)N_c$ gates to the nominal circuit with N_c CNOT gates. We propose a solution to this challenge by promoting the n_i from Eq. (1) to random variables to construct the *random identity insertion method* (RIIM). A careful choice of probability mass functions for the n_i can effectively mitigate depolarizing noise as well as or better than FIIM using far fewer quantum gates. Both RIIM and FIIM do not require any knowledge of the underlying quantum computer noise model.

This paper is organized as follows. Section II reviews linear ZNE in the presence of depolarizing noise. The RIIM technique is introduced in Sec. III. The potential of nonlinear fits is discussed in Sec. IV, connecting nonlinear fits and the superposition methods discussed in Ref. [21]. Sections V and VI extend the analysis to include other sources of quantum noise as well as statistical uncertainties, respectively.

*andrehe@lbl.gov
†bpnachman@lbl.gov
‡wadejong@lbl.gov
§cwbauer@lbl.gov

Numerical results with a simple two-qubit circuit and the quantum harmonic oscillator are presented in Sec. VII. The paper ends with conclusions and outlook in Sec. VIII.

II. LINEAR FIT USING FIIM IN THE DEPOLARIZING NOISE MODEL

One can build a concept for the impact of identity insertions analytically using a depolarizing noise model. In the density matrix formalism, the noisy CNOT operation between two qubits k and l in the state ρ is given by [7]

$$\rho \to \left(1 - \sum_{ij} \frac{\epsilon_{ij}^{(kl)}}{16}\right) U_C^{(kl)} \rho \, U_C^{(kl)}$$
$$+ \sum_{i,j=0}^{3} \frac{\epsilon_{ij}^{(kl)}}{16} \sigma_i^{(k)} \sigma_j^{(l)} \rho \, \sigma_i^{(k)} \sigma_j^{(l)}, \qquad (2)$$

where $U_C^{(kl)}$ is the CNOT operation controlled on qubit k and targeting qubit l, $\epsilon_{ij}^{(kl)} \ll 1$ quantifies the amount of noise, and $\sigma_i^{(k)}$, $\sigma_i^{(l)}$ are the set of single-qubit Pauli gates acting on qubits k and l, respectively. The noise on qubits other than k and l is assumed to be negligible. Equation (2) can accommodate any quantum map, but the focus will be on depolarizing noise, which is the most well studied and widely used (both in simulations and in characterizing quantum hardware).

The depolarizing noise model corresponds to the case where all noise parameters $\epsilon^{(kl)} \equiv \epsilon_{ij}^{(kl)}$ are equal to one another, in which case Eq. (2) becomes[1]

$$\rho \mapsto (1 - \epsilon^{(kl)}) U_C^{(kl)} \rho \, U_C^{(kl)} + \epsilon^{(kl)} \left(\frac{I_4^{(kl)}}{4} \otimes \rho_{\cancel{kl}}\right), \qquad (3)$$

where $I_4^{(kl)}$ is the 4×4 identity matrix on qubits k and l, and $\rho_{\cancel{kl}}$ is all of ρ aside from the kl qubits (partial trace with respect to qubits k and l). Equation (3) has the clear interpretation that with probability $\epsilon^{(kl)}$, ρ is equally likely to be in any of the four possible states: $|kl\rangle\langle kl| \oplus \rho_{\cancel{kl}} \in \{|00\rangle\langle 00|, |01\rangle\langle 01|, |10\rangle\langle 10|, |11\rangle\langle 11|\} \oplus \rho_{\cancel{kl}}$.

Suppose that two CNOT operations are applied sequentially on the same two qubits k and l. The impact on the state is given by

$$\rho \mapsto (1 - \epsilon^{(kl)})^2 \rho + [1 - (1 - \epsilon^{(kl)})^2]\left(\frac{I_4^{(kl)}}{4} \otimes \rho_{\cancel{kl}}\right). \quad (4)$$

Note that in the noiseless limit $\epsilon^{(kl)} \to 0$, Eq. (4) correctly reproduces the fact that the two CNOT gates form the identity, such that the density matrix is unaffected. Adding a third CNOT gate, one finds

$$\rho \mapsto (1 - \epsilon^{(kl)})^3 U_C^{(kl)} \rho \, U_C^{(kl)}$$
$$+ \left[1 - \left(1 - \epsilon^{(kl)}\right)^3\right]\left(\frac{I_4^{(kl)}}{4} \otimes \rho_{\cancel{kl}}\right). \qquad (5)$$

Extending the pattern of Eqs. (3)–(5), applying the same CNOT $r_i = 1 + 2n_i$ times in a row has the same effect as

---

[1]One can also derive the depolarizing channel from a microscopic master equation formalism [23].

applying it once with the noise amplified by r_i

$$\rho \mapsto (1 - \epsilon_i)^{r_i} U_C^i \rho \, U_C^i$$
$$+ [1 - (1 - \epsilon_i)^{r_i}]\left(\frac{I_4^{(kl)}}{4} \otimes \rho_{\cancel{kl}}\right), \qquad (6)$$

where the ith CNOT gate connects qubits k and l and to simplify notation, $\epsilon_i = \epsilon^{(kl)}$. The Taylor expansion of Eq. (6) around $\epsilon_i = 0$ to $\mathcal{O}(\epsilon_i)$ is given by

$$\rho \mapsto (1 - r_i \epsilon_i) U_C^{(kl)} \rho \, U_C^{(kl)} + r_i \epsilon_i \left(\frac{I_4^{(kl)}}{4} \otimes \rho_{\cancel{kl}}\right). \quad (7)$$

Thus, the action of r_i CNOT gates in a row is the same as the action of a single CNOT gate, but with the noise parameter amplified by a factor of r_i. In FIIM, all of the r_i are set to the same value r.

Let M be an observable and in a circuit containing $i = 1 \ldots N_c$ CNOT gates, and consider performing a measurement of the expectation value of M: $\langle M \rangle = \text{Tr}(M\rho)$. Using Eq. (6) results, the expectation value in the presence of depolarizing noise is given by

$$\langle M \rangle(r) = \left(1 - r \sum_{i=1}^{N_c} \epsilon_i\right) \langle M \rangle_{\text{ex}} + r \sum_{i=1}^{N_c} \epsilon_i \langle M \rangle_{\text{dep}_i}$$
$$+ O\left(\left(r \sum_{i=1}^{N_c} \epsilon_i\right)^2\right), \qquad (8)$$

where $\langle M \rangle_{\text{ex}}$ is the expectation value of the observable in the absence of noise, $\langle M \rangle_{\text{dep}_i}$ denotes the expectation value of the noiseless observable if the CNOT i is replaced with the depolarizing channel, and $r = 1, 3, \ldots$ is the same factor for every CNOT gate in the circuit.

From Eq. (8), the noiseless value of the expectation value is given by the measurement at $r = 0$

$$\langle M \rangle_{\text{ex}} = \langle M \rangle(0). \qquad (9)$$

Of course, it is not possible to directly perform a measurement at $r = 0$, since all circuits have noise. The idea of ZNE is to extract the noiseless limit by measuring the result of $\langle M \rangle(r)$ for various values of r and extrapolating to the value at $r = 0$. By construction, a linear fit is effective when the $O(\epsilon^2)$ terms in Eq. (8) are subdominant (the "linear regime"). In this regime, one expects to remove the dominant $O(\epsilon)$ terms with a linear fit so that after linear FIIM

$$\langle M \rangle_{\text{FIIM}} = \langle M \rangle_{\text{ex}} + O\left(\left(r_{\text{max}} \sum_{i=1}^{N_c} \epsilon_i\right)^2\right), \qquad (10)$$

where r_{max} is the maximum r value so that the circuit is still in the linear regime.
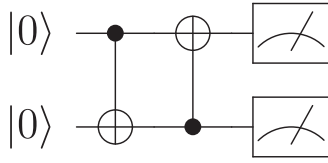
FIG. 1. An illustration of the the simple double-gate circuit described in Sec. II.

To provide further insight, it is useful to consider an explicit example where the density matrix is easy to compute for arbitrary r. Consider the simple circuit presented in Fig. 1. Because of the small number and simple orientation of gates, this model can be solved completely analytically.

