Folding-Free ZNE: A Comprehensive Quantum Zero-Noise Extrapolation Approach for Mitigating Depolarizing and Decoherence Noise

Hrushikesh Pramod Patil
NC State University
Raleigh, NC, USA
hpatil2@ncsu.edu

Peiyi Li
NC State University
Raleigh, NC, USA
pli11@ncsu.edu

Ji Liu
Argonne National Laboratory
Lemont, IL, USA
ji.liu@anl.gov

Huiyang Zhou
NC State University
Raleigh, NC, USA
hzhou@ncsu.edu

***Abstract*—Quantum computers in the NISQ era are prone to noise. A range of quantum error mitigation techniques has been proposed to address this issue. Zero-noise extrapolation (ZNE) stands out as a promising one. ZNE involves increasing the noise levels in a circuit and then using extrapolation to infer the zero noise case from the noisy results obtained. This paper presents a novel ZNE approach that does not require circuit folding or noise scaling to mitigate depolarizing and/or decoherence noise.**

To mitigate depolarizing noise, we propose leveraging the extreme/infinite noisy case, which allows us to avoid circuit folding. Specifically, the circuit output with extreme noise becomes the maximally mixed state. We show that using circuit-reliability metrics, simple linear extrapolation can effectively mitigate depolarizing noise. With decoherence noise, different states decay into the all-zero state at a rate that depends on the number of excited states and time. Therefore, we propose a state- and latency-aware exponential extrapolation that does not involve folding or scaling. When dealing with a quantum system affected by both decoherence and depolarizing noise, we propose to use our two mitigation techniques in sequence: first applying decoherence error mitigation, followed by depolarizing error mitigation.

A common limitation of ZNE schemes is that if the circuit of interest suffers from high noise, scaling-up noise levels could not provide useful data for extrapolation. We propose using circuit-cut techniques to break a large quantum circuit into smaller sub-circuits to overcome this limitation. This way, the noise levels of the sub-circuits are lower than the original circuit, and ZNE can become more effective in mitigating their noises.

I. Introduction

Recent years have seen exciting advances in the development of quantum computers. However, current NISQ (Noisy Intermediate Scale Quantum) devices are highly susceptible to noise and have limited numbers of qubits with constrained connectivity. Imperfections in hardware and the interaction with the environment inevitably lead to computations riddled with errors and noise. Noise in quantum computers is the fundamental challenge facing the reliable execution of quantum algorithms. To deal with this problem, quantum error correction (QEC) methods along with fault-tolerant quantum computation have been proposed. Although the physical error rates of quantum devices are approaching the critical thresholds, where QEC can achieve lower logical error rates than the physical one, the required number of qubits is overwhelming and beyond the capacity of current devices.

Quantum error mitigation (QEM) can be viewed as an alternative or complementary to QEC. Generally, QEM methods involve multiple runs of a circuit with a few parameters, such as the circuit depth, pulse length, etc., being varied. The results of these multiple runs are post-processed to obtain a noise-mitigated result. Popular error mitigation methods include Zero Noise Extrapolation (ZNE) [10], [12], [22], Probabilistic Error Cancellation (PEC) [22], Clifford Data Regression (CDR) [7], etc. Machine learning-based quantum error mitigation approaches like QRAFT (Quantum circuit Reversal for Attaining Full Truth) [19] have also been proposed recently.

ZNE was first introduced in [12] and [22]. The basic idea is to deliberately run the circuit with different noise levels and use the measured results to extrapolate the noise-free result.

This paper improves the efficacy of ZNE with the following novel designs. First, we propose to change the way how the extrapolation is performed such that we can leverage the extreme noise case. The quantum states under extreme noise are independent of the circuit structure but dependent on the noise source. With depolarizing noise, the quantum state under extreme noise becomes the maximally mixed state. In existing ZNE schemes, the extrapolation is based on noise levels or circuit folding factors. As a result, it is impractical to incorporate extreme noise cases, whose noise levels/folding factors are infinite.

Second, we choose to use different extrapolation algorithms to mitigate different noises. We show that using circuit-reliability metrics, simple linear extrapolation is sufficient to mitigate depolarizing noise, while exponential extrapolation is needed for mitigating decoherence noise. For systems suffering from both depolarizing and decoherence noise, we propose a serial approach, which first mitigates the decoherence noise followed by depolarization error mitigation. In all these cases, there is no need for circuit folding for extrapolation.

Third, rather than scaling up the noise levels for extrapolation, we propose to scale down the noise levels using circuit cutting techniques, which break a large circuit into smaller

sub-circuits and rely on classical processing to combine the execution results from the sub-circuits. This way, our scheme can tackle the long-standing problem of all existing ZNE schemes, which become ineffective for circuits with high noise levels as further scaling up the noise fails to produce meaningful data for extrapolation.

The remainder of the paper is organized as follows. Section II presents the background on quantum noises and discusses the related work. Section III motivates our approach. Section IV details our proposed schemes. Sections V and VI discuss the experimental methodology and the results, respectively. Section VII concludes.

II. Background

A. Noise in Quantum Computers

Noise in quantum computers can lead to errors during computation. Depolarizing, decoherence or thermal relaxation, state preparation, measurement, and cross-talk errors are the most common errors in a quantum computer. Among them, state preparation and measure (SPAM) errors can be suppressed with careful calibration and post-processing based on the calibration data (e.g., an inverse of the measurement error matrix). Cross-talk errors depend on the device topology and can be mitigated by scheduling two-qubit gates judiciously in a circuit [15]. In this work, we focus on depolarizing and decoherence noise.

The depolarizing error can be modeled by the following error channel,

$$E(\rho) = (1-p)\rho + p\frac{I}{2^n} \tag{1}$$

where ρ is the density matrix of a quantum state, p is the depolarizing probability, $E(\rho)$ is noisy state after the channel, I is the identity matrix, and n is the number of qubits [17]. Conceptually, a depolarizing error channel results in a weighted mixture of the noise-free quantum state and the maximally mixed state. With infinite noise, $p = 1$ and $E(\rho) = \frac{I}{2^n}$, which is the maximally mixed state.

Thermal relaxation to the ground state is parameterized by T_1 and T_2 decoherence times [27]. The relaxation time T_1 dictates the decay rate of a qubit in state $|1\rangle$ to $|0\rangle$. It is given by,

$$P_{|1\rangle}(t) = P_{|1\rangle}(0)e^{\frac{-t}{T_1}} \tag{2}$$

where $P_{|1\rangle}(t)$ is the probability of the state $|1\rangle$ at time t. We can see that at $t = \infty$, the probability of the excited state approaches zero. Similarly, T_2 relaxation time dictates the decay rate of the phase information of a quantum state and has a similar exponential relationship.

Based on Eq. 2, we can further derive the decoherence impact on a multi-qubit state $|q_1 q_2 ... q_n\rangle$. Assuming that each qubit decays independently [6], the decay rate would have an exponential relationship with the number of 1s (i.e., excited qubits) in the state. For example, for a 2-qubit state $|11\rangle$, the probability of it staying in $|11\rangle$ at time t can be computed as follows.

$$P_{|11\rangle}(t) = Prob_{(q_1=|1\rangle)}(t) * Prob_{(q_2=|1\rangle)}(t)$$
$$= (P_{|1\rangle}(0)e^{\frac{-t}{T_1}})(P_{|1\rangle}(0)e^{\frac{-t}{T_1}}) = P_{|11\rangle}(0)e^{\frac{-2t}{T_1}}$$

To illustrate the noise impact, we use a circuit to prepare a 3-qubit state $|\psi\rangle = \frac{1}{2}|000\rangle + \frac{1}{2}|001\rangle + \frac{1}{2}|110\rangle + \frac{1}{2}|111\rangle$ and then vary the noise by appending the circuit with different numbers of CNOT pairs. With a quantum noise simulator, we first only enable depolarizing noise, and the probabilities of the measured final state are shown in Fig 1a. In the figure, the x-axis shows the noise level, quantified with $1 - reliability$, where the reliability is computed as the estimated success probability (ESP) (Eq. 3). The figure shows that with more and more depolarizing noise, the 3-qubit state $|\psi\rangle$ decays to the maximally mixed state, where each 3-qubit state in $|000\rangle, |001\rangle, |010\rangle, ..., |111\rangle$ has the same probability of 1/8.

We also repeat the experiment with decoherence noise, and the results are shown in Fig. 1b. The x-axis is the time. We can see that with higher and higher latency resulting from the increased circuit depths, there would be stronger decoherence noise and the 3-qubit state $|\psi\rangle$ gradually decays into the $|000\rangle$ state, i.e., only $|000\rangle$ has a high probability while others have low probabilities. From Fig. 1, we can see the exponential decay rate that depends on the number of 1s in a state.

