

Comparison of Pulse Stretching and Gate Insertion for Zero Noise Extrapolation

Eesh Gupta*
Rutgers University
(Dated: January 5, 2021)

ABSTRACT

Some of the most promising near-term applications of quantum computers lie in solving classically intractable problems in chemistry, leading to potential breakthroughs in, for example, disease prevention. However, quantum computers of today are both highly susceptible to errors and limited by the number of qubits to correct them. For the near term, scientists are exploring error mitigation techniques to reduce the effects of quantum computational noise. One such popular technique is Zero Noise Extrapolation (ZNE) which approximates the ‘zero error’ solution by deliberately increasing the noise in the device. Yet precise noise amplification remains a challenge. In this paper, we compare two such techniques, pulse stretching and gate insertion, in mitigating noise in computation of ground states of simple molecules.

I. OUTLINE

Accurate simulation of molecular systems has been a defining goal of modern chemistry. Through probing energies of molecules, we can predict their properties and their interactions with other systems which are of great industrial importance. Consuming 2% of world’s total energy output, nitrogen fixation could be made less expensive if we understood the structure of a catalyst called nitrogenase. Similarly, complete analysis of copper-oxygen layers in high temperature cuprate superconductors could deliver advances in renewable energy[14]. Chemical simulations of the energy levels of such compounds rely on solving the time independent Schrodinger equation. While many computational methods have been proposed, this task requires resources scaling exponentially with the size of system, making it classically intractable for systems with more than 100 electrons.

However using quantum computers, Peruzzo et al. [5] have shown that the task of finding molecular ground state energies is achievable. Their proposed algorithm, Variational Quantum Eigensolver (VQE), re-

lies on quantum computers to efficiently prepare and measure the energies of parameterized wavefunctions, while trusting classical computers with the parameter optimization. Scaling polynomially with the number of resources, this algorithm is promising. However, due to errors in near term quantum devices, VQE seldom outputs energies within chemical accuracy. While the quantum computing literature is rich with error correction protocols, most methods are too expensive in number of qubits to be practical on near-term, noisy devices.

Given these constraints, there has been a large effort in mitigating errors with minimal overhead of resources. One such technique is Zero Noise Extrapolation (ZNE) which aims to first amplify noise and then use these noisy energies to improve upon the result [7]. Here, proper amplification of noise is crucial and done using either pulse stretching or gate insertion. In pulse stretching, we rescale the time needed to implement gates by stretching the microwave pulses. If we assume that the noise is constant in time, then rescaling the circuit-time will rescale the amount of noise that acts on that circuit. On the other hand, in gate insertion, we add redundant identity gates UU^\dagger after every unitary gate U . These insertions not only increase the number of noisy gates, but also increase the circuit-time, scaling the overall noise in the quantum circuit.

The goal of our work is to compare the performance of these noise amplification techniques in extrapolating to zero noise energies in VQE computations. For sake of simplicity, we only extrapolate after obtaining the optimized parameters. Nevertheless, the experiments discussed are instructive in highlighting the merits and demerits of each technique. This paper is organized as follows. In section ??, we begin with a brief overview of classical chemistry methods and a summary of the VQE algorithm. In section ??, we explore a few quantum errors and do a study on their effects on VQE computations. In section ??, we provide the necessary background on zero noise extrapolation as well as the noise amplification techniques. The remaining sections discuss the performance of the two techniques.

II. QUANTUM CHEMISTRY BACKGROUND

The aim of quantum chemistry is to solve the time-independent Schrodinger equation for molecular systems and hence predict their properties. If we ignore the

* Correspondence email address: eesh.gupta@rutgers.edu

interactions between electrons in these systems, then classical approaches like Hartree Fock theory are able to solve those equations and produce energy eigenvalues. But for strongly correlated systems of great industrial importance, these computed energies don't achieve chemical accuracy. On the other hand, if we do account for interactions between electrons, even the best classical chemical methods require resources scaling exponentially with the size of molecules. In other words, exact simulation of large molecular systems is classically intractable.

To relieve classical computers of this burden, Peruzzo et. al [5] proposed a hybrid quantum-classical algorithm to compute the ground state energies of molecular systems. The Variational Quantum Eigensolver is an iterative algorithm that relies on quantum computers to prepare and measure parameterized states while using classical computers to optimize the parameters. The algorithm stops when the energy cannot be minimized any further. Assuming a good starting point and an ideal quantum computer, the resultant energy is within chemical accuracy.