Letting $r = r_1 = r_2 = 1 + 2n$ and $\epsilon = \epsilon_1 = \epsilon_2$, applying Eq. (6) to Fig. 1 results in the following mapping:

$$\rho \rightarrow (1 - \epsilon)^{N_c r} U_C^{(12)} U_C^{(21)} \rho \, U_C^{(21)} U_C^{(12)}$$
$$+ [1 - (1 - \epsilon)^{N_c r}] \frac{I_4}{4} \,, \tag{11}$$

where $N_c = 2$ denotes the total number of CNOT gates in the circuit and one needs to remember that r is an odd integer,

$$r = 1 + 2n \,. \tag{12}$$

Thus, starting from the initial state $|00\rangle$, one measures each of the four possible states with probability

$$P(|00\rangle) = 1 - \frac{3}{4}x \qquad P(|ij\rangle \neq |00\rangle) = \frac{x}{4}, \tag{13}$$

where the second equation holds for all three possibilities and

$$x_{\text{FIIM}}(\epsilon, n) = 1 - (1 - \epsilon)^{N_c(1+2n)}. \tag{14}$$

Suppose that one wants to measure $\langle q_0 + q_1 \rangle$, where q_i is the ith qubit in Fig. 1. The result of this measurement gives

$$\langle q_0 + q_1 \rangle = x_{\text{FIIM}}(\epsilon, 0)$$
$$= 1 - (1 - \epsilon)^{N_c}$$
$$= N_c \, \epsilon + O(\epsilon^2), \tag{15}$$

and is therefore linear in $N_c \epsilon$, as expected from Eq. (8). Using CNOT noise mitigation, one can remove the linear term in $N_c \epsilon$. In the linear FIIM method, one performs the measurement for various values of $n = 0, ..., n_{\max}$ and then extrapolates to the value $n = -1/2$ ($r = 0$). A linear fit with these data is a solution to the equation

$$Y = X\beta, \tag{16}$$

where

$$Y = \begin{pmatrix} x_{\text{FIIM}}(\epsilon, 0) \\ x_{\text{FIIM}}(\epsilon, 1) \\ \vdots \\ x_{\text{FIIM}}(\epsilon, n_{\max}) \end{pmatrix} \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ n_{\max} & 1 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix}. \tag{17}$$

The well-known least-squares solution to Eq. (17) is $\beta = (X^T X)^{-1} X^T Y$. This results in the fitted values $\hat{\beta}$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(i - \frac{n}{2}\right) x_{\text{FIIM}}(\epsilon, N_c(1+2i))}{\frac{1}{2}n(n+1)\left[\frac{1}{3}(2n+1) - \frac{n}{2}\right]}, \tag{18}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} \left[\frac{1}{6}n(2n+1) - \frac{ni}{2}\right] x_{\text{FIIM}}(\epsilon, N_c(1+2i))}{\frac{1}{2}n(n+1)\left[\frac{1}{3}(2n+1) - \frac{n}{2}\right]} \,. \tag{19}$$

Taylor expanding Eqs. (18) and (19) to $O(\epsilon^2)$ gives

$$\hat{\beta}_1 = 2(N_c\epsilon) + \left(-2n - 2 + N_c^{-1}\right)(N_c\epsilon)^2 + O(N_c\epsilon^3), \tag{20}$$

$$\hat{\beta}_0 = N_c\epsilon + \left[\frac{n(n-1)}{3} - \frac{1 - N_c^{-1}}{2}\right](N_c\epsilon)^2 + O(N_c\epsilon^3) \,. \tag{21}$$

The resulting equation is then

$$\langle q_0 + q_1 \rangle_{\text{FIIM[lin,}n_{\max}]} = \hat{\beta}_0 + \hat{\beta}_1 x, \tag{22}$$

where the subscript FIIM[lin, $n_{\max}$] denotes a linear fit performed with the first $n_{\max}$ values of n. Inserting Eqs. (20) and (21) into Eq. (22) and evaluating at $x = -1/2$ results in

$$\langle q_0 + q_1 \rangle_{\text{FIIM[lin,}n_{\max}]} = \left(\frac{2n_{\max}^2 + 4n_{\max} + 3}{6}\right)(N_c\epsilon)^2$$
$$+ O(N_c\epsilon^3) \,. \tag{23}$$

Using more data points makes the extrapolated result worse rather than better. This can be understood by the fact that using more data points requires more CNOT gates, pushing the measurement into the nonlinear regime. One should therefore expect that the error grows with the largest number of CNOT gates used, which is given by $r_{\max} N_c$. This can clearly be seen by rewriting the result of Eq. (23):

$$\langle q_0 + q_1 \rangle_{\text{FIIM[lin,}n_{\max}]} \xrightarrow{r_{\max} \rightarrow \infty} \frac{1}{12}(r_{\max} N_c \epsilon)^2 + O(N_c\epsilon^3) \,. \tag{24}$$

The best result is therefore obtained using a linear fit with two points, giving

$$\langle q_0 + q_1 \rangle_{\text{FIIM[lin,}1]} = \frac{3}{2}(N_c\epsilon)^2 + O(N_c\epsilon^3) \,. \tag{25}$$

A main drawback of linear FIIM is that it requires

$$r_{\max} \sum_{i=1}^{N_c} \epsilon_i \sim r_{\max} N_c \, \epsilon \ll 1 \,. \tag{26}$$

While this works well for circuits for which $N_c \epsilon$ is small enough that even after multiplication with $(1 + 2n)$ it is still a valid expansion parameter, for moderately deep circuits this condition can easily be invalid, in the sense that while one might trust an expansion in $N_c \epsilon$, the expansion breaks down for $3N_c \epsilon$ or $5N_c \epsilon$. This implies that a linear fit is no longer adequate to extrapolate to the noiseless ($r = 0$) limit.

III. LINEAR FIT USING RIIM IN THE DEPOLARIZING NOISE MODEL

The main challenge with the linear fit in the FIIM method is that the extrapolated zero noise result is only accurate to $O((r_{\max} N_{\text{CNOT}} \epsilon)^2)$, with $r_{\max}$ having to be at least equal to 3. Thus, for sufficiently deep circuits where $3N_c \epsilon \sim 1$, this method fails to give an accurate result for the zero-noise extrapolation.

Since the accuracy of the ZNE depends on the maximum number of CNOT gates required, a method that uses fewer

total CNOT gates should perform better. Instead of inserting the same number of identity operators for every CNOT gate, suppose instead that identities were randomly inserted. This gives raise to the *random identity insertion method* (RIIM). For this approach, one generalizes Eq. (8) such that each CNOT gate gets an independent factor r_i:

$$
\langle M \rangle (r_1, ..., r_{N_c}) = \left(1 - \sum_{i=1}^{N_c} r_i \, \epsilon_i \right) \langle M \rangle_{\mathrm{ex}} + \sum_{i=1}^{N_c} r_i \, \epsilon_i \langle M \rangle_{\mathrm{dep}_i}
$$
$$
+ O\left(\left(\sum_{i=1}^{N_c} r_i \, \epsilon_i \right)^2 \right), \tag{27}
$$

Next, the $r_i = 1 + 2n_i$ in Eq. (1) are promoted to random variables. For example, one could choose $n_i \sim \mathrm{Poisson}(\nu)$. As $\nu \to 0$, a given circuit will have at most one CNOT gate replaced. We will show that even in this case, one can still perform a linear fit and thus remove the $O(\epsilon)$ term with only $N_c + 2$ gates instead of $3N_c$ as in linear FIIM.