When both depolarizing and decoherence noise are enabled, the state change patterns with respect to either the noise level or latency are less obvious, as shown in Fig. 1c. Although $|000\rangle$ has a higher probability as a result of decoherence noise, other states have non-trivial probabilities due to depolarizing noise.

B. Quantum Error Mitigation

There is a plethora of QEM protocols. ZNE is one of the most popular QEM protocols. It works by scaling the circuit's noise and executing the circuits with different noise scales. Then it infers/extrapolates the zero-noise case from the measurement results.

Clifford Data Regression (CDR) introduced by Czarnik et al. [7] is a learning-based QEM. CDR uses a training set of noisy and ideal results of quantum circuits comprised entirely of Clifford gates. Simulation of such circuits is efficient on classical computers. We can hence efficiently obtain the ideal noise-free results of such quantum circuits. The set of Clifford circuits is then run on a quantum device to obtain the noisy results. Both ideal and noisy results form a training set. A linear ansatz, $a_1 X^{noisy}_{|\psi\rangle} + a_2$, is deduced to correct the noisy circuit, where a_1 and a_2 are the parameters trained by machine learning methods from the training set. A recent proposal by Lowe et al. [14], variable-noise Clifford data regression (vnCDR), combines the ideas of both ZNE and CDR and achieves better mitigation results.

Introduced by Temme et al. [22], Probabilistic Error Cancellation uses a linear combination of noisy quantum gates to represent ideal gates in the circuits. Quasi-probability representations are used to represent linear combinations. The

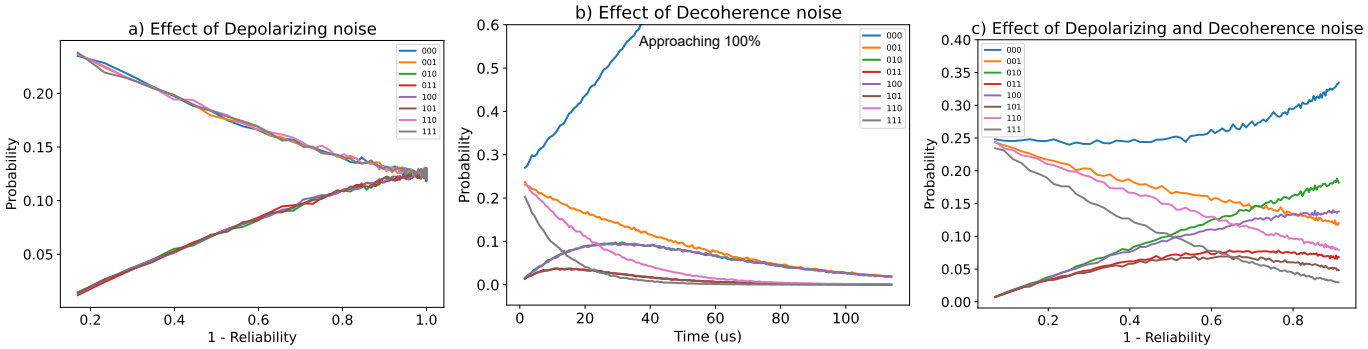


Fig. 1: Effect of depolarizing (a), thermal relaxation (decoherence) (b), and a combination of depolarizing and thermal relaxation errors (c) on a 3-qubit circuit. The noise free state is $|\psi\rangle = \frac{1}{2}|000\rangle + \frac{1}{2}|001\rangle + \frac{1}{2}|110\rangle + \frac{1}{2}|111\rangle$. The noise in the circuit is increased by appending pairs of CNOT gates.

representations are determined by sampling using a Monte Carlo estimation.

A recent work, QRAFT [19], uses machine learning for quantum error mitigation. The process involves executing the quantum circuit to be mitigated and preparing a circuit appended with its inverse. In an ideal noise-free scenario, the circuit and its inverse should yield the initial state, but in a noisy scenario, this is not the case. Multiple circuits (circuit and circuit + inverse runs) and their features constitute the training set for a machine learning-based prediction model. This model is then used to predict error-mitigated outputs from the noisy outputs of new quantum circuits.

While the previously mentioned QEM methods are not limited by the types of error they try to mitigate, there are QEM methods that aim to mitigate a particular type of error. State Preparation and Measurement (SPAM) error mitigation methods [16] are such an example. In addition, Dynamical decoupling (DD) [25] [28] is proposed as QEM to mitigate idling errors [8] as well as cross-talk errors [24]. Unlike previously mentioned QEM methods, DD does not involve running an ensemble of circuits or any post-processing step.

III. Motivation

The key idea of ZNE is to scale up the noise levels of a quantum circuit and then extrapolate to the zero-noise case based on the measurements of these different noise levels. One way to scale the noise is to time-scale the same unitary evolution by controlling the pulses of each gate [22]. Another way is to replace a unitary gate or a subcircuit or the complete circuit U with $U(U^\dagger U)^n$ [10], where n is the circuit scale/folding factor. The underlying assumption is that the noise level scales in proportion to the circuit scale/folding factor, which is accurate if the circuit noise is dominated by depolarising noise [10]. Based on the assumption that errors in an operation are stochastic [12] [9], Li and Benjamin [12] introduced Richardson extrapolation, while Endo et al. [9] introduced exponential extrapolation. Giurgica-Tiron et al. [10] explored various extrapolation models, such as the polynomial and the poly-exponential models, and also introduced adaptive extrapolation models.

The above-mentioned ZNE approaches have the following limitations. First, the relationship between the expectation value and the circuit scale/folding factor is nonlinear [10]. As a result, different curve fitting methods have been used to extrapolate the zero-noise case, and it is difficult to decide which one should be used in practice. Second, extrapolation based on the circuit scale factor ignores the circuit complexity. For two circuits with different complexity, e.g., a single-qubit circuit with a limited depth vs. a multi-qubit circuit with a high depth, the same scaling factor may have drastically different noise impacts. Third, scaling/folding the circuit does not yield any meaningful changes in the output when the noise is high, thereby making extrapolation useless. For example, when the depolarizing noise level is very high, the output is close to the maximally mixed state. Further scaling up the noise levels would not have much impact on the output state anyway.

Next, we describe our proposed improvements to overcome these limitations.

IV. A Comprehensive Approach for Zero Noise Extrapolation

A. Overview

Our proposed ZNE integrates different extrapolation models to mitigate different types of noise and uses quantum circuit-cut techniques [20], [21] to lower the noise level when the circuit suffers from high noise. The overall flow of our approach is shown in Fig. 2. We start by checking the circuit noise level or reliability. If the noise level/reliability is higher/lower than a threshold θ, meaning that further scaling up noise levels would not lead to meaningful results, we apply circuit cut techniques to break the original circuit into subcircuits. Then, for each subcircuit, we check whether depolarizing noise is dominant. If yes, we can skip the process of decoherence noise mitigation. Otherwise, we first apply decoherence noise mitigation and then depolarizing noise mitigation. In the last step, we combine the mitigated state results of each subcircuit to produce the result for the original circuit, if the circuit was cut previously.

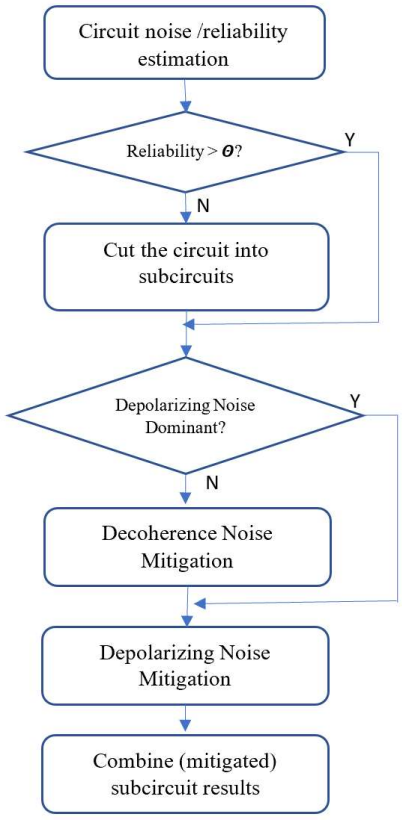
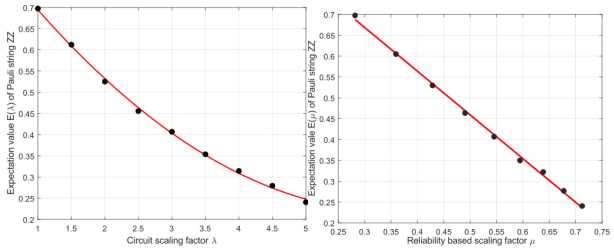


Fig. 2: Flowchart of our proposed ZNE scheme.

