

Resource Efficient Zero Noise Extrapolation 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, CA 94720, USA*

²*Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

(Dated: March 12, 2020)

In addition to readout errors, two-qubit gate noise is the main challenge for complex quantum algorithms on noisy intermediate-scale quantum (NISQ) computers. These errors are a significant challenge for making accurate calculations for quantum chemistry, nuclear physics, high energy physics, and other emerging scientific and industrial applications. There are two proposals for mitigating two-qubit gate errors: error-correcting codes and zero-noise extrapolation. This paper focuses on the latter, studying it in detail and proposing modifications to existing approaches. In particular, we propose a random identity insertion method (RIIM) that can achieve competitive asymptotic accuracy with far fewer gates than the traditional fixed identity insertion method (FIIM). For example, correcting the leading order depolarizing gate noise requires $n_{\text{CNOT}} + 2$ gates for RIIM instead of $3n_{\text{CNOT}}$ gates for FIIM. This significant resource saving may enable more accurate results for state-of-the-art calculations on near term quantum hardware.

I. INTRODUCTION

Gate and readout errors currently limit the efficacy of moderately deep circuits on existing noisy intermediate-scale quantum (NISQ) computers [1]. Readout errors can be mitigated with unfolding techniques [2]. Two-qubit gates are the most important source of gate noise and the most basic two-qubit gate is the controlled NOT operation (‘CNOT’). One strategy for mitigating these errors is to build in error correcting components into the quantum circuit. Quantum error correction [3–7] is non-trivial because qubits cannot be cloned [8–10]. As a result, there is a significant overhead in the additional number of qubits and gates required to make a circuit error-detecting or error-correcting. This has been demonstrated for simple quantum circuits [11–20], but is currently infeasible for current qubit counts and moderately deep circuits.

Another strategy for mitigating multigate errors is to find a way to vary the size of the error, measure the result at various values of the error, and then extrapolate to the zero-error result (Zero Noise Extrapolation or ZNE). With hardware level control of qubit operations, one can enlarge the size of the errors by the gate operation time [21]. Such precise hardware level control, however, is often not feasible. Instead, one can try to increase the error algorithmically by modifying the circuit operations. If the noise model is known, one can insert random Pauli gates to a circuit [22]. For Hamiltonian evolution with some general assumptions on the noise, one can rescale time [23] to amplify the noise by a desired amount. An approach that does not require knowledge of the noise model is to replace the i^{th} CNOT with

$$r_i = 2n_i + 1 \quad (1)$$

CNOT gates, for $n_i \geq 0$. The focus here is on the CNOT, but the method generalizes to any unitary operation with arbitrary $U^\dagger U$ insertions for unitary operation U . Identity insertion is illustrated in Fig. 1. Since CNOT^2 is the identity, the addition of an even number of CNOT operations should not change the circuit output, but does amplify the noise. When $n_i = n$ for all i , this is the *fixed identity insertion method* (FIIM). The application of FIIM was first proposed in Ref. [24] using a linear fit and an exponential fit were studied in Ref. [25]. Linear superpositions of enlarged noise circuits were also studied in Ref. [23], which will be similar to our results on higher order fit ZNE with FIIM. One challenge with FIIM is that it requires a large number of gates. We propose a new solution to this challenge by promoting the n_i from Eq. 1 to random variables to construct the *random identity insertion method* (RIIM).

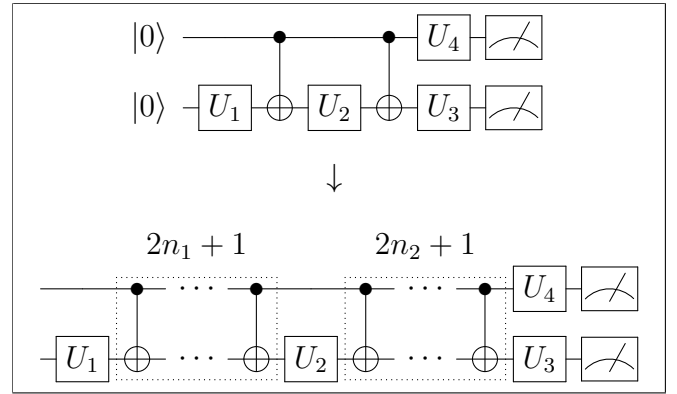


FIG. 1. An illustration of identity insertion for a generic controlled unitary operation with two qubits. The U_i represent unitary matrices and the n_i are non-negative integers.

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 non-linear fits is discussed in Sec. IV. Sections V and VI

* andrehe@lbl.gov

† bpnachman@lbl.gov

‡ wadejong@lbl.gov

§ cwbauer@lbl.gov

extend the discussion to include other sources of quantum noise as well as statistical uncertainties, respectively. 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 an intuition 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]:

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

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 $\sigma_i^{(k,l)}$ is the set of single qubit Pauli gates acting on qubits k and l .

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

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

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. Equation (3) has the clear interpretation that with probability $\epsilon^{(kl)}$, ρ is equally likely to be in any of the four possible states: $|kl\rangle \oplus \rho_{\cancel{kl}} \in \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\} \oplus \rho_{\cancel{kl}}$.