Using Eq. (6) similarly to Eq. (8), one can compute the expectation value of M for RIIM over both the quantum and classical (from sampling n) sources of stochasticity:

$$
\langle\langle M \rangle\rangle (\nu) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_{N_c}=0}^{\infty} \prod_{i=1}^{N_c} \mathrm{Pr}(n_i | \nu)
$$
$$
\times \left\{ \left[1 - \sum_i \epsilon_i (1 + 2n_i) \right] \langle M \rangle_{\mathrm{ex}} \right.
$$
$$
+ \sum_i \epsilon_i (1 + 2n_i) \langle M \rangle_{\mathrm{dep}_i}
$$
$$
\left. + O\left(\left(\sum_i (1 + 2n_i) \epsilon_i \right)^2 \right) \right\}, \tag{28}
$$

where the $\langle\langle \cdot \rangle\rangle$ represents an expectation value over the quantum fluctuations as well as the classical randomness from sampling different circuits. Since each gate is independently sampled, one can replace

$$
\sum_{n_i=0}^{\infty} \mathrm{Pr}(n_i | \nu) \, n_i = \nu, \tag{29}
$$

which immediately reduces Eq. (28) to

$$
\langle\langle M \rangle\rangle (\rho) = \left[1 - \rho \sum_i \epsilon_i \right] \langle M \rangle_{\mathrm{ex}} + \rho \sum_i \epsilon_i \langle M \rangle_{\mathrm{dep}_i}
$$
$$
+ O\left(\left(\rho \sum_i \epsilon_i \right)^2 \right), \tag{30}
$$

where $\rho = 1 + 2\nu$. Thus, Eq. (30) has the same feature as FIIM, only the integer n is now replaced by the noninteger value $\nu \geqslant 0$. By performing measurements at various values of ν and extrapolating to $\nu = -1/2$, one can extract the noiseless value. However, since the value ν is not restricted to be integer as in the FIIM case, the expansion does not have to hold for $3N_c \epsilon$, $5N_c \epsilon$, etc., but only for $\rho N_c \epsilon$, where one

can choose different values of ν to get a reasonable fit region without making ρ too far from unity.

IV. NONLINEAR FITS IN THE DEPOLARIZING NOISE MODEL

So far we have only discussed linear fits and showed that they can eliminate the $O(\epsilon)$ noise contribution to a given observable, leaving only quadratic dependence on the noise. In this section, we will generalize this result and show that one can in principle eliminate the depolarizing noise to all orders. This can be done for both the FIIM and RIIM methods, which we now discuss in turn.

A. FIIM method

We begin by revisiting the linear fit in the FIIM method, by writing it in a different way. Starting again from Eq. (8) and setting all $\epsilon \equiv \epsilon_i$ to be equal to one another, we can write

$$
\langle M \rangle (1) = \langle M \rangle_{\mathrm{ex}} + N_{\mathrm{CNOT}} \epsilon \left[\sum_i \langle M \rangle_{\mathrm{dep}_i} - \langle M \rangle_{\mathrm{ex}} \right]
$$
$$
+ O(\epsilon^2)
$$
$$
\langle M \rangle (3) = \langle M \rangle_{\mathrm{ex}} + 3 N_{\mathrm{CNOT}} \epsilon \left[\sum_i \langle M \rangle_{\mathrm{dep}_i} - \langle M \rangle_{\mathrm{ex}} \right]
$$
$$
+ O(\epsilon^2). \tag{31}
$$

One can immediately see that the linear combination

$$
\tfrac{3}{2} \langle M \rangle (1) - \tfrac{1}{2} \langle M \rangle (3) = \langle M \rangle_{\mathrm{ex}} + O(\epsilon^2). \tag{32}
$$

This is exactly what the linear fit to $r = 0$ using the two points at $r = 1, 3$ would give.

Generalizing these results, one can obtain linear combinations that remove higher order terms in ϵ as well. This fact has been observed before [21] and is an application of the Richardson extrapolation [24,25]. We will still review the results here, since they have not been used in ZNE using CNOT multiplication as a way to increase noise, and will prove useful later. Taking a particular linear combination of the terms with $r = 1, 3, \ldots, r_{\mathrm{max}}$, one can eliminate all terms up to $O(\epsilon^{n_{\mathrm{max}}+1})$ with

$$
n_{\mathrm{max}} = \frac{r_{\mathrm{max}} - 1}{2}. \tag{33}
$$

We begin by writing a general linear combination of measurements $\langle M \rangle (r)$ with different values of r and require that this linear combination eliminates all terms up to $O(\epsilon^{n_{\mathrm{max}}+1})$:

$$
\sum_{n=0}^{n_{\mathrm{max}}} a(n) \langle M \rangle (1 + 2n) = \langle M \rangle_{\mathrm{ex}} + O(\epsilon^{n_{\mathrm{max}}+1}). \tag{34}
$$

Ensuring that for any choices of $a(r)$ the coefficient of $\langle M \rangle_{\mathrm{ex}}$ is equal to one gives the constraint

$$
\sum_{n=0}^{n_{\mathrm{max}}} a(n) = 1. \tag{35}
$$

The expression for $\langle M \rangle (r)$ in the depolarizing noise model to all orders in ϵ can be obtained from Eq. (6) and one

finds

$$
\langle M \rangle(r) = (1-\epsilon)^{N_c r} \langle M \rangle_{\mathrm{ex}} + (1-\epsilon)^{(N_c-1)r}[1-(1-\epsilon)^r]
$$
$$
\times \sum_i \langle M \rangle_{\mathrm{dep}_i} + (1-\epsilon)^{(N_c-2)r}[1-(1-\epsilon)^r]^2
$$
$$
\times \sum_{i_1,i_2} \langle M \rangle_{\mathrm{dep}_{i_1 i_2}} + \cdots + [1-(1-\epsilon)^r]^{N_c}
$$
$$
\times \sum_{i_1,\ldots i_{N_c}} \langle M \rangle_{\mathrm{dep}_{i_1 \ldots i_{N_c}}}, \tag{36}
$$

where $\langle M \rangle_{\mathrm{dep}_{i_1,\ldots i_n}}$ is the exact result of a circuit in which the CNOT gates $i_1 \ldots i_n$ have been replaced by the depolarizing channel. Then, one can write

$$
(1-\epsilon)^{(N_c-i)r}[1-(1-\epsilon)^r]^i
$$
$$
= (1-\epsilon)^{(N_c-i)r} \sum_{j=0}^i \binom{i}{j}(1-\epsilon)^{jr}(-1)^j
$$
$$
= \sum_{j=0}^i \binom{i}{j}(1-\epsilon)^{(N_c-i+j)r}(-1)^j
$$
$$
= \sum_{j=0}^i \binom{i}{j}[(1-\epsilon)^{(N_c-i+j)r}-1](-1)^j + \sum_{j=0}^i \binom{i}{j}(-1)^j
$$
$$
= \sum_{j=0}^i \binom{i}{j}f_{N_c-i+j}(r,\epsilon)(-1)^{j+1}, \tag{37}
$$

where

$$
f_n(r,\epsilon) = 1 - (1-\epsilon)^{nr}. \tag{38}
$$

Therefore, one can write Eq. (36) as

$$
= \langle M \rangle_{\mathrm{ex}} - f_{N_c}(r,\epsilon)\langle M \rangle_{\mathrm{ex}}
$$
$$
+ [f_{N_c}(r,\epsilon) - f_{N_c-1}(r,\epsilon)] \sum_i \langle M \rangle_{\mathrm{dep}_i}
$$
$$
- [f_{N_c}(r,\epsilon) - 2f_{N_c-1}(r,\epsilon) + f_{N_c-2}(r,\epsilon)] \sum_{i_1,i_2} \langle M \rangle_{\mathrm{dep}_{i_1 i_2}}
$$
$$
+ \cdots - (-1)^{N_c}[f_{N_c}(r,\epsilon) - \cdots + f_0(r,\epsilon)]\langle M \rangle_{\mathrm{dep}_{i_1 \ldots i_{N_c}}}. \tag{39}
$$

It is important to remember that the values of $\langle M \rangle_{\mathrm{ex}}$, $\langle M \rangle_{\mathrm{dep}_i}$, $\langle M \rangle_{\mathrm{dep}_{i_1 \ldots i_{N_c}}}$, etc. are the results of observables measured in a noiseless circuit which one does not have access to. This means that when taking linear superposition of the form Eq. (34) the all terms up to $O(\epsilon_{\max}^n)$ have to cancel for each line separately.