Due to the different natures of depolarizing and decoherence noise, we propose different extrapolation schemes, linear extrapolation for depolarizing noise mitigation, and exponential extrapolation for decoherence noise mitigation. These two mitigation schemes are discussed in Section IV.B and IV.C, respectively. Section IV.D discusses how we scale down the noise level using the quantum circuit cut techniques.

B. Reliability-based ZNE for Depolarizing Noise Mitigation

We propose a novel reliability-based ZNE (RZNE) for depolarizing noise mitigation. Existing ZNE schemes such as digital ZNE [10] perform extrapolation based on the circuit scale factor. In contrast, we propose to perform extrapolation based on the reliability, r, of the scaled quantum circuits. The motivation is to take the circuit complexity into consideration. Since it is preferred to extrapolate to zero noise, we choose to perform extrapolation using $\mu = 1 - r$ rather than r. Zero noise means μ being 0 or r being 100%, i.e., perfect reliability. Fig.3a shows an illustrating example. The y-axis is the ZZ expectation values, $E(\lambda)$, of a 2-qubit quantum circuit with respect to different circuit folding factors λ. The circuits, including the original U and the folded ones $U(U^{\dagger}U)^{(\lambda-1)}$, run on a noise simulator using the noisy model of a 5-qubit device ibmq_quito [3]. The circuit scaling is implemented using the Mitiq framework [11]. The expectation value of the noise-free execution of the circuit is 1. $E(1)$ is the expectation



(a) $E(\lambda)$ with different circuit folding factor λ

(b) $E(\mu)$ with different reliability-based scaling factor $\mu = 1 - r$

Fig. 3: Expectation values of a 2-qubit circuit with different scaling factors and the ZNE results

value based on the measurement of the original circuit and $E(0)$ is the extrapolated expectation value corresponding to the zero-noise case. In comparison, RZNE changes the X-axis from the circuit folding factor λ to $\mu = 1 - r$, where $\mu = 1 - r$ can be viewed as a metric of the (reliability-based) noise scaling factor, as shown in Fig.3b. Here, we use the simple Estimated Success Probability (ESP) [18] as the metric of the reliability of the quantum circuit and its noise-scaled variants. It is defined as follows.

$$ESP = \prod_{i=1}^{N_{gates}} (1 - g_i^e) * \prod_{j=0}^{N_{meas}} (1 - m_j^e) \tag{3}$$

In the equation, g_i^e is the gate error rate, and m_i^e is the measurement error rate that are obtained from device calibration. Other metrics, such as reliability estimation using machine learning models [13], [26] or those considering more comprehensive noise effects [5], can also be used for more accurate estimation.

A unique advantage of using reliability-based extrapolation is that it readily enables extrapolation from the extremely noisy case, which means scaling up the circuit by an infinite factor, i.e., $\lambda = \infty$. With an infinite scale factor or noise, the reliability becomes 0 and the noise scale $\mu = 1 - r = 1$. In comparison, such an infinite number or even a very large scale factor would be difficult to handle in curving-fitting based on λ, especially linear curves. Furthermore, the expectation value $E(\lambda = \infty)$ or $E(\mu = 1)$ does not require execution of the circuit. The reason is that in this infinite noise case, the output state is the maximally mixed state when depolarizing noise is dominant, and the measured state vector would follow the uniform distribution. In our example, the ZZ expectation based on the maximally mixed state or the uniformly distributed state is $\langle\varphi| Z \otimes Z |\varphi\rangle$ = 0, if φ is the maximally mixed state.

Another important advantage of RZNE is that a simple linear extrapolation would suffice, which we will prove next.

As shown in Eq.1, depolarizing error leads to a linear combination of the state itself and the maximally mixed state. The factor p is the depolarizing probability. When the depolarizing noise is dominant, we can replace the term $1 - p$

with the circuit reliability, r, and p becomes $1-r$. Hence we will have,

$$E(\rho) = r\rho + (1-r)\frac{I}{2^n} = r\rho + \mu\frac{I}{2^n} \tag{4}$$

Thus depolarizing noise is linear with respect to r. At infinite noise, $r_{\infty noise} = ESP_{\infty noise} = 0$, the state is the maximally mixed state $\rho_{\infty noise} = \frac{I}{2^n}$. For the noiseless case, $r_{zeronoise} = ESP_{zeronoise} = 1$, the state $\rho_{zeronoise} = \rho$, which can be derived as.

$$\rho = \frac{E(\rho)}{r} - \frac{\mu}{r}\frac{I}{2^n} = \frac{E(\rho)}{r} - \frac{1-r}{r}\rho_\infty \tag{5}$$

where r is the reliability of the original circuit, and $E(\rho)$ is the noisy result/state from the original circuit.

The implication of Eq. 5 is that one can extrapolate the zero noise state simply by connecting the infinite noise state and the noisy state from the original circuit. RZNE can be performed either on the output state or on the expectation value. When extrapolating for the entire output state, the cost is 2^N extrapolations, where N is the number of qubits. However, most quantum applications have few dominant states. Thus, we propose only correcting a quantum system's Top K states. In other words, only the K highest probability states/coefficients are extrapolated.

It is worth noting that the infinite noise case, i.e., the maximally mixed state, is independent of quantum devices or different mappings on a device. As a result, we can extend our RZNE approach to combine measurement results from different devices/mappings using linear curve fitting with the condition that the linear curve must pass the point corresponding to the infinite noise case, where multiple points are circuit executions on different devices or different mappings.

Algorithm 1: Reliability-based ZNE (RZNE)

Data: The circuit C; Y_C containing the execution results corresponding to the circuit
Result: An expectation value after noise mitigation
begin
  $r \leftarrow ComputeReliability(C)$;
  $\mu \leftarrow 1 - r$;
  $Y_C \leftarrow RunCircuit(C)$;
  /* μ_{m+1} is the factor for the infinite noise case which has reliability of 0 */
  $\mu_{m+1} \leftarrow 1 - 0$;
  /* Compute expectation value based on maximally mixed state */
  $y_{m+1} \leftarrow ComputeExpectationValue(\mu_{m+1})$;
  /* $E(\mu; a, b) = a + b\mu$ is a linear model with fixed point(μ_{m+1}, y_{m+1}) */
  $a, b \leftarrow BestFit(E(\mu; a, b), (\mu, \mathbf{Y}))$;
  return $E(0; a, b)$;
end

In summary, we list our RZNE algorithm in algorithm 1. In the algorithm, the set C is the original circuit and the same circuit with different qubit mappings or different devices, if they are used. Their reliability is computed using ESP due to its simplicity. The set Y are the execution results of the circuits, i.e., either noisy states or expectation values. During curve fitting, the fitted linear curve is required to pass the point corresponding to the infinite noise case, (μ_{m+1}, y_{m+1}). If $m = 1$, the curve fitting result is simply the line connecting the infinite noise case and the noisy result of the original circuit. The algorithm's output is the extrapolated result corresponding to zero noise or reliability being 1 (i.e., $\mu = 0$).

C. State- and Latency-Aware ZNE for Decoherence Noise Mitigation

From eq. 2, at $t = 0$, the term $e^{\frac{-t}{T_1}}$ approaches 1, thereby $P_{|1\rangle}(t) = P_{|1\rangle}(0)$, indicating no effect of decoherence noise. In other words, we can obtain the error-free values if we extrapolate to time $t = 0$. As the decay rate depends on both the time and the number of excited 1s in a multi-qubit state, we propose the following state- and latency-aware ZNE, SLZNE, to mitigate decoherence noise.

$$P_s(0) = P_s(t)/e^{\frac{-t}{T_1}} \tag{6}$$

In the equation, t is the circuit latency, c is the number of 'ones' in the quantum state s, and $P_s(t)$ is the probability of state s, i.e., the measured probability of the state s. However, the state with all zeros grows with time with probability $P_s(0) \leftarrow \frac{1-P_s(t)}{e^{-t/T_1}}$.

SLZNE repeats extrapolation on every output state, which comes at the cost of 2^n calculations. To reduce this complexity, we resort to the same TOP K approach discussed in Section IV.B. Our SLZNE algorithm for mitigating the decoherence noise is shown in Algorithm 2.