In the following sections, we discuss classical methods and the various stages of VQE in a little more detail.

A. Classical Chemistry Methods

Consider a molecular system of N electrons interacting with nucleus fixed at the centre. Since the nucleons are larger than electrons by 4 order of magnitude, we will assume them to have no kinetic energy (Born Oppenheimer Approximation). Then the Hamiltonian for the system can be written as

$$H = H_{\text{core}} + \sum_i^N \sum_{j>i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

Here H_{core} consists of the kinetic of electrons, coulomb repulsions among nucleons and coulomb attraction between electrons and nucleons. If the H was equal to this term alone, then we could write N Schrodinger equations, one for each electron, and solve them independently to obtain N wavefunctions. However, the second term of the molecular Hamiltonian H denoting the electron-electron repulsion renders those N equations coupled. For example, the solution for first equation $\psi(\mathbf{r}_1)$ depends on the solutions of the other $N - 1$ equations $\{\psi(\mathbf{r}_2), \psi(\mathbf{r}_3), \dots, \psi(\mathbf{r}_N)\}$ because the region occupied by electron 1 is affected by the positions of the other $N - 1$ electrons.

1. Hartree Fock Theory

In the Hartree Fock method, we average over the repulsions felt by each electron due to other electrons in the system. For example, suppose we only have 2 electrons, electron 1 in spin orbital a and electron 2 is spin orbital b. Then the two electron operator describing their interaction $\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}$ should depend on both the spatial coordinates of electron 1 \mathbf{r}_1 and that of electron 2 \mathbf{r}_2 . However, Hartree Fock theory converts that two electron operator into the following one electron operator

$$v(\mathbf{r}_1) = \int \frac{|\psi_b(\mathbf{x}_2)|^2}{\mathbf{r}_1 - \mathbf{r}_2} d\mathbf{x}_2$$

(equation 3.7 in [1]). Here we compute the repulsions by averaging over all the spin and spatial coordinates of electron 2 \mathbf{x}_2 .

While this treatment may save some computational resources, it does not get rid of the coupling between the N Schrodinger equations. In the previous case, we need the solution for the second electron $\psi_b(\mathbf{x}_2)$ to construct the hamiltonian operator for the first electron. Thus, we use the following iterative technique.

1. *Guess:* From a given set of M Spin Orbitals , choose N .
2. *Construct:* Using the N spin orbitals $\{\psi_i\}$, construct the fock operator for each electron i.e. the electronic Hamiltonian with the mean field approximation as described above.
3. *Solve:* Using these N fock operators, solve the Schrodinger equation for each electron. Call these solutions $\{\chi_i\}$.
4. *Check* If the previous step does not produce the spin orbitals used to construct the operators in step 2 i.e. $\psi_i \neq \chi_i$ for any i, repeat step 2 with $\{\chi_i\}$ as the initial spin orbitals.
5. *Self Consistency Achieved!* Otherwise, end algorithm and output $\{\chi_i\}$

Note that we first input a set of atomic orbitals to this algorithm and the resultant molecular orbitals are combinations of those initial single particle atomic orbitals [14]. These resultant wavefunctions correspond to the *ground* state of the molecule. They are *self consistent* in the sense that they are themselves the solutions to the operators that they generated. Although the energy corresponding to this configuration accounts for 99% of the total energy, the remaining 1% from the certain electron correlation effects is however important to be within chemical accuracy.

2. Post Hartree-Fock Methods

To improve upon Hartree Fock theory, we must do away with "pigeon-holing" electrons into spin orbitals. Even in the ground state, electrons don't *always* occupy the lowest energy spin orbitals. Some may be excited to higher energy spin orbitals due to their interactions with others. Recognizing this notion, Full Configuration Interaction (FCI) accounts for all the possible ways of arranging electrons in $2K$ spin orbitals and produces the exact solution within a given basis. Within the second quantization framework, we can express the FCI wavefunction as

$$\Psi_{FCI} = \Phi_0 + T_1\Phi_0 + T_2\Phi_0 + \dots$$

where

$$T_1 = \sum_{\substack{i \in \text{occ} \\ a \in \text{virt}}} t_a^i a_a^\dagger a_i$$

$$T_2 = \sum_{\substack{i > j \in \text{occ} \\ a > b \in \text{virt}}} t_{ab}^{ij} a_a^\dagger a_b^\dagger a_i a_j$$