This means that the requirement on the coefficients $a(n)$ must satisfy the general equation

$$
\sum_{n=0}^{n_{\max}} a(n)f_k(1+2n,\epsilon) = 1 + O(\epsilon^{n_{\max}+1}), \tag{40}
$$

for all values of k. After some lines of algebra, one can show that this is indeed possible with the coefficients [21]

$$
a(i) = \prod_{j=0,j\neq i}^{n_{\max}} \frac{(1+2j)}{2(j-i)}
$$
$$
= \frac{2^{-2n_{\max}}}{i!}\frac{(-1)^i}{1+2i}\frac{(1+2n_{\max})!}{n_{\max}!(n_{\max}-i)!}, \tag{41}
$$

for all $i \in 1 \ldots n_{\max}$. Note that the coefficient for $i \sim n_{\max}/2$ is the largest and satisfies the scaling

$$
\max_i[a(i)] \sim a(n_{\max}/2) \sim \frac{2^{n_{\max}+1}}{n_{\max}}. \tag{42}
$$

To summarize, by using values $\langle M \rangle(r)$ with $r = 1, 3, \ldots, r_{\max}$ and taking the linear combination $\sum_{n=0}^{n_{\max}} a(n)\langle M \rangle(1+2n)$, one obtains the noiseless value of the observable up to corrections given by $O(\epsilon^{n_{\max}+1})$.

One alternative approach with a natural interpretation is performing a polynomial fit with degree $n_{\max}-1$ to measurements of $\langle M \rangle(r)$ with $r = 1, 3, \ldots, r_{\max}$. A polynomial fit uses the same setup for the linear fit, with Eq. (16), only now X and β are augmented:

$$
X = \begin{pmatrix} 0^{n_{\mathrm{fit}}} & \cdots & 0 & 1 \\ 1^{n_{\mathrm{fit}}} & \cdots & 1 & 1 \\ \vdots & \vdots & & \\ n_{\max}^{n_{\mathrm{fit}}} & \cdots & n_{\max} & 1 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{n_{\mathrm{fit}}} \\ \vdots \\ \beta_1 \\ \beta_0 \end{pmatrix}, \tag{43}
$$

where n_{fit} is the order of the polynomial. One can show that extrapolating the resulting fit

$$
\langle M \rangle_{\mathrm{FIIM}[n_{\mathrm{fit}},n_{\max}]} = \sum_{i=1}^{n_{\mathrm{fit}}} \hat{\beta}_i x^i, \tag{44}
$$

to $x = -\frac{1}{2}$ removes the $O(\epsilon^{n_{\max}+1})$ component of the depolarizing error when $n_{\mathrm{fit}} = n_{\max}-1$. Both the polynomial fit and the superposition from Eq. (41) give rise to the same linear combinations of the values measured at various values of r. One can show this with some symbolic manipulation

$$
\langle M \rangle_{\mathrm{FIIM}[n_{\mathrm{fit}},n_{\max}]} = \sum_{i=0}^{n_{\mathrm{fit}}} \hat{\beta}_i \left(-\frac{1}{2}\right)^i
$$
$$
= \sum_{i=0}^{n_{\mathrm{fit}}} \sum_{j=0}^{n_{\max}} ((X^T X)^{-1}X^T)_{ij} Y_j \left(-\frac{1}{2}\right)^i
$$
$$
= \sum_{j=0}^{n_{\max}} \left[\sum_{i=0}^{n_{\mathrm{fit}}} ((X^T X)^{-1}X^T)_{ij}\left(-\frac{1}{2}\right)^i\right] Y_j
$$
$$
= \sum_{n=0}^{n_{\max}} \left[\sum_{i=0}^{n_{\mathrm{fit}}} ((X^T X)^{-1}X^T)_{in}\left(-\frac{1}{2}\right)^i\right]
$$
$$
\times \langle M \rangle(1+2n)
$$
$$
\equiv \sum_{n=0}^{n_{\max}} \tilde{a}(n)\langle M \rangle(1+2n). \tag{45}
$$

We have verified that the $\tilde{a}(n)$ in Eq. (45) are equivalent to the $a(n)$ in Eq. (41).

TABLE I. Table giving the coefficients for higher order RIIM fits.

n_{max}	$a_{\{3\}}$	$a_{\{5\}}$	$a_{\{3,3\}}$	$a_{\{7\}}$	$a_{\{5,3\}}$	$a_{\{3,3,3\}}$	$a_{\{9\}}$	$a_{\{7,3\}}$	$a_{\{5,5\}}$	$a_{\{5,3,3\}}$	$a_{\{3,3,3,3\}}$
1	$-\frac{1}{2}$										
2	$-\frac{N_c+4}{4}$	$\frac{3}{8}$	$\frac{1}{4}$								
3	$-\frac{N_c^2+10N_c+24}{16}$	$\frac{3(N_c+6)}{16}$	$\frac{N_c+6}{8}$	$-\frac{5}{16}$	$-\frac{3}{16}$	$-\frac{1}{8}$					
4	$-\frac{N_c^3+18N_c^2+104N_c+192}{96}$	$\frac{3N_c^2+32N_c+154}{64}$	$\frac{N_c^2+14N_c+59}{32}$	$-\frac{45}{32}$	$-\frac{3N_c+29}{32}$	$-\frac{N_c+8}{16}$	$\frac{35}{128}$	0	$\frac{29}{64}$	$\frac{3}{32}$	$\frac{1}{16}$

B. RIIM method

The RIIM method uses a different value of r_i for each CNOT gate. Applying Eq. (6) with the full ϵ dependence leads to the analog of Eq. (39) from FIIM:

$$\langle M \rangle (r_1, \ldots, r_{N_c})$$
$$= \prod_j (1-\epsilon)^{r_j} \left[\langle M \rangle_{\text{ex}} + \sum_i \frac{f_1(r_i, \epsilon)}{(1-\epsilon)^{r_i}} \langle M \rangle_{\text{dep}_i} \right.$$
$$+ \sum_{i_1 > i_2} \frac{f_1(r_{i_1}, \epsilon)}{(1-\epsilon)^{r_{i_1}}} \frac{f_1(r_{i_2}, \epsilon)}{(1-\epsilon)^{r_{i_2}}} \langle M \rangle_{\text{dep}_{i_1 i_2}} + \cdots$$
$$\left. + \sum_{i_1 > \ldots > i_{N_c}} \frac{f_1(r_{i_1}, \epsilon)}{(1-\epsilon)^{r_{i_1}}} \cdots \frac{f_1(r_{i_{N_c}}, \epsilon)}{(1-\epsilon)^{r_{i_{N_c}}}} \langle M \rangle_{\text{dep}_{i_1 \ldots i_{N_c}}}, \quad (46)$$

where all of the multisums are over all sets of indices i_j from the original $(1, \ldots, N_c)$ that satisfy the ordering relation given under the sum symbol. To eliminate all terms up to order $\epsilon^{n_{max}}$, one needs to include all possible combinations of $r_1, \ldots, r_{N_c}$ with $\sum_i r_i = N_c + 2n_{max}$. To write a generic solution we require a bit of notation. Denote by $O(\{e_1, \ldots e_n\})$ the sum of all operators with the r_i given by permutations of 1 and the various values of $e_i = 3, 5, 7, \ldots$. So

$$O(\{\}) = O(1, \ldots, 1),$$
$$O(\{e_1\}) = O(e_1, 1, \ldots, 1) + O(1, e_1, \ldots, 1) + \cdots,$$
$$O(\{e_1, e_2\}) = O(e_1, e_2, 1, \ldots, 1) + O(e_1, 1, e_2, \ldots, 1) + \cdots, \quad (47)$$

and so on.