D. Circuit-Cut based ZNE

A quantum circuit with very low reliability has its output state close to the infinite noise case. Noise-boosting steps for such circuits result in minor changes in the circuit's output, thereby limiting the effectiveness of ZNE. To handle such cases, we propose to leverage circuit-cutting techniques to divide a circuit into less noisy sub-circuits. Each sub-circuit has higher reliability than the original circuit since the number of gates and/or qubits in each sub-circuit is reduced. We can apply ZNE upon each sub-circuit and then combine their results to obtain the output of the original circuit, i.e., mitgate-then-combine or CutQC-MC.

In our scheme, we use the CutQC scheme proposed by Tang et al. [21] to cut the circuit into sub-circuits and then recombine the sub-circuit results. CutQC has a post-processing overhead that scales exponentially with the number of cuts. We adopt the design from [21], which aims to find the cut locations to minimize this post-processing overhead.

Our Cut-ZNE algorithm is shown in algorithm 3. The input is a quantum circuit C. The set S contains the circuit-cut results, i.e., a set of sub-circuits and the boolean variable MC is set to True. If the number of cuts is lower than a threshold, max_num_cuts, we perform ZNE on each sub-circuit, first

Algorithm 2: State- and Latency-Aware ZNE (SLZNE)

Data: Circuit C; T_1 is the thermal relaxation time of the device
Result: An expectation value after noise mitigation
begin
  $t = ComputeLatency(C)$;
  $Y_C \leftarrow RunCircuit(C)$;
  /* Extrapolate for each state using $P_s(0) = P_s(t)/e^{-ct/T_1}$ */
  for $s \in Y_C$ **do**
    /* Get the number of excited qubits in state s. */
    $c \leftarrow CalculateNumberOfOnes(s)$;
    if $c == 0$ **then**
      $P_s(0) \leftarrow \frac{1-P_s(t)}{e^{-t/T_1}}$;
    end
    else
      $P_s(0) \leftarrow \frac{P_s(t)}{e^{-ct/T_1}}$;
    end
    $Append(Y_{corrected}, P_s)$
  end
  $E \leftarrow CalculateExpectationValue(Y_{corrected})$;
  return E;
end

SLZNE if the decoherence noise is significant, and then RZNE for mitigating depolarizing noise. The mitigated results of each sub-circuit are combined through the $CombineResult$ function.

Since mitigating every sub-circuit may prove costly, we also propose combine-then-mitigate (CutQC-CM), i.e., combining the noisy sub-circuit results and then mitigating the final state/expectation value. As our RZNE requires the circuit reliability for extrapolation, we use the geometric mean of all the sub-circuit reliabilities as the reliability of the combined one. The algorithm for this approach is presented in Alg. 3 with $MC = False$.

V. Experimental Methodology

A. Benchmarks

To evaluate our proposed folding-free ZNE schemes, we run the experiments on the benchmarks shown in Table I. SupermarQ [23] is used to generate our benchmarks.

In our experiments, for the Top K versions of RZNE and SLZNE, K is set to a constant 10. The threshold for applying RZNE is set to $ESP > 0.10$ ($\theta = 0.10$ in Fig. 2), and the heuristic threshold for selectively applying SLZNE is set to $0.7 \times T_1$.

B. Experimental Setup

Our experiments are performed on the IBM Qiskit QASM simulator and real quantum devices. The simulator uses a noise model from IBM's 16 qubit quantum device IBM Guadalupe. The number of shots is set to 32768.

ZNE can be applied in the following two ways when the desired output is an expectation value. The first is to apply ZNE upon each individual coefficients of the output state, normalize the coefficients, and then compute the expectation value based on the noise-mitigated state. The second is to apply ZNE directly upon the expectation value, which means to extrapolate from the expectation values of the scaled circuits. In our experiments, we report the results using the first way.

We demonstrate Cut-ZNE with only the Hamiltonian Simulation, VQE and GHZ benchmark. The number of cuts required by CutQC to separate the circuits for QAOA was

Algorithm 3: Cut-QC based RZNE (CutQC-MC & CutQC-CM)

Data: the circuit C
Result: An error mitigated result
begin
  /* $\{S\}$ is a set of subcircuits of the original circuit, $\{R\}$ is their ESPs and cuts is the number of cuts required to separate circuit into sub circuits. When variable MC is set, CutQC-MC is used; when not set, CutQC-CM is used. */
  $\{S\}, cuts \leftarrow CutCircuit(C)$;
  if $cuts < max\_num\_cuts$ **then**
    $Y \leftarrow \emptyset$;
    for $s_j \in S$ **do**
      $Y_j \leftarrow \emptyset$;
      $Y_j \leftarrow RunCircuits(s_j)$;
      if MC *is* $True$ **then**
        $y_j \leftarrow RZNE(\{s_j\}, \{y_j\})$;
      end
      Append(Y, y_j);
    end
    $y \leftarrow CombineResults(S, Y)$;
    if MC *is* $False$ **then**
      $r \leftarrow GeometricMean(\{R\})$;
      $y \leftarrow RZNE(y, r)$
    end
    return y;
  end
end

Benchmark	Parameters
Hamiltonian Simulation (HS)	No. of Qubits, No. of time steps
Vanilla QAOA (QA)	No. of Qubits
Fermionic Swap QAOA (QS)	No. of Qubits
Variational Quantum Eigensolver (VQE)	No. of Qubits, No. of Layers
GHZ State Preparation (GHZ)	No. of Qubits

TABLE I: Application benchmarks.

more than our predetermined threshold of 2. The reasoning behind our choice to limit cuts is the classical post-processing, which scales with 4^K with K being the number of cuts. Given the exponential scaling, 2 was chosen as the maximal number of allowable cuts to demonstrate the effectiveness of Cut-ZNE. We used methods provided by authors of CutQC [21] for circuit cutting and sub-circuit result recombination. We show the effectiveness of approaches proposed in Alg. 3 on the aforementioned benchmarks.

We also compare our approaches with the state-of-the-art noise mitigation schemes shown in Table II. For DZNE, CDR, and vnCDR, we use the software package Mitiq [11]. For QRAFT, we used the code provided by the authors [2]. For DZNE, we also explore full state extrapolation, in which case the parameters are the same as in II, but DZNE is applied for every element in the quantum state.

We open-sourced our code at https://github.com/hrushikesh890/FoldFreeZNE

Method	Parameters
Digital ZNE (DZNE)	Scaling Factors = {1, 2, 3, 4, 5}, Extrapolation method = Linear
Clifford Data Regression (CDR)	Training set = 64 circuits, Fit method = Linear
variable noise CDR (vnCDR)	Training set = 64 circuits, Fit method = Linear, Scale Factors = 1, 2, 3, 4
QRAFT	Training set = 1000 random circuits

TABLE II: Error mitigation methods used for comparison.

C. Evaluation Criteria

To evaluate the effectiveness of different QEM schemes, we use the absolute observable error (ABE) [14], which is defined as follows, to quantify the error impact on the observable.

$$ABE(\langle O \rangle) = |\langle O \rangle_{ideal} - \langle O \rangle_{observed}| \tag{7}$$

To compute the improvement from error mitigation methods over the existing errors, we propose the following metric termed as Absolute Error Ratio (ABR),

$$ABR = \left| \frac{\langle O \rangle_{ideal} - \langle O \rangle_{mitigated}}{\langle O \rangle_{ideal} - \langle O \rangle_{noisy}} \right| \tag{8}$$

The lower the value of ABR, the better the performance of the error mitigation method. If for a mitigation scheme, its ABR is greater than or equal to '1', it means that the error mitigation method in question does not successfully mitigate the error compared to the unmitigated case. The reason is that the Absolute Error Ratio for an unmitigated circuit is always 1.

For the GHZ benchmark, we report the Hellinger fidelity [1] instead of the ABE or ABR as GHZ does not have an observable.