. Here Φ_0 is the Hartree Fock state with all electrons in lowest energy spin orbitals, $T_1\Phi_0$ are the singly excited states, $T_2\Phi_0$ are the double excited states and so on. In the ground state, we would not expect the doubly excited states to be as important as the hartree fock state i.e. $t_{ab}^{ij} 0$. But in accounting for these excited states, we are more likely to achieve chemical accuracy. The goal then is to compute the coefficients t 's. However, the factorial dependence of the number of states on the number of spin orbitals makes FCI classically intractable. A possible approximation may involve truncating excitations. For example, Configuration Interactions Singles and Doubles (CISD) ignores excited states larger than doubles.

$$\Psi_{CISD} = (1 + T_1 + T_2)\Phi_0$$

Tractable yet inaccurate for strongly correlated systems, this method can be refined to accommodate the higher excited states. Promoting the linear operator to an exponential, we obtain

$$\Psi_{CCSD} = e^{T_1+T_2}\Phi_0$$

where the expansion

$$e^{T_1+T_2} = 1 + T_1 + (T_1^2/2 + T_2) + \dots$$

generates approximations to higher excited states. For example T_1T_2 achieves a triple excitation. Known as Coupled Cluster Singles and Doubles, this method differs from FCI in minimizing the number of parameters t 's. Although this method still fails for large molecular systems, we will use this method along with quantum computers to achieve chemically accurate simulations for smaller systems.

B. Variational Quantum Eigensolver (VQE)

1. State Preparation

The convergence of the optimization algorithm depends on the choice of initial state and parameterization. Traditionally, we would use Hartree Fock as our starting point and Unitary Coupled Cluster Singles and Doubles wavefunction for parameterization. To illustrate, consider a system of ν electrons and N spin orbitals. If we rely on Jordan Wigner encoding to map our chemistry problem onto the quantum computer, each qubit corresponds to one of the N spin orbitals. Then preparing the Hartree Fock state $|\Phi\rangle = a_1^\dagger a_2^\dagger \dots a_n^\dagger |0\rangle$ translates into initializing a set of ν qubits in the $|1\rangle$ state i.e $|1\rangle^{\otimes \nu} |0\rangle^{\otimes N-\nu}$ [9]. Here qubits in $|1\rangle$ are *occupied* and those in $|0\rangle$ are unoccupied i.e. represent the possible spin orbitals that electrons could be excited to. After that initial iteration the VQE procedure, we act on $|\Phi\rangle$ with the UCCSD operator e^T nontrivially where T is dictated by the parameters $\mathbf{t} = (t_1, \dots, t_k)$.

Initializing in the Hartree Fock state ensures proximity to the global minimum. And for small molecules like H_2 , use of *UCCSD* ansatz on ideal quantum computers gives us ground state energies within chemical accuracy. However, for larger molecular systems, these chemistry methods may strain computational resources. One may suggest to use a different quantum encoding scheme. Whereas Jordan Wigner requires $\mathcal{O}(N)$ qubits to simulate N spin orbitals, there exists Bravyi Kitaev encoding that requires only $\mathcal{O}(\log n)$ qubits [3]. Yet, even with this efficient encoding method, the number of gate required for mapping the UCCSD operator grow as $\mathcal{O}(N^2\nu^2)$, rendering it impractical for even systems as small as LiH .

Motivated to control the gate count, Kandala et. al [6] proposed a hardware efficient ansatz which, along with error mitigation strategies, was found to be effective for small molecular systems [10]. This method dismisses the details of chemical system and treats parameters as tools to explore the Hilbert space rather than weights of different excitations as in UCCSD. While this approach yields promising results for small systems, further work is required to establish its efficacy for larger systems. Further, McClean et. al [8] challenge this approach by underscoring the irregularities associated with random initial parameters. Unlike UCCSD, Kandala's approach cannot efficiently use Hartree Fock as its initial state. Hence, his method is limited to randomly initialized states, which for large number of qubits renders the energy gradient to be nonzero only for a small subset of directions, increasing the likelihood of being stuck in local minima.