To eliminate all terms up to $\epsilon^{n_{max}}$, one includes all operators $O(\{e_1 \ldots e_n\})$ with $\sum_i e_i \leqslant 2n_{max} + N_c$, each with its own coefficient. One then determines the coefficients by demanding that all terms up to $\epsilon^{n_{max}}$ vanish. So, for example, to eliminate the linear term in ϵ, one includes the operator $O(\{\})$ and $O(\{3\})$. Solving the equations yields

$$a_{\{\}} O(\{\}) + a_{\{3\}} O(\{3\}) = 0 + O(\epsilon^2), \quad (48)$$

with

$$a_{\{\}} = 1 - a_{\{3\}} N_c. \quad (49)$$

Solving this equation, one finds

$$a_{\{3\}} = -\frac{1}{2}, \quad (50)$$

which again reproduces the result of the linear fit discussed in Sec. III. To eliminate the linear and quadratic term in ϵ, one

includes the operators $O(\{\})$, $O(\{3\})$, $O(\{5\})$, and $O(\{3, 3\})$ and solves the equation

$$a_{\{\}} O(\{\}) + a_{\{3\}} O(\{3\}) + a_{\{5\}} O(\{5\}) + a_{\{3,3\}} O(\{3, 3\})$$
$$= 0 + O(\epsilon^3), \quad (51)$$

again with the constraint

$$a_{\{\}} = 1 - a_{\{3\}} N_c - a_{\{5\}} N_c - a_{\{3,3\}} \binom{N_c}{2}. \quad (52)$$

In general, the coefficient for each term a in the constraint expressions can be determined by

$$\binom{N_c}{|K|} \frac{|K|!}{K_1! K_2! \ldots K_n!}$$

where K is the set of extra CNOT gates to be randomly inserted in the circuit, and K_n is the number of times each unique element in K is repeated.

Solving the resulting set of equations gives

$$a_{\{3\}} = -\frac{N_c + 4}{4}, \qquad a_{\{5\}} = \frac{3}{8}, \qquad a_{\{3,3\}} = \frac{1}{4}. \quad (53)$$

While we have not been able to derive a closed-form expressions for the coefficients yet, we report valid choices for the various coefficients with[2] $n_{max} = 1, 2, 3, 4$ in Table I. These results allow us to remove depolarizing noise with corrections arising at $\epsilon^{n_{max}+1}$ using $N_c + 2n_{max}$ gates. This should be compared with the FIIM method where the same noise reduction requires $(2n_{max} + 1)N_c$ gates.

For relatively shallow circuits, one could feasibly perform the measurements for all permutations required for $O(\{e_1, \ldots, e_n\})$. For example, to remove the $O(\epsilon)$ error, one would need to perform $N_c + 1$ sets of measurements. However, this quickly becomes impractical. This can be circumvented by randomizing: For each measurement that goes into $O(\{e_1, \ldots, e_n\})$, randomly pick one of the $N_{\{e_1, \ldots, e_n\}}$ operations.

---

[2] There are enough equations so that solutions exist for all n_{max} and these can be found numerically.

TABLE II. A comparison of the gate count needed for a given order of depolarization error correction for FIIM and RIIM.

Method	Remainder	No. of CNOTs
FIIM	$O(\epsilon^n)$	$(2n - 1)N_c$
RIIM	$O(\epsilon^n)$	$N_c + 2(n - 1)$

Table II provides an overview of the gate count required for FIIM and RIIM in the removal of depolarization noise at a given order in ϵ.

V. BEYOND THE DEPOLARIZING NOISE MODEL

Equation (2) introduced the full Kraus representation of a noisy CNOT gate. Let $\epsilon_{ij} = \epsilon + \delta_{ij}$. The depolarizing error model is the case where $\delta_{ij} = 0$ and is what has been considered thus far. In reality, there will be some nonzero δ_{ij}, though the nondepolarizing error has been less studied in the literature and less characterized on current hardware platforms. While the methods studied in the previous sections are able to suppress the depolarizing error to $O(\epsilon^{n_{\max}})$, they do not remove the $O(\delta)$ term. This means that it is not useful to go beyond $O(\epsilon^2)$, unless $\delta < \epsilon^2$.

There are many other sources of noise, important examples being amplitude damping and decoherence noise. The latter can be well approximated as an exponential random variable per operation, where the gate has some fidelity (time constant) and requires some finite time to perform. We leave the study of such noise to future investigations, but we anticipate that methods similar to those studied here can be used to remove noise other than depolarizing noise as well. In fact, in Ref. [21] it was argued that similar methods also apply to amplitude damping noise.

VI. STATISTICAL UNCERTAINTY

All results presented so far were in the limit where one can measure the value of an observable with arbitrary precision. This is, of course, not true, since any measurement on a quantum computer is probabilistic in nature, such that most measurements have a statistical uncertainty associated with them, which depends inversely on the square root of the number of runs used to perform the measurement.

Using the results of the previous sections, one can quantify the impact of the statistical uncertainty. Recall that the noiseless value $\langle M \rangle_{\mathrm{ex}}$ is obtained by taking linear combinations of measurements with different values of r, and that in the limit of zero statistical uncertainty the final uncertainty on the noiseless value is given by the maximum of δ and $\epsilon^{n_{\max}+1}$. In the presence of statistical uncertainty, each measurement of $\langle M \rangle(r)$ can only be determined up to a statistical uncertainty

$$\Delta(r) \sim \frac{1}{\sqrt{n_{\mathrm{meas}}}}, \tag{54}$$

where n_{meas} denotes the number of measurements that are performed in the measurement of each value $\langle M \rangle(r)$. The full uncertainty is more complicated and depends on the exact observable. This can be derived by writing the generic form of the ZNE expectation value:

$$\langle M \rangle_{\mathrm{ZNE}} = \sum_{\vec{r}} a(\vec{r}) \langle M \rangle(\vec{r})$$

$$= \sum_{\vec{r}} a(\vec{r}) \left(\frac{1}{n_{\mathrm{meas}}} \sum_{i=1}^{2^{n_{\mathrm{qubits}}}} \beta_i m_i(\vec{r}) \right),$$

$$= \frac{1}{n_{\mathrm{meas}}} \sum_{\vec{r}} \sum_{i=1}^{2^{n_{\mathrm{qubits}}}} (a(\vec{r}) \beta_i) m_i(\vec{r}), \tag{55}$$

where $a(\vec{r})$ is either of the form $a(r)$ for FIIM or $a_{\{r_1,\ldots,r_n\}}$ for RIIM. Going from the first to the second line, we have used that any observable is constructed from the measured counts $m_i(\vec{r})$ of the $2^{n_{\mathrm{qubit}}}$ states, where the weighting of each count m_i by β_i is observable dependent.

The m_i are random variables with n_{meas} total draws with probability vector $\vec{p} \in [0, 1]^{n_{\mathrm{qubits}}}$ with $p_i(\vec{r}) = p_{i,\mathrm{ex}} + O(\epsilon) \tilde{p}_i(\vec{r})$, $i = 1, \ldots, 2^{n_{\mathrm{qubits}}}$. The probability vectors satisfy $\sum_i p_i(\vec{r}) = 1$. The exact result $p_{i,\mathrm{ex}}$ is the probability of measuring state i without any depolarizing error and the $\vec{r}$ dependence is in the $O(\epsilon)$ term, and the observable dependence is now in the values of the $p_i(\vec{r})$. Thus, the $m_i(\vec{r})$ are drawn from a multinomial random variables, with variances and covariances given by $\mathrm{Var}(m_i) = n_{\mathrm{meas}} p_i (1 - p_i)$ and $\mathrm{Cov}(m_i, m_j) = -n_{\mathrm{meas}} p_i p_j$, respectively. Using this information, one can then compute the total standard deviation for $\langle M \rangle_{\mathrm{ZNE}}$:

$$\Delta_{\mathrm{stat}} = \frac{1}{\sqrt{n_{\mathrm{meas}}}} \sqrt{\sum_{\vec{r}} a(\vec{r})^2 M(\{\beta_i, p_i(\vec{r})\})}. \tag{56}$$

with

$$M(\{\beta_i, p_i\}) = \sum_{i=1}^{2^{n_{\mathrm{qubits}}}} \beta_i^2 p_i \left((1 - p_i) - \sum_{j=i+1}^{2^{n_{\mathrm{qubits}}}} \beta_j^2 p_j \right). \tag{57}$$

For observables for which only a single-qubit state contributes, such that only a single β_i is nonvanishing, the probability vector only has a single entry such that $\beta_k = 1$ for one $k \in \{1, \ldots, 2^{n_{\mathrm{qubits}}}\}$ and $p_{i \neq k, \mathrm{ex}} = 0$. Choosing $k = 1$ and $\beta_1 = 1$, Eq. (56) for FIIM becomes

$$\Delta_{\mathrm{stat}} = \frac{1}{\sqrt{n_{\mathrm{meas}}}} \sqrt{\sum_{n=0}^{n_{\max}} a(n)^2 p_1 (1 - p_1)}$$

$$\approx \frac{1}{\sqrt{n_{\mathrm{meas}}}} \sqrt{\sum_{n=0}^{n_{\max}} O(\epsilon) a(n)^2}$$

$$\sim \frac{\sqrt{\epsilon}}{\sqrt{n_{\mathrm{meas}}}} \frac{2^{n_{\max}}}{n_{\max}}, \tag{58}$$

where the last line is only true in the limit of large $n_{\max}$, since we have used that the sum is dominated by its largest values, given in Eq. (42).