VI. Experimental Results

A. Mitigating Depolarizing Noise

We first examine the effectiveness of RZNE in mitigating errors arising from depolarizing noise. Depolarizing noise is simulated using Qiskit [4] noise model. The noise model is constructed from the backend information of the `ibmq_guadalupe` machine. Within the noise model, we disable the thermal relaxation error and readout error while keeping only the depolarizing error. We evaluate our method on Hamiltonian Simulation (HS), VQE, QAOA (QA), and Swap-QAOA (QS) benchmarks. Each benchmark has 4 circuits with different ESP values. For each run, we calculate the expectation value from the ideal noise-free simulation, noisy simulation, RZNE-mitigated noisy simulation, and Top 10 RZNE-mitigated simulation. As discussed in Section V.C., the absolute error ratio (ABR) is used to quantify the error impact. Our experimental results are shown in Fig. 5 and 4. It is evident from the results that both variants of RZNE effectively reduce the error caused by noise for all benchmarks. On average, RZNE demonstrates an 88% reduction in error when compared to unmitigated results, while RZNE Top 10 demonstrates a 72% reduction. As shown in Fig. 5, RZNE outperforms RZNE Top 10, but requires more post-processing. The advantage of RZNE is that it corrects all the states, reducing the impact of states with higher probabilities due to noise. On the other hand, RZNE Top K only operates on a portion of the states, leaving some states uncorrected, resulting in worse performance than RZNE. In Fig. 4, we observe that for the GHZ benchmarks with a higher number of qubits, RZNE Top K begins to perform significantly better than RZNE when the number of qubits increases. This is because the GHZ benchmark has two dominant states, and extrapolating only a few states prevents unintentional boosting of incorrect states. We can thus conclude that if there is prior information on the number of dominant states, the RZNE Top K method is the preferred one.

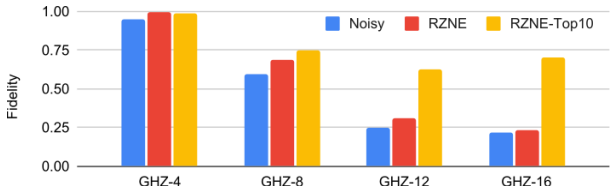


Fig. 4: The fidelity of the GHZ state with depolarizing noise mitigated by RZNE and RZNE-Top 10 (Higher the better).

B. Mitigating Decoherence Noise

In this experiment, we examine the effectiveness of SLZNE on mitigating decoherence errors. We use Qiskit [4] noise model with only thermal relaxation noise enabled. The median T_1 is $91.99\mu s$. Fig. 7 and Fig. 6 report our results of SLZNE and its variant SLZNE Top-10. We also include the results using RZNE as a reference to compare the effectiveness of SLZNE.

The results in Figure 7 show that while all methods were successful in mitigating decoherence noise, SLZNE Top 10 is the most effective one on average. SLZNE is less optimal because it extrapolates all states, which causes some non-dominant states to be extrapolated incorrectly. When states with high number of 'ones' (excited states) decay into 'in-

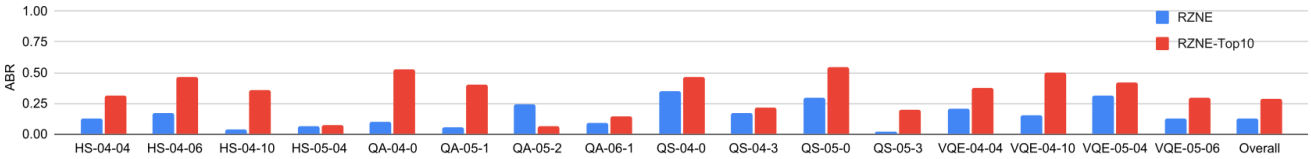


Fig. 5: Absolute error ratios for RZNE and RZNE-Top 10 on various benchmarks under depolarizing noise. (Lower the better).

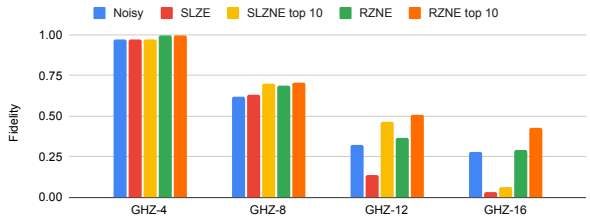


Fig. 6: The fidelity of the GHZ state with decoherence noise mitigated by SLZNE and RZNE.

termediate states', which are generally close in Hamming distance, due to decoherence. Some of these 'intermediate states' may not be present in the ideal noiseless output. However, when SLZNE corrects all the states, it causes these intermediate states to incorrectly have higher probabilities. This problem is avoided by only correcting a few of the states with the highest probabilities. Then, when we re-normalize the coefficients, the intermediate states have their probabilities reduced. Another important observation is that VQE, which has higher latency, had the best performance using SLZNE, while QAOA and GHZ benchmarks with low-latency circuits resulted in poor performance of SLZNE. This can be attributed to the fact that with higher latency, the 'intermediate states' are more likely to decay into the all zeros state, thereby reducing the chance that intermediate states get incorrectly mitigated.

C. Mitigating both Depolarizing and Decoherence Noise

In this experiment, both the depolarizing error and thermal relaxation error were enabled within the noise model. The results are presented in Figures 9 and 8, which include the outcomes of both RZNE and SLZNE+RZNE, as well as their respective Top K versions. The results we obtained are consistent with previous experiments in mitigating depolarizing and decoherence noises. Specifically, we have observed that SLZNE-based methods work best for circuits with high latency such as VQE benchmark circuits, while RZNE is effective for short-depth circuits like QAOA. Based on these observations, we apply SLZNE only when the circuit latency exceeds $0.7 \times T_1$ in the remaining experiments.

D. Comparison with the State-of-the-art QEM Approaches

In this experiment, we test our methods against the state-of-the-art QEM methods including Digital Zero Noise Extrapolation (DZNE), Clifford Data Regression (CDR), variable noise Clifford Data Regression (vnCDR), and QRAFT. We enable all but readout error noise sources of the Qiskit noise simulator and the noise model is constructed from the backend information of the `ibmq_guadalupe` machine. The experimental results are reported in Fig. 10 for the Hamiltonian Simulation, VQE, QAOA and Fermionic Swap QAOA and Fig. 11 for the GHZ benchmark.

We observe that RZNE consistently results in lower median error than all other state-of-the-art (SOA) methods. Our heuristic of using $0.7 \times T_1$ for Selective SLZNE + RZNE can only be applied to Hamiltonian Simulation and VQE benchmarks, as only these benchmarks have circuit latencies above the heuristic threshold. In particular, RZNE performs exceptionally well for the Hamiltonian simulation benchmark compared to other error mitigation methods. The performance gain of RZNE over Digital ZNE (DZNE) is due to two reasons: 1) DZNE extrapolates over the expectation value of the observable, while RZNE extrapolates over the state; and 2) by utilizing reliability as a metric, RZNE determines how aggressively to apply the correction. It is difficult to predict the effect of noise on the expectation value of an observable. This is primarily why we need folding. We also need to choose, or find, the extrapolation model that best fits the data. However, the model that best fits the data runs a risk of overfitting leading to incorrect correction. As we cannot always guarantee prior knowledge of the correct observable, the efficacy of chosen extrapolated method cannot be guaranteed. RZNE avoids all these issues by correcting states, as we can reasonably determine the evolution of a state in the presence of specific noise factors. Thus, we can fix the choice of extrapolation method for RZNE. Another downside of utilizing DZNE is that folding a circuit that has been significantly affected by noise does not yield any meaningful change in the value of the observable, resulting in a near-flat line for extrapolation that does not provide any scope for significant error mitigation. RZNE rectifies this by not having fixed scale factors, and for low reliability values, RZNE extrapolation is aggressive, resulting in better error mitigation. Our method also performs well against the learning based approaches like QRAFT and CDR.

E. Mitigating Noise on Real Superconducting Device

In this experiment, we perform tests on various IBM machines. On the real device, we reduced the number of shots to 8192. We report the observed absolute error ratios in Table III. We compare our method with DZNE (Scale factors = 1, 3, 5), CDR and vnCDR (training set = 32 circuits). We observe that the results on real machine agree with the simulator results presented in Section VI-D. Important to note is that only VQE 7 qubit experiment passed our heuristic threshold for $0.7 \times T_1$.

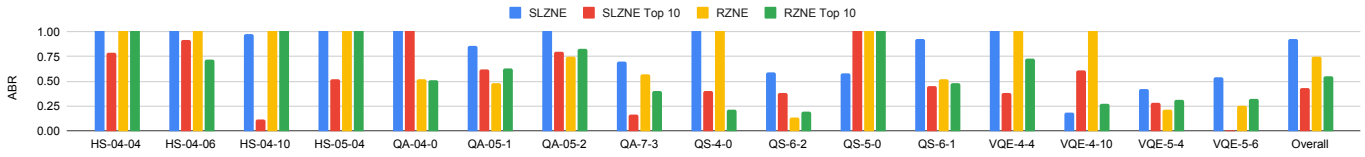


Fig. 7: Absolute error ratios for SLZNE and RZNE on various benchmarks under decoherence noise.