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

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

Note that in the noiseless limit $\epsilon^{(kl)} \rightarrow 0$, 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

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

Extending the pattern of Eq. (3)-(5), applying the same CNOT $r_i = 1 + 2n_i$ times in a row has the same effect as applying it once with the noise amplified by r_i

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

where the i^{th} 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

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

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 $i = 1 \dots N_c$ CNOT gates, 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

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

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 observable if the CNOT i is replaced with the depolarizing channel, and $r = 1, 3, \dots$ 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 $r = 0$

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

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 $\mathcal{O}(\epsilon^2)$ terms in Eq. (8) are subdominant (the ‘linear regime’). In this regime, one expects to remove the dominant $\mathcal{O}(\epsilon)$ terms with a linear fit so that after linear FIIM

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

where r_{max} is the maximum r value so that the circuit is still in the linear regime.

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. 2. Due to the small number and simple orientation of gates, this model can be solved completely analytically.

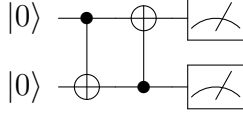


FIG. 2. An illustration of the the simple double gate circuit described in Sec. II.

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

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

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

$$r = 1 + 2n. \quad (12)$$

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

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

where

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

Suppose that one wants to measure $\langle q_0 + q_1 \rangle$, where q_i is the i^{th} qubit in Fig. 2. The result of this measurement gives

$$\begin{aligned} \langle q_0 + q_1 \rangle &= x_{\text{FIIM}}(\epsilon, 0) \\ &= 1 - (1 - \epsilon)^{N_c} \\ &= N_c \epsilon + \mathcal{O}(\epsilon^2), \end{aligned} \quad (15)$$

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 $n = 0, \dots, n_{\text{max}}$ and then extrapolates to the value $n = -1/2$ ($r = 0$). A linear fit with these data is a solution to the equation

$$Y = X\beta, \quad (16)$$

where

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

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

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

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

Taylor expanding Eq. (18) and (19) to $\mathcal{O}(\epsilon^2)$ gives

$$\hat{\beta}_1 = 2(N_c \epsilon) + (-2n - 2 + N_c^{-1})(N_c \epsilon)^2 + \mathcal{O}(N_c \epsilon^3) \quad (20)$$

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

The resulting equation is then

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

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

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

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 non-linear regime. One should therefore expect that the error grows with the largest number of CNOT gates used, which is given by $r_{\text{max}} N_c$. This can clearly be seen by rewriting the result of Eq. (23)

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

The best result is therefore obtained using a linear fit with 2 points, giving

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

A main drawback of linear FIIM is that it requires

$$r_{\text{max}} \sum_{i=1}^{N_{\text{CNOT}}} \epsilon_i \sim r_{\text{max}} N_c \epsilon \ll 1. \quad (26)$$

While this works well for circuits for which $N_{\text{CNOT}} \epsilon$ 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 $N_{\text{CNOT}} \epsilon$, the expansion breaks down for $3N_{\text{CNOT}} \epsilon$ or $5N_{\text{CNOT}} \epsilon$. 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 $\mathcal{O}((r_{\max} N_{\text{CNOT}} \epsilon)^2)$, with r_{\max} having to be at least equal to 3. Thus, for deep enough circuits where $3N_{\text{CNOT}} \epsilon \sim 1$ this method completely 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 less total CNOT gates should perform much better. Instead of inserting the same number of identity operators for every CNOT gate, suppose instead that identities were randomly inserted. This gives rise 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 :

$$\begin{aligned} \langle M \rangle(r_1, \dots, r_{N_c}) &= \left(1 - \sum_{i=1}^{N_c} r_i \epsilon_i \right) \langle M \rangle_{\text{ex}} + \sum_{i=1}^{N_c} r_i \epsilon_i \langle M \rangle_{\text{dep}_i} \\ &\quad + \mathcal{O}\left(\left(\sum_{i=1}^{N_c} r_i \epsilon_i \right)^2 \right), \end{aligned} \quad (27)$$

Next, the $r_i = 1 + 2n_i$ in Eq. 1 are promoted to random variables. For example, one could choose $n_i \sim \text{Poisson}(\nu)$. As $\nu \rightarrow 0$, 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 $\mathcal{O}(\epsilon)$ term with only $N_{\text{CNOT}} + 2$ gates instead of $3N_{\text{CNOT}}$ 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:

$$\begin{aligned} \langle \langle M \rangle \rangle(\nu) &= \sum_{n_1=0}^{\infty} \cdots \sum_{n_{N_{\text{CNOTS}}}=0}^{\infty} \prod_{i=1}^{N_c} \text{Pr}(n_i|\nu) \\ &\quad \times \left\{ \left[1 - \sum_i \epsilon_i (1 + 2n_i) \right] \langle M \rangle_{\text{ex}} \right. \\ &\quad \left. + \sum_i \epsilon_i (1 + 2n_i) \langle M \rangle_{\text{dep}_i} \right. \\ &\quad \left. + \mathcal{O}\left(\left(\sum_i (1 + 2n_i) \epsilon_i \right)^2 \right) \right\}. \end{aligned} \quad (28)$$

Since each gate is independently sampled, one can replace

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

which immediately reduces Eq. (28) to

$$\begin{aligned} \langle \langle M \rangle \rangle(\rho) &= \left[1 - \rho \sum_i \epsilon_i \right] \langle M \rangle_{\text{ex}} + \rho \sum_i \epsilon_i \langle M \rangle_{\text{dep}_i} \\ &\quad + \mathcal{O}\left(\left(\rho \sum_i \epsilon_i \right)^2 \right), \end{aligned} \quad (30)$$

where $\rho = 1 + 2\nu$. Thus, Eq. (30) has the same feature as FIIM, only the integer n is now replaced by the non-integer value $\nu \geq 0$. 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 $3N_{\text{CNOT}} \epsilon$, $5N_{\text{CNOT}} \epsilon$, etc., but only for $\rho N_{\text{CNOT}} \epsilon$, where one can choose different values of ν to get a reasonable fit region without making ρ too far from unity.