Combining the statistical uncertainty with the uncertainties discussed before, the final uncertainties in the FIIM and RIIM methods are given by

$$\Delta_{\mathrm{FIIM/RIIM}}[\epsilon, \delta; n_{\max}, n_{\mathrm{meas}}] \sim \max[\delta, \epsilon^{n_{\max}}, \Delta_{\mathrm{stat}}]. \tag{59}$$

In practice, randomizing over circuits for RIIM will introduce a contribution to the statistical uncertainty. This may be optimized by a clever allocation of experiments. With limited resources, this summarizes RIIM versus FIIM as a tradeoff between statistical uncertainty and gate depth. On NISQ computers, gate depth is often the limiting factor and so this is a desirable tradeoff. One may want to use multiple computers or multiple sets of qubits on the same computer to perform RIIM (or FIIM) order to further improve the statistical precision. Such configurations will be studied in future work.
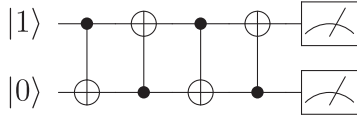
FIG. 2. A simple circuit with four CNOT gates used in this section.

TABLE III. Device error parameters for `ibmq-ourense` qubits 0 and 1.

Qubit no.	T1 (μs)	T2 (μs)	CNOT error rate
0	102.01	81.04	0.021
1	91.34	33.42	0.021

VII. NUMERICAL RESULTS

We use `qiskit` [26] to simulate the quantum circuits described below and demonstrate FIIM and RIIM. Section VII A studies the simple CNOT-only circuit from Fig. 1 and Sec. VII B examines a more complicated case of time evolution for the quantum simple harmonic oscillator.

A. Simple circuit

The simple circuit shown in Fig. 1 was particularly useful because of its analytical tractability. In particular, because one can compute the expectation values analytically, it is possible to consider the $n_{meas} \to \infty$ limit. In this section, we use a slight modification of this simple circuit shown in Fig. 2, which uses four CNOT gates, which are started in the initial state $|10\rangle$. In the noiseless limit, the final state is given by $|11\rangle$. Four gates are used in order to demonstrate the potential for removing depolarization errors up to ϵ^5, and we use a different initial state such that decoherence, discussed later in the section, is not driving the result towards the final expectation.

Figure 3 illustrates the scaling of the error and gate count for RIIM and FIIM for this circuit. As desired, the error decreases with the order of the error correction. The number of qubits required for RIIM is much lower than FIIM for a fixed

order of error correction. For example, correcting the $O(\epsilon^4)$ requires 8 total gates for RIIM but FIIM requires 36. In fact, for a fixed correction order, the coefficient of the subleading depolarizing error is also smaller for RIIM than for FIIM. Note that one needs $n_{fit} = n_{max} + 1$.

`qiskit` can be used to study the impact of other sources of noise, such as thermal relaxation. A full noise model from the IBMQ device is used, which includes depolarizing and decoherence errors. Table III provides an overview of the parameters used to generate a full noise model using `qiskit` for the `ibmq-ourense` device.

In Fig. 4, we show the result where the measured observable is the expected value of the output string, converting from binary numbers to integers ($00 \to 0, 01 \to 1, 10 \to 2, 11 \to 3$). In the noiseless limit, the expectation value is 3, corresponding to $|11\rangle$. Fixed identity insertions (but no corrections yet) are applied up to $r_{max} = 31$. The observable decays at a quicker rate in the case with the full noise model as expected, as the circuit feels the effect of thermal relaxation (which drives the system toward the $|00\rangle$ state) as well as the depolarizing noise, which drives the system to the completely mixed state.

Figure 5 compares the extrapolation error obtained from FIIM and RIIM under the action of full and purely depolarizing noise models. The extrapolation error for both FIIM and
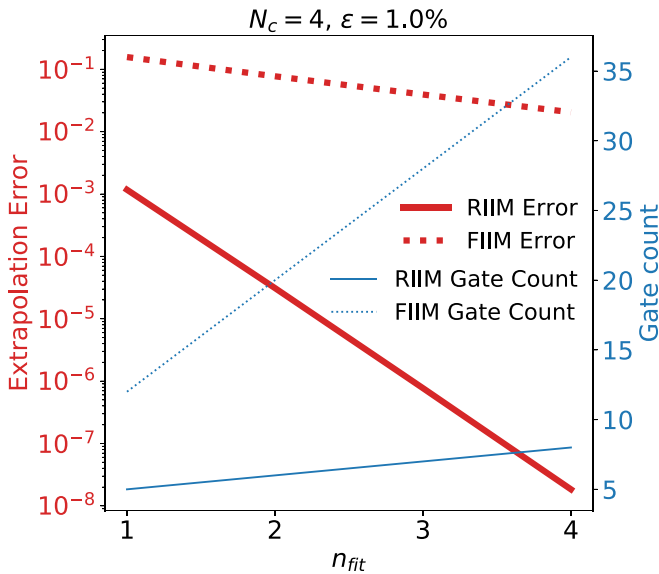

FIG. 3. Numerical results based on the higher order fits described in Sec. IV using the four-CNOT-gate circuit presented in Fig. 2. The horizontal axis is the order of the depolarizing error that is being removed. The left axis is the error on $\langle \sum_{i=0}^{N_c} q_i \rangle$ as ϵ is extrapolated to zero. The right axis is the number of gates requires to make the correction. Only depolarizing noise is considered and $n_{meas} = \infty$.
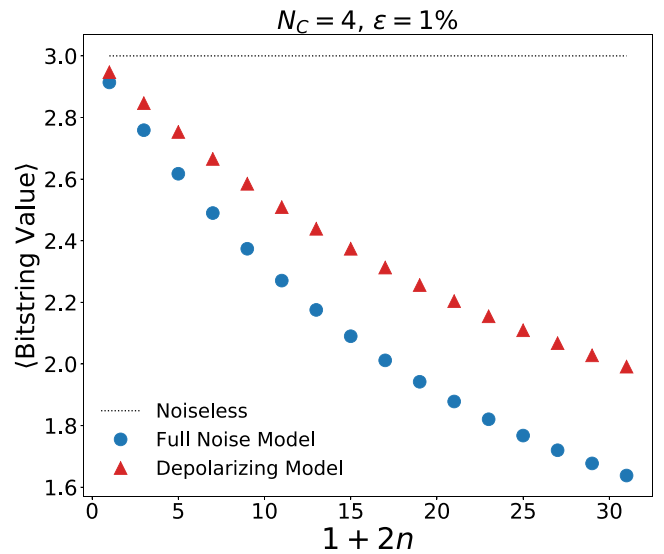