Benchmark	Machine	Qubits	ESP	RZNE	RZNE Top 10	SLZNE	SLNE Top 10	DZNE	DZNE full state	CDR	vnCDR
Hamiltonian Simulation	ibm_kolkata	4	0.860	0.217	0.571	0.217	0.571	0.656	0.439	0.758	0.682
		4	0.740	0.424	0.192	0.424	0.192	0.223	0.051	2.099	1.055
		7	0.654	0.924	0.933	0.924	0.933	1.075	0.764	1.106	1.105
		7	0.332	0.849	0.901	0.849	0.901	0.975	0.927	0.433	1.124
VQE	ibm_guadalupe	4	0.797	0.507	0.766	0.507	0.766	0.637	0.672	0.506	0.630
		4	0.687	0.606	0.712	0.606	0.712	0.742	0.862	0.417	0.772
		7	0.081	1.072	1.189	1.072	1.189	0.999	1.001	0.947	0.995
		7	0.009	1.170	1.229	0.314	0.387	0.998	0.990	1.251	1.106
Swap QAOA	ibm_mumbai	4	0.878	0.274	0.721	0.274	0.721	0.771	0.370	3.533	3.319
		4	0.878	0.727	0.842	0.727	0.842	0.855	0.643	1.882	1.836
		7	0.625	1.006	1.022	1.006	1.022	1.002	1.013	1.011	1.027
QAOA	ibm_hanoi	4	0.578	0.387	0.689	0.387	0.689	0.556	0.659	2.173	2.239
		4	0.747	0.831	0.889	0.831	0.889	0.760	0.766	1.608	1.579
		7	0.454	0.784	0.649	0.784	0.649	1.012	1.074	0.735	1.108

TABLE III: Absolute error ratios of state of the art error mitigation methods and our method on real quantum devices.

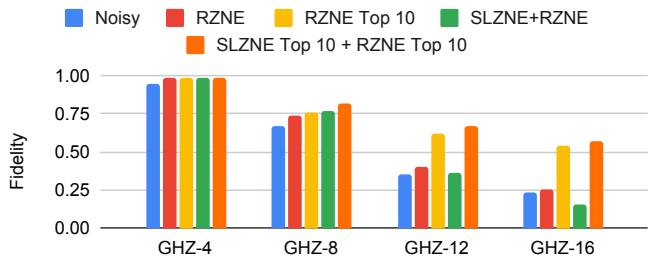


Fig. 8: The fidelity of the GHZ state with deploarizing and decoherence noise mitigated by SLZNE and RZNE.

Benchmark	Machine	Qubits	Reliability	Fidelity		
			ESP	Noisy	RZNE	RZNE Top 10
GHZ	ibm_cairo	4	0.805	0.583	0.648	0.635
		16	0.683	0.457	0.458	0.585

TABLE IV: The fidelity of the GHZ state with our error mitigation schemes on a real quantum device.

Finally, having accurate estimate of gate error is important for RZNE. At times, the IBM backend may have CNOT errors equal to 1 which gives ESP to be 0. This leads to RZNE failing to extrapolate. From Table III, we observe that our methods consistently mitigate the noises on real devices. For GHZ benchmark, the state fidelities are reported in Table IV but we do not include the comparison with DZNE, CDR and vnCDR as they operate on observables instead of quantum states.

F. Results of Cut-ZNE

Circuits with low ESP can be cut into smaller sub-circuits. The sub-circuits can be executed separately and the results can be recombined using classical processing to form the output state of the original circuit. This way, their noise levels/ESPs can be reduced/increased as demonstrated in Tables V and VI. In these experiments, circuits were cut into 2 sub-circuits using up to 2 cuts. We apply Cut-ZNE on the Hamiltonian Simulation, VQE and GHZ benchmarks running on the Qiskit noise simulator with all noise sources but readout error enabled and the results are shown in Tables. V and VI along with the real device, IBM Kolkata, results for Hamiltonian Simulation, VQE benchmarks and IBM Hanoi, for GHZ benchmarks. We also include the RZNE as a reference point for performance of Cut-ZNE methods. We also show the effect of simply cutting the circuit and recombining the execution results of the sub-circuits (labeled 'CutQC Unmitigated') leads to error mitigation by itself. Subsequently applying RZNE in one of two ways reduces error even further.

In the tables, CutQC-CM (combine-then-mitigate) refers to the method in Alg. 3 with $MC = False$ and CutQC-MC (mitigate-then-combine) is with $MC = True$. As seen in tables all methods successfully mitigate errors as their Absolute Error Ratio is less than 1. For VQE and GHZ the best approach observed was CutQC-MC while for Hamiltonian simulation it was the CutQC-CM. RZNE proved to be better than applying CutQC for the Hamiltonian Simulation at least. Most importantly, we can see from both tables that the mitigation of errors is not just due to CutQC, and by applying our ZNE approaches, the errors are further mitigated, i.e., CutQC-CM or CutQC-MC vs. CutQC Unmitigated.

VII. Conclusions

In this paper, we propose a comprehensive scheme for mitigating both decoherence and depolarizing noise. The key novelty includes (a) we propose reliability-based ZNE (RZNE), which performs extrapolation upon the circuit reliability such that we can easily utilize the extreme noise case and avoid folding the circuits; (b) we also show that with extrapolation upon circuit reliability, simple linear extrapolation suffices for mitigating depolarizing noise; (c) we propose SLZNE, which is an exponential extrapolation model based on circuit latency

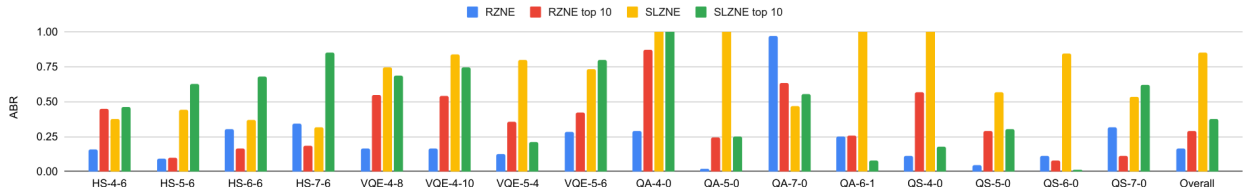


Fig. 9: Absolute error ratios for RZNE and SLZNE on various benchmarks under decoherence and depolarizing noise.

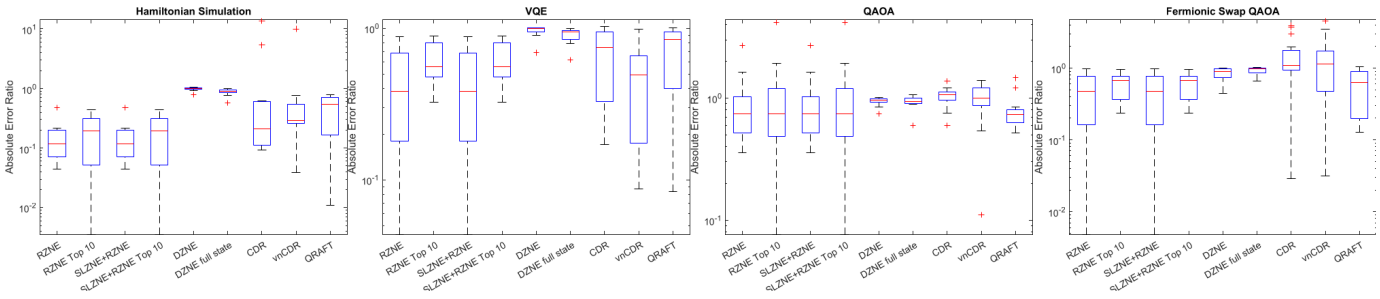


Fig. 10: Comparison of absolute error ratios of RZNE and SLZNE with various state of art error mitigation methods.

Benchmark	Qubits	Machine	Reliability			Absolute Error Ratio			
			ESP	$ESP_{subcircuit1}$	$ESP_{subcircuit2}$	RZNE	CutQC (Unmitigated)	CutQC-CM	CutQC-MC
Hamiltonian Simulation	5	IBM qasm simulator	0.376	0.632	0.579	0.371	0.886	0.029	0.971
	6		0.287	0.632	0.359	0.086	0.806	0.136	0.592
	7		0.212	0.585	0.359	0.1	0.926	0.25	0.535
	7	ibmq_kolkata	0.649	0.980	0.894	0.923	0.341	0.307	1.024
	8		0.577	0.964	0.894	0.916	0.997	0.978	0.909
VQE	8	IBM qasm simulator	0.117	0.796	0.344	0.666	0.956	0.831	0.342
	9		0.087	0.494	0.344	0.589	0.663	0.596	0.399
	10		0.07	0.136	0.477	0.559	0.983	0.763	0.569
	7	ibmq_kolkata	0.644	0.988	0.918	0.913	0.687	0.754	0.635
	8		0.550	0.919	0.979	0.817	0.669	1.005	0.488

TABLE V: Absolute error ratios for Cut-ZNE error mitigation schemes.