IV. NON-LINEAR FITS IN THE DEPOLARIZING NOISE MODEL

So far we have only discussed linear fits and showed that they can eliminate the $\mathcal{O}(\epsilon)$ 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 method, 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

$$\begin{aligned} \langle M \rangle(1) &= \langle M \rangle_{\text{ex}} + N_{\text{CNOT}} \epsilon \left[\sum_i \langle M \rangle_{\text{dep}_i} - \langle M \rangle_{\text{ex}} \right] \\ &\quad + \mathcal{O}(\epsilon^2) \\ \langle M \rangle(3) &= \langle M \rangle_{\text{ex}} + 3N_{\text{CNOT}} \epsilon \left[\sum_i \langle M \rangle_{\text{dep}_i} - \langle M \rangle_{\text{ex}} \right] \\ &\quad + \mathcal{O}(\epsilon^2). \end{aligned} \quad (31)$$

One can immediately see that the linear combination

$$\frac{3}{2} \langle M \rangle(1) - \frac{1}{2} \langle M \rangle(3) = \langle M \rangle_{\text{ex}} + \mathcal{O}(\epsilon^2). \quad (32)$$

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

Generalizing these results one can immediately obtain linear combinations that remove higher order terms in ϵ as well. This fact has been observed before [23], and is an application of the Richardson extrapolation [26, 27].

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 $r = 1, 3, \dots, r_{\max}$ one can eliminate all terms up to $\mathcal{O}(\epsilon^{n_{\max}+1})$ with

$$n_{\max} = \frac{r_{\max} - 1}{2}. \quad (33)$$

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 $\mathcal{O}(\epsilon^{n_{\max}+1})$

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

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

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

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

$$\begin{aligned} \langle M \rangle(r) &= (1-\epsilon)^{N_c r} \langle M \rangle_{\text{ex}} \\ &+ (1-\epsilon)^{(N_c-1)r} [1 - (1-\epsilon)^r] \sum_i \langle M \rangle_{\text{dep}_i} \\ &+ (1-\epsilon)^{(N_c-2)r} [1 - (1-\epsilon)^r]^2 \sum_{i_1, i_2} \langle M \rangle_{\text{dep}_{i_1 i_2}} \\ &+ \dots \\ &+ [1 - (1-\epsilon)^r]^{N_c} \sum_{i_1, \dots, i_{N_c}} \langle M \rangle_{\text{dep}_{i_1, \dots, i_{N_c}}} \\ &= \langle M \rangle_{\text{ex}} - f_{N_c}(r, \epsilon) \langle M \rangle_{\text{ex}} \\ &+ [f_{N_c}(r, \epsilon) - f_{N_c-1}(r, \epsilon)] \sum_i \langle M \rangle_{\text{dep}_i} \\ &+ [f_{N_c}(r, \epsilon) - f_{N_c-2}(r, \epsilon)] \sum_{i_1, i_2} \langle M \rangle_{\text{dep}_{i_1 i_2}} \\ &+ \dots \\ &+ f_{N_c}(r, \epsilon)^{N_c} \sum_{i_1, \dots, i_{N_c}} \langle M \rangle_{\text{dep}_{i_1, \dots, i_{N_c}}}, \end{aligned} \quad (36)$$

where

$$f_n(r, \epsilon) = 1 - (1-\epsilon)^{nr}. \quad (37)$$

It is important to remember that the values of $\langle M \rangle_{\text{ex}}$, $\langle M \rangle_{\text{dep}_i}$, $\langle M \rangle_{\text{dep}_{i_1, \dots, i_{N_c}}}$ 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 $\mathcal{O}(\epsilon^{n_{\max}})$ have to cancel for each line separately.

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

$$\sum_{n=0}^{n_{\max}} a(n) f_k(1+2n, \epsilon) = 1 + \mathcal{O}(\epsilon^{n_{\max}+1}), \quad (38)$$

for all values of k . After some lines of algebra, one can show that this is indeed possible with the coefficients [23]

$$\begin{aligned} a(i) &= \prod_{j=0, j \neq i}^{n_{\max}} \frac{(1+2j)}{2(j-i)} \\ &= \frac{2^{-2n_{\max}} (-1)^i (1+2n_{\max})!}{i! (1+2i) n_{\max}! (n_{\max}-i)!}, \end{aligned} \quad (39)$$

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

$$\max_i [a(i)] \sim a(n_{\max}/2) \sim \frac{2^{n_{\max}+1}}{n_{\max}} \quad (40)$$

To summarize, by using values $\langle M \rangle(r)$ with $r = 1, 3, \dots, r_{\max}$ 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 $\mathcal{O}(\epsilon^{n_{\max}+1})$.