2. Energy Measurement

Once a state has been prepared, we act on it with the Hamiltonian H and measure the associated energy eigenvalue. Before mapping this operator onto quantum computers, we first convert it to its second quantization form

$$H = h_n u c + \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pq} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s$$

where $\{a_i^\dagger, a_i\}$ are the creation and annihilation operators [9], $h_n u c$ is nuclear repulsion energy, h_{pq} correspond to single electron integrals and h_{pqrs} correspond to two electron integrals. These creation and annihilation operator can be mapped onto a quantum circuit using pauli matrices. For example, in Jordan Wigner encoding, the following operator adds an electron to the first spin orbital:

$$a_1^\dagger = \frac{X - iY}{2} \otimes I \otimes I \dots \otimes I$$

Since the first spin orbital corresponds to the first qubit in this encoding scheme, the creation operator a_1^\dagger only acts on the first qubit nontrivially. In a similar fashion, we can map each term in the hamiltonian, using some approximations as detailed in [4], as

$$O_i = \sigma_0 \otimes \sigma_1 \dots \otimes \sigma_{N-1} \text{ where } \sigma_i \in \{I_i, X_i, Y_i, Z_i\}$$

Here, σ_i is a pauli matrix acting on the i 'th qubit. Measuring the expectation value of each term or local Hamiltonian O_i gives the energy eigenvalue corresponding to the prepared state

$$E = \langle H \rangle = \sum_i h_i \langle O_i \rangle$$

For example, in the case of the hydrogen molecule using parity basis qubit encoding, the qubit hamiltonian is

$$H = -1.05II + 0.04IZ - 0.04ZI - 0.011ZZ + 0.181XX$$

3. Parameter Optimization

Once we measure the energy corresponding to a state, we require a classical optimizer to improve upon this result. In each iteration, the optimizer suggests a set of parameters, receives the computed energy and updates the parameters to minimize that energy. The algorithm converges once the difference between computed energies in successive iterations are within some threshold. Note that this is just one example of a stopping condition for the algorithm.

Fortunately, there are numerous algorithms in the literature that can assist in such optimization of energy. However, the chosen algorithm must meet the following criteria for efficient optimization, as detailed in [13]:

1. Ability to find a global minima in the presence of quantum noise
2. Scalability with the number of parameters
3. Small number of samples from quantum circuit measurements

We also want the algorithm to converge with minimal iterations to avoid unwanted effects from T_1, T_2 fluctuations. Popular algorithms include Constrained Optimization by Linear Optimization (COBYLA) and Simultaneous Perturbation by Stochastic Approximation (SPSA). Analyzing the effects of noise and error mitigation techniques on optimization is however outside the scope of this paper.

III. NOISE AND VQE

A. Types of Noise

In this section, we will only cover the major sources of noise, ignoring correlated as well as non-physical errors.

1. Coherent

In IBM's quantum computers, the basis set of gates consists of $U1(\lambda)$, $U2(\phi, \lambda)$, and $U3(\theta, \phi, \lambda)$ gates, all specified in terms of angles. Due to experimental errors, microwave pulses implementing these gates may cause an over-rotation or under-rotation by some angle $\delta\theta, \delta\phi, \delta\lambda$. On IBMQ devices, we find single qubit gate errors rates to be $0.1 - 0.5\%$ and two qubit gate error rates to be $1 - 2\%$.

2. Incoherent

Incoherent errors are broadly classified into bit flips and phase flips. A bit flip corresponds to $|0\rangle \rightarrow |1\rangle$ and a $|1\rangle \rightarrow |0\rangle$. On the other hand, phase flips corresponds to $|+\rangle \rightarrow |-\rangle$ and $|-\rangle \rightarrow |+\rangle$. Simple and elegant, this description may falsely suggest incoherent errors as discrete processes. Thus, a better classification of these errors is in terms of the following continuous processes:

- *Amplitude Damping*: Since quantum systems lose inevitably lose energy to the environment, it is natural to expect the excited state $|1\rangle$ decaying

to the lower energy state $|0\rangle$ over time. Assuming environment to be at zero temperature (same temperature as $|0\rangle$), qubits in state $\text{ket}0$ should, however, not be affected **I don't like this claim.** This process is called amplitude damping, transforming the bloch vector as

$$\begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \rightarrow \begin{bmatrix} e^{-t/2T_1}r_x \\ e^{-t/2T_1}r_y \\ (1 - e^{-t/T_1}) + e^{-t/T_1}r_z \end{bmatrix}$$

where t is time and T_1 is the speed of the process [10.5555/1972505]. If $r_z = 1$ or the bloch vector points toward the north pole ($|0\rangle$), then r_z is invariant to this transformation. Otherwise, r_z decays and is gradually replaced by $r_z = 1$. Hence, the *flow* is directed towards north pole. If environment were to be at finite temperature, this *flow* would be directed towards the mixed state. Such general amplitude damping describes the T_1 relaxation process.