Benchmark	Qubits	Machine	Reliability			Fidelity				
			ESP	$ESP_{subcircuit1}$	$ESP_{subcircuit2}$	Noisy	RZNE	CutQC Unmitigated	CutQC-CM	CutQC-MC
GHZ	4	IBM qasm simulator	0.929	0.958	0.944	0.936	0.980	0.918	0.943	0.996
	8		0.604	0.815	0.930	0.580	0.662	0.732	0.764	0.931
	12		0.380	0.604	0.815	0.270	0.321	0.521	0.544	0.676
	16		0.286	0.737	0.420	0.213	0.227	0.363	0.377	0.432
	8	ibm_hanoi	0.944	0.976	0.983	0.665	0.695	0.813	0.820	0.818
	12		0.906	0.944	0.976	0.537	0.550	0.630	0.633	0.662
	16		0.826	0.967	0.919	0.405	0.406	0.530	0.533	0.552

TABLE VI: The fidelity of GHZ benchmark using Cut-ZNE error mitigation schemes.

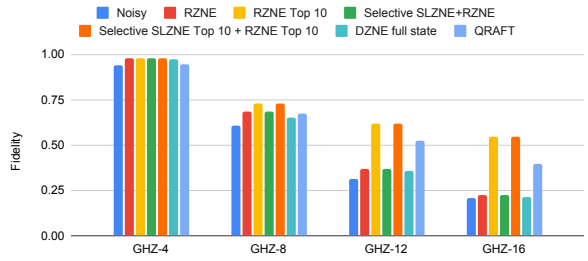


Fig. 11: The fidelity of the GHZ benchmarks with different error mitigation methods.

and the number of 1s in the state to mitigate decoherence noise; (d) we propose a heuristics to integrate RZNE and SLZNE: if the circuit latency is lower than $0.7 \times T_1$, we use RZNE; otherwise, we first apply SLZNE then RZNE; and (e) we propose to leverage quantum circuit cut techniques to scale down the noise level so as to make ZNE approaches more effective for circuits suffering from very high noises. Our experiments on both noise simulators and real quantum devices demonstrate the effectiveness of our methods against the state-of-the-art quantum error mitigation schemes on commonly used benchmarks.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their insightful comments. This work is partly funded by NSF grants 1818914 (with a subcontract to NC State University from Duke University), and 2120757 (with a subcontract to NC State University from the University of Maryland). It is also supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers.

REFERENCES

[1] "Hellinger fidelity." [Online]. Available: https://qiskit.org/documentation/stubs/qiskit.quantum_info.hellinger_fidelity.html

[2] *QRAFT ASPLOS 21 Code and Dataset*. Zenodo, Feb. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.4657000

[3] "Ibm quantum," https://quantum-computing.ibm.com/, 2021.

[4] M. S. ANIS, Abby-Mitchell, H. Abraham, AduOffei, R. Agarwal, G. Agliardi, M. Aharoni, I. Y. Akhalwaya, G. Aleksandrowicz, T. Alexander, M. Amy, S. Anagolum, Anthony-Gandon, E. Arbel, A. Asfaw, A. Athalye, A. Avkhadiev, C. Azaustre, P. BHOLE, A. Banerjee, S. Banerjee, W. Bang, A. Bansal, P. Barkoutsos, A. Barnawal, G. Barron, G. S. Barron, L. Bello, Y. Ben-Haim, M. C. Bennett, D. Bevenius, D. Bhatnagar, A. Bhobe, P. Bianchini, L. S. Bishop, C. Blank, S. Bolos, S. Bopardikar, S. Bosch, S. Brandhofer, Brandon, S. Bravyi, N. Bronn, Bryce-Fuller, D. Bucher, A. Burov, F. Cabrera, P. Calpin, L. Capelluto, J. Carballo, G. Carrascal, A. Carriker, I. Carvalho, A. Chen, C.-F. Chen, E. Chen, J. C. Chen, R. Chen, F. Chevallier, K. Chinda, R. Cholarajan, J. M. Chow, S. Churchill, CisterMoke, C. Claus, C. Clauss, C. Clothier, R. Cocking, R. Cocuzzo, J. Connor, F. Correa, Z. Crockett, A. J. Cross, A. W. Cross, S. Cross, J. Cruz-Benito, C. Culver, A. D. Córcoles-Gonzales, N. D, S. Dague, T. E. Dandachi, A. N. Dangwal, J. Daniel, M. Daniels, M. Dartiailh, A. R. Davila, F. Debouni, A. Dekusar, A. Deshmukh, M. Deshpande, D. Ding, J. Doi, E. M. Dow, P. Downing, E. Drechsler, E. Dumitrescu, K. Dumon, I. Duran, K. EL-Safty, E. Eastman, G. Eberle, A. Ebrahimi, P. Eendebak, D. Egger, ElePT, Emilio, A. Espiricueta, M. Everitt, D. Facoetti, Farida, P. M. Fernández, S. Ferracin, D. Ferrari, A. H. Ferrera, R. Fouilland, A. Frisch, A. Fuhrer, B. Fuller, M. GEORGE, J. Gacon, B. G. Gago, C. Gambella, J. M. Gambetta, A. Gammanpila, L. Garcia, T. Garg, S. Garion, J. R. Garrison, J. Garrison, T. Gates, H. Georgiev, L. Gil, A. Gilliam, A. Giridharan, J. Gomez-Mosquera, Gonzalo, S. de la Puente González, J. Gorzinski, I. Gould, D. Greenberg, D. Grinko, W. Guan, D. Guijo, J. A. Gunnels, H. Gupta, N. Gupta, J. M. Günther, M. Haglund, I. Haide, I. Hamamura, O. C. Hamido, F. Harkins, K. Hartman, A. Hasan, V. Havlicek, J. Hellmers, Ł. Herok, S. Hillmich, H. Horii, C. Howington, S. Hu, W. Hu, J. Huang, R. Huisman, H. Imai, T. Imamichi, K. Ishizaki, Ishwor, R. Iten, T. Itoko, A. Ivrii, A. Javadi, A. Javadi-Abhari, W. Javed, Q. Jianhua, M. Jivrajani, K. Johns, S. Johnstun, Jonathan-Shoemaker, JosDenmark, JoshDumo, J. Judge, T. Kachmann, A. Kale, N. Kanazawa, J. Kane, Kang-Bae, A. Kapila, A. Karazeev, P. Kassebaum, T. Kehrer, J. Kelso, S. Kelso, V. Khanderao, S. King, Y. Kobayashi, Kovi11Day, A. Kovyrshin, R. Krishnakumar, V. Krishnan, K. Krsulich, P. Kumkar, G. Kus, R. LaRose, E. Lacal, R. Lambert, H. Landa, J. Lapeyre, J. Latone, S. Lawrence, C. Lee, G. Li, J. Lishman, D. Liu, P. Liu, Lolcroc, A. K. M, L. Madden, Y. Maeng, S. Maheshkar, K. Majmudar, A. Malyshev, M. E. Mandouh, J. Manela, Manjula, J. Marecek, M. Marques, K. Marwaha, D. Maslov, P. Maszota, D. Mathews, A. Matsuo, F. Mazhandu, D. McClure, M. McElaney, C. McGarry, D. McKay, D. McPherson, S. Meesala, D. Meirom, C. Mendell, T. Metcalfe, M. Mevissen, A. Meyer, A. Mezzacapo, R. Midha, D. Miller, Z. Minev, A. Mitchell, N. Moll, A. Montanez, G. Monteiro, M. D. Mooring, R. Morales, N. Moran, D. Morcuende, S. Mostafa, M. Motta, R. Moyard, P. Murali, D. Murata, J. Müggenburg, T. NEMOZ, D. Nadlinger, K. Nakanishi, G. Nannicini, P. Nation, E. Navarro, Y. Naveh, S. W. Neagle, P. Neuweiler, A. Ngoueya, T. Nguyen, J. Nicander, Nick-Singstock, P. Niroula, H. Norlen, NuoWenLei, L. J. O'Riordan, O. Ogunbayo, P. Ollitrault, T. Onodera, R. Otaolea, S. Oud, D. Padilha, H. Paik, S. Pal, Y. Pang, A. Panigrahi, V. R. Pascuzzi, S. Perriello, E. Peterson, A. Phan, K. Pilch, F. Piro, M. Pistoia, C. Piveteau, J. Plewa, P. Pocreau, A. Pozas-Kerstjens, R. Pracht, M. Prokop, V. Prutyanov, S. Puri, D. Puzzuoli, J. Pérez, Quant02, Quintiii, R. I. Rahman, A. Raja, R. Rajeev, I. Rajput, N. Ramagiri, A. Rao, R. Raymond, O. Reardon-Smith, R. M.-C. Redondo, M. Reuter, J. Rice, M. Riedemann, Rietesh, D. Risinger, M. L. Rocca, D. M. Rodríguez, RohithKarur, B. Rosand, M. Rossmannek, M. Ryu, T. SAPV, N. R. C. Sa, A. Saha, A. Ash-Saki, S. Sanand, M. Sandberg, H. Sandesara, R. Sapra, H. Sargsyan, A. Sarkar, N. Sathaye, B. Schmitt, C. Schnabel, Z. Schoenfeld, T. L. Scholten, E. Schoute, M. Schulterbrandt, J. Schwarm, J. Seaward, Sergi, I. F. Sertage, K. Setia, F. Shah, N. Shammah, R. Sharma, Y. Shi, J. Shoemaker, A. Silva, A. Simonetto, D. Singh, D. Singh, P. Singh, P. Singkanipa, Y. Siraichi, Siri, J. Sistos, I. Sitdikov, S. Sivarajah, Slavikmew, M. B. Sletfjerding, J. A. Smolin, M. Soeken, I. O. Sokolov, I. Sokolov, V. P. Soloviev, SooluThomas, Starfish, D. Steenken, M. Stypulkoski, A. Suau, S. Sun, K. J. Sung, M. Suwama, O. Słowik, H. Takahashi, T. Takawale, I. Tavernelli, C. Taylor, P. Taylour, S. Thomas, K. Tian, M. Tillet, M. Tod, M. Tomasik, C. Tornow, E. de la Torre, J. L. S. Toural, K. Trabing, M. Treinish, D. Trenev, TrishaPe, F. Truger, G. Tsilimigkounakis, D. Tulsi, W. Turner, Y. Vaknin, C. R. Valcarce, F. Varchon, A. Vartak, A. C. Vazquez, P. Vijaywargiya, V. Villar, B. Vishnu, D. Vogt-Lee, C. Vuillot, J. Weaver, J. Weidenfeller, R. Wieczorek, J. A. Wildstrom, J. Wilson, E. Winston, WinterSoldier, J. J. Woehr, S. Woerner, R. Woo, C. J. Wood, R. Wood, S. Wood, J. Wootton, M. Wright, L. Xing, J. YU, B. Yang, U. Yang, J. Yao, D. Yeralin, R. Yonekura, D. Yonge-Mallo, R. Yoshida, R. Young, J. Yu, L. Yu, C. Zachow, L. Zdanski, H. Zhang, I. Zidaru, B. Zimmermann, C. Zoufal, aeddins ibm, alexzhang13, b63, bartek bartlomiej, bcamorrison, brandhsn, charmerDark, deeplokhande, dekel.meirom, dime10, dlasecki, ehchen, fanizzamarco, fs1132429, gadial, galeinston, georgezhou20, georgios ts, gruu, hhorii, hykavitha, itoko, jeppevinkel, jessica angel7, jezerjojo14, jliu45, jscott2, klinvill, krutik2966, ma5x, michelle4654, msuwama, nico lgrs, nrhawkins, ntgiwsvp, ordmoj, sagar pahwa, pritamsinha2304, ryancocuzzo, saktar unr, saswati qiskit, septembrr, sethmerkel, sg495, shaashwat, smturro2, sternparky, strickroman, tigerjack, tsura crisaldo, upsideon, vadebayo49, welien, willhbang, wmurphy collabstar, yang.luh, and M. Čepulkovskis, "Qiskit: An open-source framework for quantum computing," 2021.