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 $r = 1, 3, \dots, r_{\max}$. A polynomial fit uses the same setup for the linear fit, with Eq. (16), only now X and β are augmented:

$$X = \begin{pmatrix} 0^{n_{\text{fit}}} & \dots & 0 & 1 \\ 1^{n_{\text{fit}}} & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ n_{\text{max}}^{n_{\text{fit}}} & \dots & n_{\text{max}} & 1 \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_{n_{\text{fit}}} \\ \vdots \\ \beta_1 \\ \beta_0 \end{pmatrix}, \quad (41)$$

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

$$\langle M \rangle_{\text{FIIM}[n_{\text{fit}}, n_{\text{max}}]} = \sum_{i=1}^{n_{\text{fit}}} \hat{\beta}_i x^i, \quad (42)$$

to $x = -\frac{1}{2}$ removes the $\mathcal{O}(\epsilon^{n_{\max}})$ component of the depolarizing error when $n_{\text{fit}} = n_{\max}$. Both the polynomial fit and the superposition from Eq. (39) give rise to the same linear combinations of the values measured at various values of r . One can show this with some symbolic

manipulation:

$$\begin{aligned}
\langle M \rangle_{\text{FIIM}[n_{\text{fit}}, n_{\text{max}}]} &= \sum_{i=0}^{n_{\text{fit}}} \hat{\beta}_i \left(-\frac{1}{2}\right)^i \\
&= \sum_{i=0}^{n_{\text{fit}}} \sum_{j=0}^{n_{\text{max}}} ((X^T X)^{-1} X^T)_{ij} Y_j \left(-\frac{1}{2}\right)^i \\
&= \sum_{j=0}^{n_{\text{max}}} \left[\sum_{i=0}^{n_{\text{fit}}} ((X^T X)^{-1} X^T)_{ij} \left(-\frac{1}{2}\right)^i \right] Y_j \\
&= \sum_{n=0}^{n_{\text{max}}} \left[\sum_{i=0}^{n_{\text{fit}}} ((X^T X)^{-1} X^T)_{in} \left(-\frac{1}{2}\right)^i \right] \langle M \rangle (1+2n) \\
&\equiv \sum_{n=0}^{n_{\text{max}}} \tilde{a}(n) \langle M \rangle (1+2n). \tag{43}
\end{aligned}$$

We have verified that the $\tilde{a}(n)$ in Eq. (43) are equivalent to the $a(n)$ in Eq. (39).

B. RIIM method

The RIIM method one 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. (36) from FIIM:

$$\begin{aligned}
\langle M \rangle (r_1, \dots, r_{N_c}) &\tag{44} \\
&= \prod_j (1 - \epsilon)^{r_j} \left[\langle M \rangle_{\text{ex}} \right. \\
&\quad + \sum_i \frac{f_1(r_i, \epsilon)}{(1 - \epsilon)^{r_i}} \langle M \rangle_{\text{dep}_i} \\
&\quad + \sum_{i_1 > i_2} \frac{f_1(r_{i_1}, \epsilon)}{(1 - \epsilon)^{r_{i_1}}} \frac{f_1(r_{i_2}, \epsilon)}{(1 - \epsilon)^{r_{i_2}}} \langle M \rangle_{\text{dep}_{i_1 i_2}} \\
&\quad + \dots \\
&\quad \left. + \sum_{i_1 > \dots > i_{N_c}} \frac{f_1(r_{i_1}, \epsilon)}{(1 - \epsilon)^{r_{i_1}}} \dots \frac{f_1(r_{i_{N_c}}, \epsilon)}{(1 - \epsilon)^{r_{i_{N_c}}}} \langle M \rangle_{\text{dep}_{i_1 \dots i_{N_c}}} \right].
\end{aligned}$$

To eliminate all terms up to order $\epsilon^{n_{\text{max}}}$, one needs to include all possible combinations of r_1, \dots, r_{N_c} with $\sum_i r_i = N_c + 2n_{\text{max}}$. To write a generic solution we require a bit of new notation. Denote by $O(\{e_1, \dots, e_n\})$ the sum of all operators with the r_i given by permutations of 1 and the various values of $e_i = 3, 5, 7, \dots$. So

$$\begin{aligned}
O(\{\}) &= O(1, \dots, 1) \tag{45} \\
O(\{e_1\}) &= O(e_1, 1, \dots, 1) + O(1, e_1, \dots, 1) + \dots \\
O(\{e_1, e_2\}) &= O(e_1, e_2, 1, \dots, 1) + O(e_1, 1, e_2, \dots, 1) + \dots
\end{aligned}$$

and so on.

To eliminate all terms up to $\epsilon^{n_{\text{max}}}$ one include all operators $O(\{e_1, \dots, e_n\})$ with $\sum_i e_i \leq 2n_{\text{max}} + n$, each with its own coefficient. One then determine the coefficients by

demanding that all terms up to $\epsilon^{n_{\text{max}}}$ vanish. So for example, to eliminate the linear term in ϵ one include include the operator $O(\{\})$ and $O(\{3\})$. Solving the equations

$$a_{\{\}} O(\{\}) + a_{\{3\}} O(\{3\}) = 0 + \mathcal{O}(\epsilon^2) \tag{46}$$

with

$$a_{\{\}} = 1 - a_{\{3\}} N_c \tag{47}$$

Solving this equation, one finds

$$a_{\{3\}} = -\frac{1}{2}, \tag{48}$$

which again reproduces the result of the linear fit discussed in Section III. To eliminate the linear and quadratic term in ϵ once includes the operators $O(\{\})$, $O(\{3\})$, $O(\{5\})$ and $O(\{3, 3\})$, and solves the equation

$$a_{\{\}} O(\{\}) + a_{\{3\}} O(\{3\}) + a_{\{5\}} O(\{5\}) + a_{\{3,3\}} O(\{3, 3\}) = 0 + \mathcal{O}(\epsilon^3), \tag{49}$$

again with the constraint

$$a_{\{\}} = 1 - a_{\{3\}} N_c - a_{\{5\}} N_c - a_{\{3,3\}} \binom{N_c}{2}. \tag{50}$$

Solving the resulting set of equations gives

$$a_{\{3\}} = -\frac{N_c + 4}{4}, \quad a_{\{5\}} = \frac{3}{8}, \quad a_{\{3,3\}} = \frac{1}{4}. \tag{51}$$

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 $n_{\text{max}} = 1, 2, 3, 4$ in Table II. These results allow to remove depolarizing noise with corrections arising at $\epsilon^{n_{\text{max}}+1}$ using $N_c + 2n_{\text{max}}$ gates. This should be compared with the FIIM method where the same noise reduction requires $(2n_{\text{max}} + 1)N_c$ gates.