FIG. 4. Numerical results from simulating the four-CNOT circuit in noiseless and noisy simulators using `qiskit`. The vertical axis shows the expectation value of the measured observable. The horizontal axis displays r, or $1 + 2n$. Noisy simulations include both full and purely depolarizing cases. The number of shots for each point is 10^7, with a standard deviation of 10^{-3}.
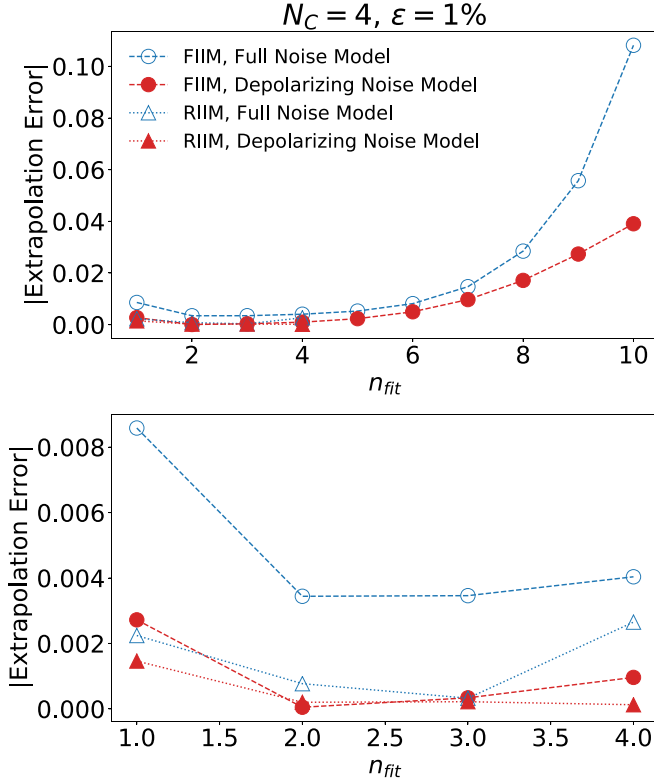
FIG. 5. Numerical results from simulating the four-CNOT circuit and applying FIIM and RIIM extrapolation in a noisy simulator. The lower plot displays a reduced range of the data in the upper plot for detail. Extrapolation error is 0 for noiseless simulation. The number of shots for each circuit is 10^7.

RIIM are higher in the case of a full noise model which has nondepolarizing elements. RIIM performs as well or better than FIIM in both noise models. The minimum extrapolation error is achieved at $n_{fit} = 2$. This can be understood in the context of Eq. (59) in Sec. VI with the parameters of $\epsilon = 1\%$ and $n_{meas} = 10^7$. The value for ϵ was chosen to be a reasonable estimate of the device error rates seen in the calibration data available on the IBMQ portal. At $n_{fit} = 1$, the dominant error $\Delta_{FIIM/RIIM}$ is determined by ϵ rather than the statistical uncertainty. However, as $n_{fit} = 2$, the statistical error Δ_{stat} begins to exceed ϵ^2, and by $n_{fit} = 3$ the dominant error becomes the statistical error, which prevents further reduction of extrapolation error and leads the the exponential scaling of the error as n_{fit} is increased further. Note that RIIM is only used to eliminate errors up to $O(\epsilon^4)$, as the circuit only contains four CNOTS.[3]

B. Hamiltonian evolution

Trotterized time evolution is a useful technique for the simulation of Hamiltonians on digital quantum computers. For the one-dimensional simple harmonic oscillator Hamiltonian,

---

[3]In principle, one could go higher by modifying the method as there are fewer terms on the right-hand side of Eq. (46) than need to be eliminated.
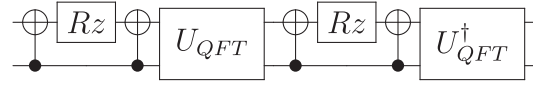

FIG. 6. Circuit diagram for a single Trotter step of the time evolution of the harmonic oscillator hamiltonian for two qubits. The total number of CNOT operations for the quantum Fourier transform U_{QFT} on two qubits is 5, giving a total of 14 CNOT operations. However, one CNOT operation from each of the U_{QFT} is canceling a CNOT from the rest of the circuit, giving a total of 10 CNOT operators per Trotter step.

time evolution is given by

$$|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle, \qquad (60)$$

where

$$H = \tfrac{1}{2}(\hat{x}^2 + \hat{p}^2) \equiv H_x + H_p. \qquad (61)$$

The Hamiltonian in Eq. (61) can be implemented on a digital quantum computer by discretizing the possible values of x to be $-x_{max}, -x_{max} + \delta_x, \ldots, x_{max} - \delta_x, x_{max}$, where $\delta_x = 2x_{max}/(2^{n_{qubits}} - 1)$ and n_{qubits} is the number of qubits. This system has been recently studied in the context quantum field theory as a benchmark $(0 + 1)$-dimensional noninteracting scalar field theory [27–34]. As discussed in these studies, the momentum operator $\hat{p}^2$ can be effectively implemented with quantum Fourier transforms. Since $[H_x, H_p] \neq 0$, one can approximate the time evolution of the Hamiltonian by using the first-order Suzuki-Trotter expansion [35–37]:

$$e^{-i(H_x + H_p)t} \approx \left[e^{-iH_x \frac{t}{n}} e^{-iH_p \frac{t}{n}} \right]^n$$
$$\equiv \left[U_n^{(H)}(t/n) \right]^n. \qquad (62)$$

The approximation in Eq. (62) can be efficiently represented as a quantum circuit block which is repeated n times to the desired number of Trotter steps, as illustrated in Fig. 6.

Time evolution of the ground state of the Harmonic oscillator gives

$$|\psi_0(t)\rangle = e^{-iHt}|\psi_0(0)\rangle = e^{-iE_0 t}|\psi_0(0)\rangle, \qquad (63)$$

where $E_0 = 1/2$. Thus, the time evolution produces a pure phase and one finds

$$\langle \psi_0(0)|\psi_0(t)\rangle = 1. \qquad (64)$$

The ground state of the harmonic oscillator is a Gaussian distribution in the variable x, which can be generated through the action of a unitary circuit (U_{State}) on the state $|0\rangle$, which is implemented with two CNOT gates:

$$|\psi_0(0)\rangle = U_{State}|0\rangle. \qquad (65)$$

Thus, the overlap can be written as

$$\lim_{n\to\infty} \langle 0|U_{State}^\dagger \left[U_n^{(H)}(t/n) \right]^n U_{State}|0\rangle = 1. \qquad (66)$$

For finite values of n, the deviation of the overlap from unity will grow with time t/n and one achieves higher accuracy for larger n

$$\langle 0|U_{State}^\dagger \left[U_n^{(H)}(t/n) \right]^n U_{State}|0\rangle = 1 + O(t^2/n^2). \qquad (67)$$
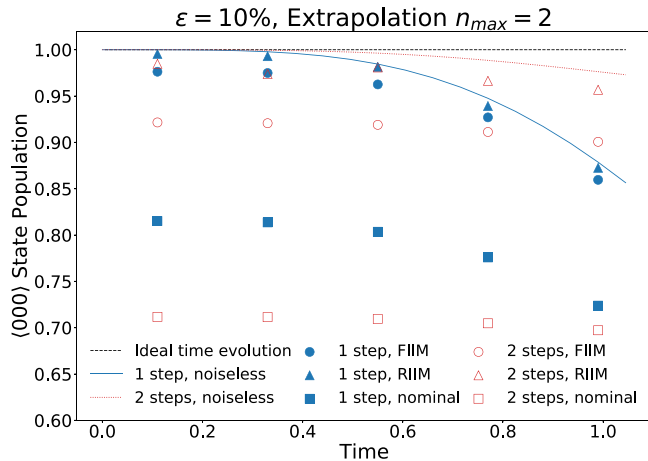
FIG. 7. The expectation value of the observable in Eq. (66) as a function of time for various numbers of Trotter steps and noise mitigation techniques. The purely depolarizing noise model is used for this simulation. In the absence of Trotterization error, the observable should be unity, independent of t.

FIG. 8. The squared error of the observable value extrapolated using FIIM plotted against n_{max}. The dashed lines indicate the error using second-order RIIM.

On the other hand, more Trotter steps requires deeper circuits, and therefore larger errors from the gate noise, in particular the CNOT noise.

We choose to simulate the harmonic oscillator with a total of two qubits, corresponding to four discrete values of x. In this case the CNOT count is given by

$$N_c = 4 + 10\,n. \qquad (68)$$

The accuracy of the approximation increases with the number of Trotter steps n. FIIM has been used to increase the accuracy of Trotterized simulation of the time evolution of Hamiltonians, but is less accurate when the depth of a single Trotter step becomes too large, as introducing three or more times as many CNOT operations as there are in the nominal circuit does not allow for the accurate extrapolation of the observable [38].

Figure 7 presents the result of one and two Trotter steps, corrected with RIIM and with FIIM up to $O(\epsilon^2)$. For both one and two steps, the RIIM extrapolations are closer to the noiseless lines than the FIIM extrapolations, indicating that the RIIM error is smaller than the FIIM one.