- *Phase Damping* Due to qubit's interaction with stray electromagnetic fields, information about the relative phase of its state may be lost. In this case, the bloch vector suffers the following transformation:

$$\begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \rightarrow \begin{bmatrix} e^{-t/2T_2}r_x \\ e^{-t/2T_2}r_y \\ r_z \end{bmatrix}$$

The x and y components contract as the z component is left unaffected. Then gradually, a $|+\rangle$ or a $|-\rangle$ state will become a superposition of $|+\rangle$ and $|-\rangle$ states. Note how r_x and r_y components also shrink in amplitude damping. That case however is different because the *flow* is directed toward $|0\rangle$ state which has neither x nor y components. Here, the *flow* is directed more generally towards the z axis.

3. Readout

Measurement errors are similar to bit flip errors in that we may read $|0\rangle$ as $|1\rangle$ and $|1\rangle$ as $|0\rangle$ when measuring states. A possible reason for these errors is that measurement times ($5\mu s$) is not significantly smaller than decoherence times ($60\mu s$), leading to bit flip errors. Another reason may be overlapping in the physical quantities used to distinguish $|0\rangle$ from $|1\rangle$ [12]. On IBMQ machines, these errors usually range from 1–2% but can sometimes be found to be as high as 30%.

4. Leakage

Transmon qubits constructed using Josephson Junction are reffered to as anharmonic quantum oscillators because they don't have equally spaced energy states. For example, the "jump" from $|0\rangle$ to $|1\rangle$ (roughly 5 GhZ)is larger in terms of energy than that from $|1\rangle$ to $|2\rangle$ (roughly .3Ghz. These unequal spacings enable us to use such superconductors as two level systems. However, due to experimental errors, some qubit transformation may result in qubit being in $|2\rangle$ state i.e. qubit leaks into the higher energy levels of the oscillator. Such errors can be avoided by proper pulse shaping.

B. Impact of Noise on VQE

We begin our investigation by analyzing how different noise sources affect VQE computations. Since different noise types cannot be isolated on the real device, the following experiments were performed on Qiskit noise model FakeBogota. Within this simulator, we can control incoherent, coherent and readout errors. However, we cannot control the underlying probabilistic or stochastic noise due to the nature of quantum measurement. Note that coherent errors in noise model are simulated via depolarizing errors, which probabilistically act after gate operations to dissolve the existing qubit states into completely mixed state.

Preparation of the UCCSD state for H_2 molecule requires three parameters. Randomly selecting these parameters, each from the range $(-\pi, \pi)$, we measure the corresponding energies under both ideal and noisy conditions. Then we plot the difference between these energies for each of the 10^3 samples. Results from figure ?? shows that incoherent noise has the least effect on changing the spread of data. On the other hand, readout errors increase the spread as well as introduce a positive bias. This shift can attributed to an assymmetry in readout errors: 1 is more likely to be measured as 0 than 0 being measures as 1.

However, if we sample parameters *close* to parameters optimized on the noise model, all three noise sources result in a positive bias as shown in Figure ???. These shifts highlight the issue central to error mitigation in quantum chemistry: ground state energy computed by VQE on noisy quantum computers can be chemically inaccurate by more than two orders of magnitude. To mitigate such effects of noise, we look at Zero Noise Extrapolation (ZNE).