For relatively shallow circuits, one could feasibly perform the measurements for all permutations required for $O(\{e_1, \dots, e_n\})$. For example, to remove the $\mathcal{O}(\epsilon)$ 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 $O(\{e_1, \dots, e_n\})$, randomly pick one of the $N_{\{e_1, \dots, e_n\}}$ operations.

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

Method	Remainder	# of CNOTs
FIIM	$\mathcal{O}(\epsilon^n)$	$(2n - 1)n_{\text{CNOT}}$
RIIM	$\mathcal{O}(\epsilon^n)$	$n_{\text{CNOT}} + 2(n - 1)$

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

V. BEYOND THE DEPOLARIZING NOISE MODEL

Equation (2) introduced the full Krauss 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 non-zero δ_{ij} , though the non-depolarizing 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 $\mathcal{O}(\epsilon^{n_{\max}})$, they do not remove the $\mathcal{O}(\delta)$ term. This means that it is not useful to go beyond $\mathcal{O}(\epsilon^2)$, 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 [23] 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 $\langle M \rangle_{\text{ex}}$ 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 $\langle M \rangle_{\text{ex}}(r)$ can only be determined up to a statistical uncertainty

$$\Delta(r) \sim \frac{1}{\sqrt{n_{\text{meas}}}}, \quad (52)$$

where n_{meas} denotes the number of measurements that are performed in the measurement of each value $\langle M \rangle(R)$. Adding the various contributions arising from the linear superposition in quadrature, one finds that the error from

statistical uncertainties is given by

$$\begin{aligned} \Delta_{\text{stat}} &= \frac{1}{\sqrt{n_{\text{meas}}}} \sqrt{\sum_{n=0}^{n_{\max}} [a(n)]^2} \\ &\sim \frac{1}{\sqrt{n_{\text{meas}}}} \frac{2^{n_{\max}}}{n_{\max}}, \end{aligned} \quad (53)$$

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. (40).

This means that the final uncertainty in the FIIM and RIIM methods are given by

$$\Delta_{\text{FIIM/RIIM}}[\epsilon, \delta; n_{\max}, n_{\text{meas}}] \sim \max[\delta, \epsilon^{n_{\max}}, \Delta_{\text{stat}}] \quad (54)$$

VII. NUMERICAL RESULTS

We use `qiskit` [28] to simulate the quantum circuits described below and demonstrate FIIM and RIIM. Section VII A studies the simple CNOT only circuit from Fig. 2 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. 2 was particularly useful because of its analytical tractability. In particular, because one can compute the expectation values analytically, it is possible to consider the $n_{\text{shots}} \rightarrow \infty$ limit. In this section, we use a slight modification of this simple circuit, which uses 4 CNOT gates, which are started in the initial state $|10\rangle$. In the noiseless limit, the final state is

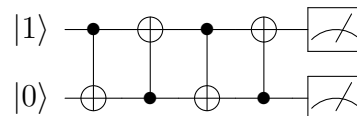


FIG. 3. A simple circuit with 4 CNOT gates used in this section.

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.

Fig. 4 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 $\mathcal{O}(\epsilon^4)$ 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.

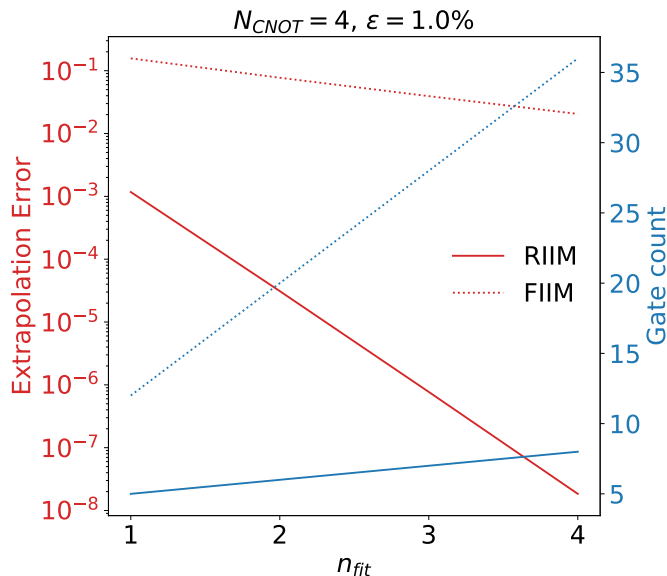


FIG. 4. Numerical results based on the higher-order fits described in Sec. IV using the four CNOT gate version of the model presented in Fig. 2. The horizontal axis is the order of the depolarizing error that is being removed. The left axis is the error on $\langle \sum_{i=0}^{N_c} q_i \rangle$ as ϵ is extrapolated to zero. The right axis is the number of gates requires to make the correction. Only depolarizing noise is considered and $n_{\text{shots}} = \infty$.

`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.