Figure 8 compares the error obtained from the FIIM and RIIM extrapolations over different values of n_{fit}. The extrapolated error from RIIM up to $O(\epsilon^2)$ is lower than any of the errors obtained through FIIM for all values of n_{fit} in the one-step case and in the two-step case.

VIII. CONCLUSIONS

We have performed a detailed study of zero-noise extrapolation for correcting gate errors in quantum circuits. The first aspect of this study was the formalization of the fixed identity insertion method (FIIM), which increases the circuit error by inserting pairs of gates after each CNOT in the circuit. This method has been studied in the past, but we derived analytic results for removing higher order depolarizing noise. These analytic results were previously known in the context of Hamiltonian evolution and are connected with the identity insertion formalism. We also made the observation that these
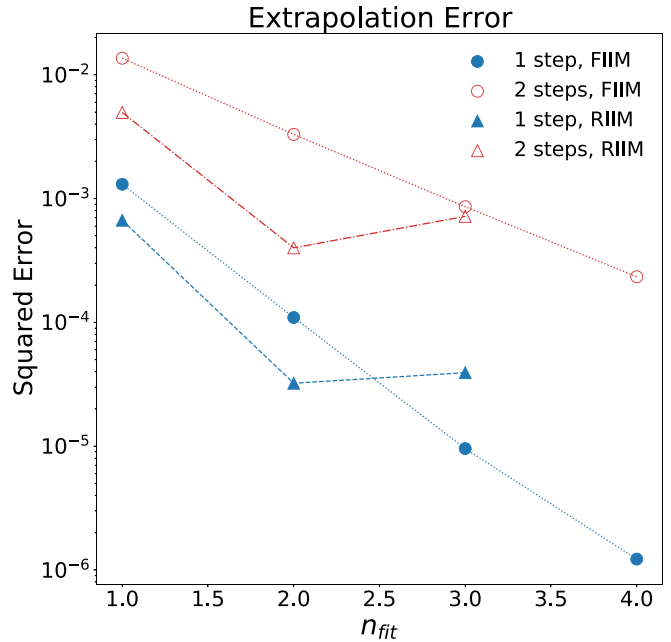
extended fits are equivalent to higher order polynomial extrapolations.

A key challenge with FIIM is that it requires a significant inflation in the gate count to achieve high precision. We proposed a method whereby identities are randomly instead of deterministically inserted. A careful choice of insertion probabilities can result in the same formal accuracy as FIIM but with far fewer gates [$(2n-1)N_c$ versus $N_c + 2(n-1)$]. This method will provide access to moderately deep circuits where FIIM is not applicable for near-term devices.

Finally, we have discussed the impact of other important sources of noise. In particular, ZNE does not remove generic nondepolarizing noise. Furthermore, large statistical noise can spoil the high-order depolarizing noise cancellation. Other techniques may be required to mitigate these sources of noise within the ZNE framework, although it is likely that these techniques can be used to mitigate amplitude damping and decoherence noise.

In the era of NISQ hardware, zero-noise extrapolation will continue to play an important role for enhancing the precision of quantum algorithms. Identity insertions provide a practical error-model agnostic and software-based approach for enhancing errors in a controlled way. The RIIM method has extended this methodology for finer control over the error scaling and will extend the efficacy of zero-noise extrapolation to moderate-depth circuits. Combined with readout error mitigation, these techniques will provide a complete package for improving the accuracy of near-term calculations on quantum devices.

[1] B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, Unfolding quantum computer readout noise, arXiv:1910.01969 [quant-ph].

[2] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[3] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, arXiv:0904.2557 [quant-ph] .

[4] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Rep. Prog. Phys. **76**, 076001 (2013).

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

[6] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, Cambridge, UK, 2013).

[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, New Yor, 2011).

[8] M. Urbanek, B. Nachman, and W. de Jong, Quantum error detection improves accuracy of chemical calculations on a quantum computer, arXiv:1910.00129 [quant-ph].

[9] J. R. Wootton and D. Loss, Repetition code of 15 qubits, Phys. Rev. A **97**, 052313 (2018).

[10] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature (London) **508**, 500 (2014).

[11] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, State preservation by repetitive error detection in a superconducting quantum circuit, Nature (London) **519**, 66 (2015).

[12] N. M. Linke, M. Gutierrez, K. A. Landsman, C. Figgatt, S. Debnath, K. R. Brown, and C. Monroe, Fault-tolerant quantum error detection, Sci. Adv. **3**, e1701074 (2017).

[13] M. Takita, A. W. Cross, A. D. Córcoles, J. M. Chow, and J. M. Gambetta, Experimental Demonstration of Fault-Tolerant State Preparation with Superconducting Qubits, Phys. Rev. Lett. **119**, 180501 (2017).

[14] J. Roffe, D. Headley, N. Chancellor, D. Horsman, and V. Kendon, Protecting quantum memories using coherent parity check codes, Quantum Sci. Technol. **3**, 035010 (2018).

[15] C. Vuillot, Is error detection helpful on IBM 5Q chips? Quantum Inf. Comput. **18**, 0949 (2018).

[16] D. Willsch, M. Willsch, F. Jin, H. De Raedt, and K. Michielsen, Testing quantum fault tolerance on small systems, Phys. Rev. A **98**, 052348 (2018).

[17] R. Harper and S. T. Flammia, Fault-Tolerant Logical Gates in the IBM Quantum Experience, Phys. Rev. Lett. **122**, 080504 (2019).

[18] 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).

[19] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski, Cloud Quantum Computing of an Atomic Nucleus, Phys. Rev. Lett. **120**, 210501 (2018).

[20] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 021050 (2017).

[21] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[22] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[23] M. Arsenijević, J. Jeknić-Dugić, and M. Dugić, Generalized Kraus operators for the one-qubit depolarizing quantum channel, Br. J. Phys. **47**, 339 (2017).

[24] L. F. Richardson and J. A. Gaunt, The deferred approach to the limit, part I, Philos. Trans. R. Soc. London A **226**, 636 (1927).

[25] A. Sid, *Practical Extrapolation Methods: Theory and Applicationsn* (Cambridge University Press, New York, 2003).

[26] IBM Research, Qiskit [https://qiskit.org].

[27] S. P. Jordan, H. Krovi, K. S. M. Lee, and J. Preskill, BQP-completeness of scattering in scalar quantum field theory, Quantum **2**, 44 (2018).

[28] S. P. Jordan, K. S. M. Lee, and J. Preskill, Quantum computation of scattering in scalar quantum field theories, Quantum Inf. Comput. **14**, 1014 (2014).

[29] S. P. Jordan, K. S. M. Lee, and J. Preskill, Quantum algorithms for quantum field theories, Science **336**, 1130 (2012).

[30] S. P. Jordan, K. S. M. Lee, and J. Preskill, Quantum algorithms for fermionic quantum field theories, arXiv:1404.7115 [hep-th].

[31] R. D. Somma, Quantum simulations of one dimensional quantum systems, Quantum Inf. Comput. **16**, 1125 (2016).

[32] A. Macridin, P. Spentzouris, J. Amundson, and R. Harnik, Electron-Phonon Systems on a Universal Quantum Computer, Phys. Rev. Lett. **121**, 110504 (2018).

[33] A. Macridin, P. Spentzouris, J. Amundson, and R. Harnik, Digital quantum computation of fermion-boson interacting systems, Phys. Rev. A **98**, 042312 (2018).

[34] N. Klco and M. J. Savage, Digitization of scalar fields for quantum computing, Phys. Rev. A **99**, 052335 (2019).

[35] H. F. Trotter, On the product of semi-groups of operators, Proc. Am. Math. Soc. **10**, 545 (1959).

[36] M. Suzuki, Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with

applications to many-body problems, Commun. Math. Phys. **51**, 183 (1976).

[37] M. Suzuki, Relationship between d-dimensional quantal spin systems and $(d + 1)$-dimensional Ising systems—Equivalence, critical exponents, and systematic approximants of the partition function and spin correlations, Prog. Theor. Phys. **56**, 1454 (1976).

[38] N. Klco, J. R. Stryker, and M. J. Savage, SU(2) non-Abelian gauge field theory in one dimension on digital quantum computers, Phys. Rev. D **101**, 074512 (2020).