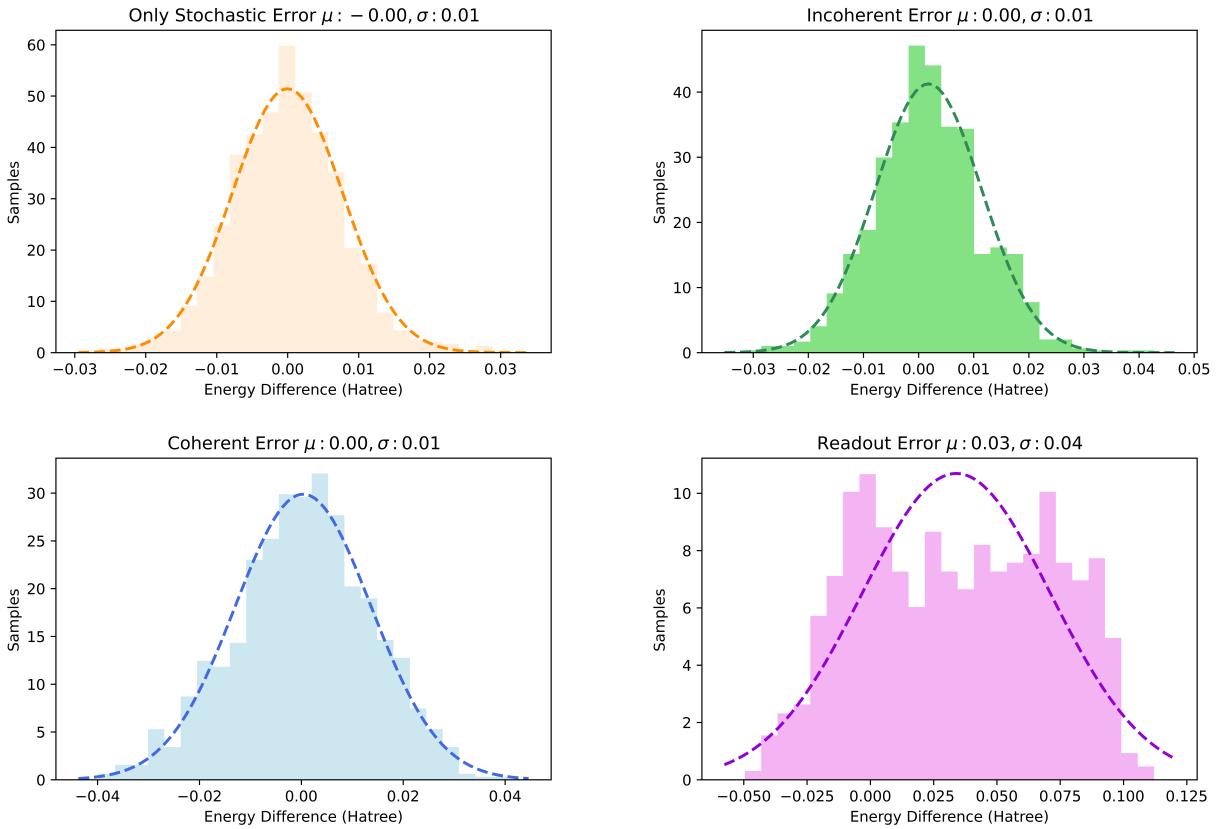


Figure 1: Impact of Noise on UCCSD circuit for hydrogen atom with randomly chosen parameters.

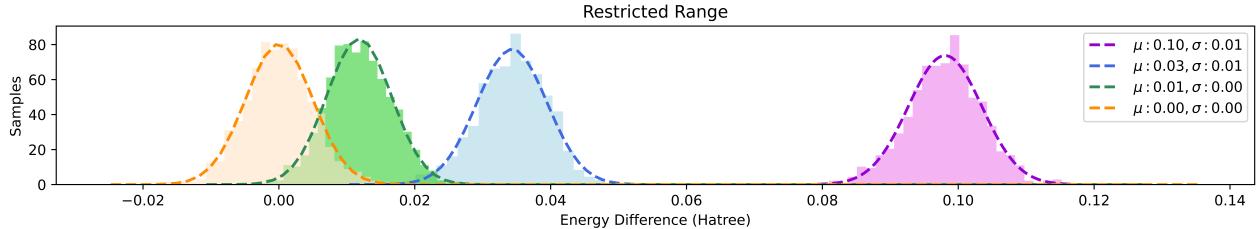


Figure 2: Impact of Noise on UCCSD circuit for hydrogen atom with parameters chosen from a tight range about their optimized values.