In Fig. 5, we show the result where the measured observable is the expected value of the output string, converting from binary numbers to integers ($00 \rightarrow 0, 01 \rightarrow 1, 10 \rightarrow 2, 11 \rightarrow 3$). In the noiseless limit, the expectation value is 3, corresponding to $|11\rangle$ Fixed identity insertions (but no corrections yet) are applied up to $r_{\text{max}} = 31$. 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 towards the $|00\rangle$ state) as well as the depolarizing noise, which drives the system to the completely mixed state.

Fig. 6 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 RIIM are higher in the case of a full noise model which has non-depolarizing elements. RIIM performs as well or better than FIIM in both noise models. The minimum extrapolation error is achieved at $n_{\text{fit}} = 2$. This can be understood in the context of Eq. (54) in Section VI with the parameters of $\epsilon = 1\%$ and $n_{\text{meas}} = 10^7$. At $n_{\text{fit}} = 1$, the dominant error $\Delta_{\text{FIIM/RIIM}}$ is determined by ϵ rather than the statistical uncertainty. However, as $n_{\text{fit}} = 2$, the statistical error Δ_{stat} begins to exceed ϵ^2 , and by $n_{\text{fit}} = 3$ the dominant error becomes the statistical error, which prevents further reduction of extrapolation

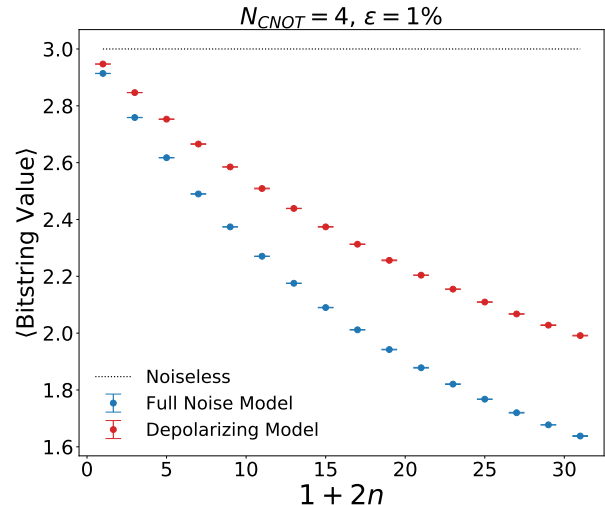


FIG. 5. Numerical results from simulating the 4-CNOT circuit in noiseless and noisy simulators using `qiskit`. The vertical axis shows the expectation value of the measured observable. The horizontal axis displays r , or $1 + 2n$. Noisy simulations include both full and purely depolarizing cases. The number of shots for each point is 10^7 , with a standard deviation of 10^{-3} .

error and leads to the exponential scaling of the error as n_{fit} is increased further. Note that RIIM is only used to eliminate errors up to $\mathcal{O}(\epsilon^4)$, as the circuit only contains 4 CNOTs.

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, time evolution is given by

$$|\psi(t)\rangle = e^{-iHt} |\psi(0)\rangle, \quad (55)$$

where

$$H = \frac{1}{2}(\hat{x}^2 + \hat{p}^2) \equiv H_x + H_p. \quad (56)$$

The Hamiltonian in Eq. (56) can be implemented on a digital quantum computer by discretizing the possible values of x to be $-x_{\text{max}}, -x_{\text{max}} + \delta_x, \dots, x_{\text{max}} - \delta_x, x_{\text{max}}$, where $\delta_x = 2x_{\text{max}}/(2^{n_{\text{qubits}}} - 1)$ 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 non-interacting scalar field theory [29–36]. 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

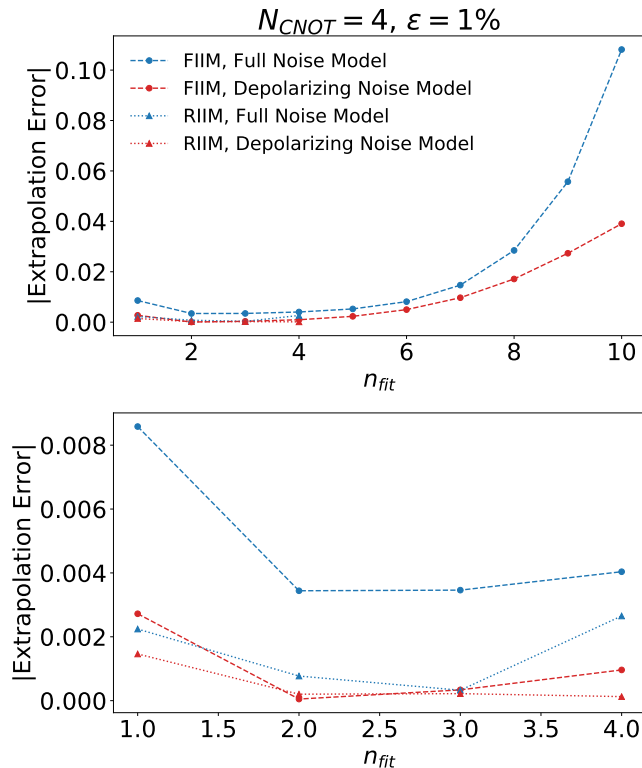


FIG. 6. Numerical results from simulating the 4-CNOT circuit and applying FIIM and RIIM extrapolation in a noisy simulator. The lower plot displays a reduced range of the data in the upper plot for detail. Extrapolation error is 0 for noiseless simulation. The number of shots for each point is 10^7 .

Suzuki-Trotter expansion [37–39]:

$$e^{-i(H_x+H_p)t} \approx \left[e^{-iH_x \frac{t}{n}} e^{-iH_p \frac{t}{n}} \right]^n \equiv \left[U_n^{(H)}(t/n) \right]^n \quad (57)$$

The approximation in Eq. (57) 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. 7.