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[6] A. Borras, A. P. Majtey, A. R. Plastino, M. Casas, and A. Plastino, "Robustness of highly entangled multiqubit states under decoherence," *Phys. Rev. A*, vol. 79, p. 022108, Feb 2009. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.79.022108

[7] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Error mitigation with Clifford quantum-circuit data," *Quantum*, vol. 5, p. 592, Nov. 2021. [Online]. Available: https://doi.org/10.22331/q-2021-11-26-592

[8] P. Das, S. Tannu, S. Dangwal, and M. Qureshi, "Adapt: Mitigating idling errors in qubits via adaptive dynamical decoupling," in *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 950–962. [Online]. Available: https://doi.org/10.1145/3466752.3480059

[9] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications," *Physical Review X*, vol. 8, no. 3, p. 031027, 2018.

[10] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, "Digital zero noise extrapolation for quantum error mitigation," in *QCE*, 2020.

[11] R. LaRose, A. Mari, S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. E. Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, P. Thakre, N. Shammah, and W. J. Zeng, "Mitiq: A software package for error mitigation on noisy quantum computers," 2020.

[12] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Physical Review X*, vol. 7, no. 2, 2017.

[13] J. Liu and H. Zhou, "Reliability modeling of nisq-era quantum computers," in *IISWC*, 2020.

[14] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Unified approach to data-driven quantum error mitigation," *Phys. Rev. Research*, vol. 3, p. 033098, Jul 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.3.033098

[15] P. Murali, D. C. Mckay, M. Martonosi, and A. Javadi-Abhari, "Software mitigation of crosstalk on noisy intermediate-scale quantum computers," ser. ASPLOS '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1001–1016. [Online]. Available: https://doi.org/10.1145/3373376.3378477

[16] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, "Scalable mitigation of measurement errors on quantum computers," *PRX Quantum*, vol. 2, p. 040326, Nov 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PRXQuantum.2.040326

[17] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[18] S. Nishio, Y. Pan, T. Satoh, H. Amano, and R. V. Meter, "Extracting success from ibm's 20-qubit machines using error-aware compilation," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 16, no. 3, pp. 1–25, 2020.

[19] T. Patel and D. Tiwari, "Qraft: Reverse your quantum circuit and know the correct program output," in *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 443–455. [Online]. Available: https://doi.org/10.1145/3445814.3446743

[20] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, "Simulating large quantum circuits on a small quantum computer," *Phys. Rev. Lett.*, vol. 125, p. 150504, Oct 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.125.150504

[21] W. Tang, T. Tomesh, M. Suchara, J. Larson, and M. Martonosi, "Cutqc: Using small quantum computers for large quantum circuit evaluations," in *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 473–486. [Online]. Available: https://doi.org/10.1145/3445814.3446758

[22] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," *Phys. Rev. Lett.*, vol. 119, Nov 2017. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.119.180509

[23] T. Tomesh, P. Gokhale, V. Omole, G. Ravi, K. N. Smith, J. Viszlai, X. Wu, N. Hardavellas, M. R. Martonosi, and F. T. Chong, "Supermarq: A scalable quantum benchmark suite," in *2022 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. Los Alamitos, CA, USA: IEEE Computer Society, apr 2022, pp. 587–603. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/HPCA53966.2022.00050

[24] V. Tripathi, H. Chen, M. Khezri, K.-W. Yip, E. Levenson-Falk, and D. A. Lidar, "Suppression of crosstalk in superconducting qubits using dynamical decoupling," *Phys. Rev. Applied*, vol. 18, p. 024068, Aug 2022. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.18.024068

[25] L. Viola, E. Knill, and S. Lloyd, "Dynamical decoupling of open quantum systems," *Phys. Rev. Lett.*, vol. 82, pp. 2417–2421, Mar 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.82.2417

[26] H. Wang, Z. Liang, J. Gu, Z. Li, Y. Ding, W. Jiang, Y. Shi, D. Z. Pan, F. T. Chong, and S. Han, "Torchquantum case study for robust quantum circuits," in *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design, ICCAD 2022, San Diego, California, USA, 30 October 2022 - 3 November 2022*, T. Mitra, E. F. Y. Young, and J. Xiong, Eds. ACM, 2022, p. 136. [Online]. Available: https://doi.org/10.1145/3508352.3561118

[27] C. J. Wood, "Special session: Noise characterization and error mitigation in near-term quantum computers," in *2020 IEEE 38th International Conference on Computer Design (ICCD)*, 2020, pp. 13–16.

[28] J. Zhang, A. M. Souza, F. D. Brandao, and D. Suter, "Protected quantum computing: Interleaving gate operations with dynamical decoupling sequences," *Phys. Rev. Lett.*, vol. 112, p. 050502, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.112.050502