IV. ZERO NOISE EXTRAPOLATION

Suppose the expectation value of an observable can be expressed in terms of the error parameter $\lambda \ll 1$ as the following power series

$$E(\lambda) = E^* + \sum_{k=1}^n a_k \lambda^k + \mathcal{O}(\lambda^{n+1}) \quad (1)$$

where E^* is the noiseless energy. Our goal in Zero Noise Extrapolation is to better approximate E^* by canceling out increasing powers of λ . To do so, Temmet et al. [7] proposed computing the expectation value at different

noise rates $c_i \lambda$ where $c_0 = 1 < c_1 < c_2 < \dots < c_n$. Then using Richardson Extrapolation, we combine those expectation values such that

$$E_n = \sum_{i=0}^n \gamma_i E(c_i \lambda) = E^* + \mathcal{O}(\lambda^{n+1}) \quad (2)$$

The coefficients γ_i are determined by the relations

$$\sum_{i=0}^n \gamma_i = 1 \quad (3)$$

,

$$\sum_{i=0}^n \gamma_i c_i^k = 0 \text{ for } k = 1, \dots, n \quad (4)$$

The first equation (??) ensures that the coefficient of E^* in equation (??) is unity, and the system of equations encoded by the second equation (??) cancels out terms containing $\lambda, \lambda^2, \dots, \lambda^n$. Thus, we have improved the error bound on our approximation for E^* from $\mathcal{O}(\lambda)$ to $\mathcal{O}(\lambda^{n+1})$.

A. Polynomial Fitting and Richardson Extrapolation

In cases where c_i are sufficiently small and the number of scaled energies n is large, using Richardson Extrapolation may produce a worse estimate of than the unamplified noisy energy $E(\lambda)$. This occurs because n 'th order Richardson extrapolation is equivalent to fitting $n + 1$ points through a n 'th degree polynomial f_n and computing $f_n(0)$. To see this relationship between polynomial fitting and Richardson method, consider the following example.

Suppose we are given the expectation value at $c_0 = 1$ and $c_1 = 2$.

$$\begin{aligned} E(\lambda) &= E^* + a_1 \lambda + \mathcal{O}(\lambda) \\ E(2\lambda) &= E^* + a_1(2\lambda) + \mathcal{O}(\lambda) \end{aligned}$$

Following the Richardson method, to get cancel the λ term so that $E_1 = \gamma_0 E(\lambda) + \gamma_1 E(2\lambda) = E^* + \mathcal{O}(\lambda^2)$, we need

$$\begin{aligned} \gamma_0 + \gamma_1 &= 1 \\ \gamma_0 + 2\gamma_1 &= 0 \end{aligned}$$

Subtracting the first equation from the second, we obtain the solution $\{\gamma_0 = 2, \gamma_1 = -1\}$ and thus, $E_1 = 2E(\lambda) + E(\lambda)$. Another way to arrive at the same solution is by connecting the points $E(\lambda), E(2\lambda)$ on the $E - \lambda$ plane and computing the y intercept of the line. The slope of such a line would be

$$m = \frac{E(2\lambda) - E(\lambda)}{2\lambda - \lambda} = \frac{E(2\lambda) - E(\lambda)}{\lambda}$$

and y intercept could be found by

$$\begin{aligned} E(\lambda) &= m\lambda + b = E(2\lambda) - E(\lambda) + b \\ b &= 2E(\lambda) - E(\lambda) \end{aligned}$$

So Richardson extrapolation of two points is equivalent to computing the y intercept b of the line connecting the two points.

If we accept this relationship between polynomial fitting and Richardson technique, then in some cases, using n 'th degree polynomial to fit $n + 1$ points may result in overfitting data and hence fail to produce an accurate zero noise value. In these circumstances, fits of lower degree polynomials will be preferable. To illustrate, assume that c_i 's are small so that a linear fit best predicts the zero noise value. Then instead of canceling terms involving $\lambda, \lambda^2, \dots, \lambda^n$ from equation [??], we will only remove the λ term. Such cancellation entails

$$\begin{aligned} \gamma_0 + \gamma_1 + \dots + \gamma_n &= 1 \\ c_0\gamma_0 + c_1\gamma_1 + \dots + c_n\gamma_n &= 0 \end{aligned}$$

With unknowns outnumbering the number of equations, we have an undetermined linear system which has infinite number of solutions (assuming distinct stretch factors $\{c_i\}$). However, through a technique called singular value decomposition, we can pick a solution $\gamma = (\gamma_0, \dots, \gamma_n)$ which minimizes $\|\gamma\|$. This least norm solution γ minimizes the coefficients of $\lambda^2, \lambda^3, \dots$ terms and hence reduces the weight of the error bound $\mathcal{O}(\lambda^2)$. Applying these coefficients γ upon the scaled energies, we get

$$E_n = \sum_{i=0}^n \gamma_i E(c_i \lambda) = E^* + \mathcal{O}(\lambda^2) \quad (5)$$

as desired. Alternatively, we could have fitted the $n + 1$ points through a least squares linear fit and the resulting y intercept would have yielded a similar result.