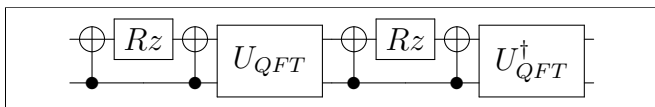


FIG. 7. Circuit diagram for a single Trotter step of the time evolution of the harmonic oscillator hamiltonian for two qubits. The total number of CNOT operations for the quantum Fourier transform U_{QFT} on 2 qubits is 5, giving a total of 14 CNOT operations. However, one CNOT operation from each of the U_{QFT} is cancelling a CNOT from the rest of the circuit, giving a total of 10 CNOT operators per Trotter step.

Time evolution of the ground state of the Harmonic

oscillator gives

$$|\psi_0(t)\rangle = e^{-iHt} |\psi_0(0)\rangle = e^{-iE_0 t} |\psi_0(0)\rangle, \quad (58)$$

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

$$\langle \psi_0(0) | \psi_0(t) \rangle = 1. \quad (59)$$

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 on the state $|0\rangle$. U_{State} is implemented with 2 CNOT gates.

$$|\psi_0(0)\rangle = U_{\text{State}} |0\rangle. \quad (60)$$

Thus, the overlap can be written as

$$\lim_{n \rightarrow \infty} \langle 0 | U_{\text{State}}^\dagger \left[U_n^{(H)}(t/n) \right]^n U_{\text{State}} |0\rangle = 1. \quad (61)$$

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

$$\langle 0 | U_{\text{State}}^\dagger \left[U_n^{(H)}(t/n) \right]^n U_{\text{State}} |0\rangle = 1 + \mathcal{O}(t^2/n^2). \quad (62)$$

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 2 qubits, corresponding to 4 discrete values of x . In this case the CNOT count is given by

$$N_c = 4 + 10n. \quad (63)$$

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 [40].

Fig. 8 presents the result of one and two Trotter steps, corrected with RIIM and with FIIM up to $\mathcal{O}(\epsilon^2)$. 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.

Fig. 9 compares the error obtained from the FIIM and RIIM extrapolations over different values of n_{max} . The extrapolated error from RIIM up to $\mathcal{O}(\epsilon^2)$ is lower than any of the errors obtained through FIIM for all values of n_{fit} in the 1-step case and in the 2-step case.

VIII. CONCLUSIONS

We have performed a detailed study of zero noise extrapolation for correcting gate errors in quantum circuits.

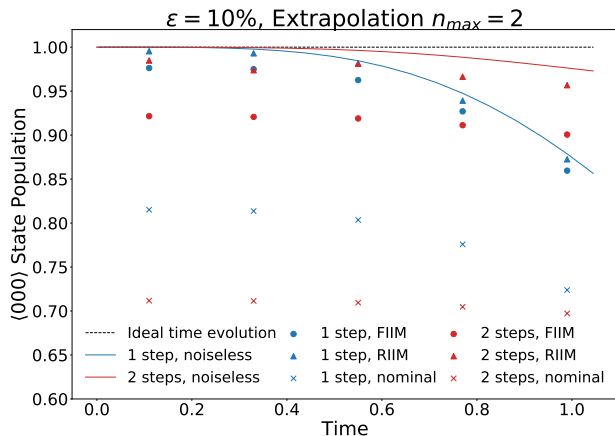


FIG. 8. The expectation value of the observable in Eq. (61) as a function of time for various numbers of Trotter steps and noise mitigation techniques. The purely depolarizing noise model is used for this simulation. In the absence of Trotterization error, the observable should be unity, independent of t .

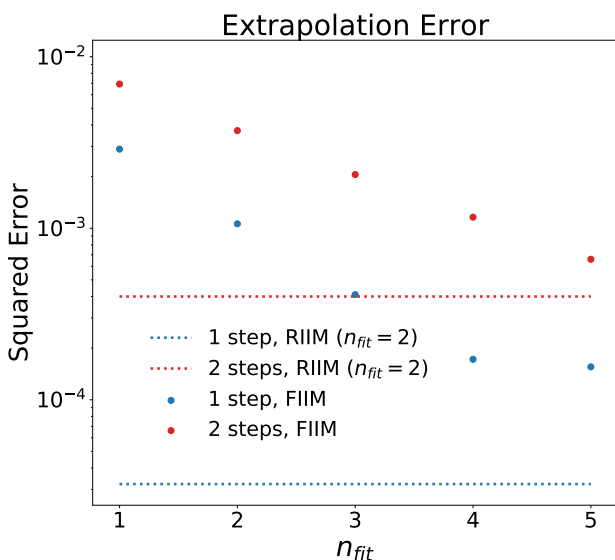


FIG. 9. The squared error of the observable value extrapolated using FIIM plotted against n_{max} . The dashed lines indicate the error using second-order RIIM.

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 make 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 propose a new 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 $[(2n - 1)n_{\text{CNOT}} \text{ versus } n_{\text{CNOT}} + 2(n - 1)]$. 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 non-depolarizing noise. Furthermore, large shot noise can spoil the high-order depolarizing noise cancellation. New techniques may be required to mitigate these sources of noise within the ZNE framework.

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 new 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.

ACKNOWLEDGMENTS

We would like to thank Andrew Christensen, Yousef Hindy, Mekena Metcalf, John Preskill, Miro Urbanek, and Will Zeng for useful discussions. This work is supported by the U.S. Department of Energy, Office of Science under contract DE-AC02-05CH11231. In particular, support comes from Quantum Information Science Enabled Discovery (QuantISED) for High Energy Physics (KA2401032) and the Office of Advanced Scientific Computing Research (ASCR) through the Accelerated Research for Quantum Computing Program.

[1] J. Preskill, *Quantum* **2**, 79 (2018).
 [2] B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, (2019), [arXiv:1910.01969](https://arxiv.org/abs/1910.01969) [quant-ph].
 [3] D. Gottesman, (2009), [arXiv:0904.2557](https://arxiv.org/abs/0904.2557) [quant-ph].
 [4] S. J. Devitt, W. J. Munro, and K. Nemoto, *Reports on*

Progress in Physics **76**, 076001 (2013).
 [5] B. M. Terhal, *Rev. Mod. Phys.* **87**, 307 (2015).
 [6] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (2013).
 [7] M. A. Nielsen and I. L. Chuang, *Quantum Computation*

- and *Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).
- [8] J. L. Park, *Found. Phys.* **1**, 23 (1970).
- [9] W. K. Wootters and W. H. Zurek, *Nature* **299**, 802 (1982).
- [10] D. Dieks, *Phys. Lett. A* **92**, 271 (1982).
- [11] B. N. Miroslav Urbanek and W. A. de Jong, (2019), [arXiv:1910.00129 \[quant-ph\]](https://arxiv.org/abs/1910.00129).
- [12] J. R. Wootton and D. Loss, *Phys. Rev. A* **97**, 052313 (2018).
- [13] 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, *Nature* **508**, 500 (2014).
- [14] 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, *Nature* **519**, 66 (2015).
- [15] N. M. Linke, M. Gutierrez, K. A. Landsman, C. Figgatt, S. Debnath, K. R. Brown, and C. Monroe, *Sci. Adv.* **3**, e1701074 (2017).
- [16] M. Takita, A. W. Cross, A. D. Córcoles, J. M. Chow, and J. M. Gambetta, *Phys. Rev. Lett.* **119**, 180501 (2017).
- [17] J. Roffe, D. Headley, N. Chancellor, D. Horsman, and V. Kendon, *Quantum Sci. Technol.* **3**, 035010 (2018).
- [18] C. Vuillot, *Quantum Inf. Comput.* **18**, 0949 (2018).
- [19] D. Willsch, M. Willsch, F. Jin, H. De Raedt, and K. Michielsen, *Phys. Rev. A* **98**, 052348 (2018).
- [20] R. Harper and S. T. Flammia, *Phys. Rev. Lett.* **122**, 080504 (2019).
- [21] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, *Nature* **567**, 491 (2019).
- [22] Y. Li and S. C. Benjamin, *Phys. Rev. X* **7**, 021050 (2017).
- [23] K. Temme, S. Bravyi, and J. M. Gambetta, *Phys. Rev. Lett.* **119**, 180509 (2017).
- [24] 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, *Phys. Rev. Lett.* **120**, 210501 (2018).
- [25] S. Endo, S. C. Benjamin, and Y. Li, *Phys. Rev. X* **8**, 031027 (2018).
- [26] L. F. Richardson and J. A. Gaunt, *Philosophica Transactions of the Royal Society of London. Series A* **226**, 636 (1927).
- [27] A. Sid., *Practical extrapolation methods: Theory and applications* (Cambridge University Press, New York, NY, USA, 2003).
- [28] IBM Research, “Qiskit,” <https://qiskit.org> (2019).
- [29] S. P. Jordan, H. Krovi, K. S. M. Lee, and J. Preskill, (2017), [arXiv:1703.00454 \[quant-ph\]](https://arxiv.org/abs/1703.00454).
- [30] S. P. Jordan, K. S. M. Lee, and J. Preskill, (2011), [*Quant. Inf. Comput.* 14,1014(2014)], [arXiv:1112.4833 \[hep-th\]](https://arxiv.org/abs/1112.4833).
- [31] S. P. Jordan, K. S. M. Lee, and J. Preskill, *Science* **336**, 1130 (2012), [arXiv:1111.3633 \[quant-ph\]](https://arxiv.org/abs/1111.3633).
- [32] S. P. Jordan, K. S. M. Lee, and J. Preskill, (2014), [arXiv:1404.7115 \[hep-th\]](https://arxiv.org/abs/1404.7115).
- [33] R. D. Somma, *Quantum Info. Comput.* **16**, 1125?1168 (2016).
- [34] A. Macridin, P. Spentzouris, J. Amundson, and R. Harnik, *Phys. Rev. Lett.* **121**, 110504 (2018).
- [35] A. Macridin, P. Spentzouris, J. Amundson, and R. Harnik, *Phys. Rev.* **A98**, 042312 (2018), [arXiv:1805.09928 \[quant-ph\]](https://arxiv.org/abs/1805.09928).
- [36] N. Klco and M. J. Savage, *Phys. Rev.* **A99**, 052335 (2019), [arXiv:1808.10378 \[quant-ph\]](https://arxiv.org/abs/1808.10378).
- [37] H. F. Trotter, *Proceedings of the American Mathematical Society* **10**, 545 (1959).
- [38] M. Suzuki, *Communications in Mathematical Physics* **51**, 183 (1976).
- [39] M. Suzuki, *Progress of Theoretical Physics* **56**, 1454 (1976).
- [40] N. Klco, J. R. Stryker, and M. J. Savage, (2019), [arXiv:1908.06935 \[quant-ph\]](https://arxiv.org/abs/1908.06935).

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	$\frac{29}{64}$	$\frac{3}{32}$	$\frac{1}{16}$

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