B. Noise Amplification

The central issue of Zero Noise Extrapolation lies in proper amplification of noise. Below we discuss two popular methods of doing so: pulse stretching and gate insertion.

1. Pulse Stretching

As the name suggests, this noise amplification technique rescales the time taken by the microwave pulses to implement gates in a circuit. If a circuit takes time T to run, then the "stretched" circuit will take time cT . If noise is constant in time and $cT < T_1$ (i.e. within coherence budget), then Temme et. al [7] claim that a circuit with rescaled pulses taking time cT is equivalent to that same circuit with rescaled noise parameter $c\lambda$ taking time T . To prove this claim, we will walk through some math developed in Section II of supplementary materials in [7].

A circuit can be represented as the following multi qubit hamiltonian.

$$K(t) = \sum_{\alpha} J_{\alpha}(t) P_{\alpha} \text{ where } P_{\alpha} \in \langle I, X_j, Y_j, Z_j \rangle_{j=1,\dots,N}$$

Even though Temme et. al consider X_j, Y_j, Z_j as Pauli matrices acting on the j'th qubit, we will consider them as rotation operators about the respective axis on the bloch sphere. Then, the coupling coefficients $J_{\alpha}(t)$ act as dimming switches for a rotation operators, activating and deactivating rotation operators to implement gate operations on the quantum state ρ . In other words, if $J_{\alpha}(t)$ was a gaussian pulse, then the area underneath denotes the amount of rotation.

Using this hamiltonian, we can express the evolution of a quantum state ρ as

$$\frac{\partial}{\partial t} \rho(t) = -i[K(t), \rho(t)] + \lambda \mathcal{L}(\rho(t)) \quad (6)$$

If the RHS were just composed of the commutator term, then the equation would be analog of schrodinger equation with a density matrix instead of a wavefunction(also known as Louiville equation). The term $\mathcal{L}(\rho)$ is the time independent Lindblad operator which generates noise (assuming $\lambda \ll 1$). Temme et. al also assume that the operator does not depend on $K(t)$ and $J_{\alpha}(t)$. Integrating both sides now,

$$\rho_{\lambda}(T) = \rho(0) - i \int_0^T [K(t), \rho(t)] dt + \lambda \int_0^T \mathcal{L}(\rho(t)) dt \quad (7)$$

where $\rho(0)$ is the initial state of the system. Now if we rescale time T as $T' = cT$, then the new coupling coefficient $J'_{\alpha}(t)$ can be expressed as

$$J'_{\alpha}(t) = c^{-1} J_{\alpha}(t/c)$$

In stretching a Gaussian pulse without changing the area under the graph, we would squish its height. Hence J_{α} term incurs a factor of c^{-1} . Further, $\rho'(t) = \rho(t/c)$ where $\rho'(t)$ is the state of the rescaled circuit after time t .

Thus, we can rewrite the integral for the rescaled circuit as

$$\rho'_{\lambda}(T') = \rho(0) - i \int_0^{cT} [K(t), \rho(t)] dt + \lambda \int_0^{cT} \mathcal{L}(\rho'(t)) dt \quad (8)$$

Switching the dummy variable t with ct' ,

$$\begin{aligned} \rho'_{\lambda}(T') &= \rho(0) - i \int_0^T \sum_{\alpha} c^{-1} J_{\alpha}(t') [P_{\alpha}, \rho(t')] c dt' + \\ &\quad \lambda \int_0^T \mathcal{L}(\rho(t')) c dt' \end{aligned}$$

$$\rho'_{\lambda}(T') = \rho(0) - i \int_0^T \sum_{\alpha} J_{\alpha}(t') [P_{\alpha}, \rho(t')] dt' + \quad (9)$$

$$c \lambda \int_0^T \mathcal{L}(\rho(t')) dt' \quad (10)$$

$$\rho'_{\lambda}(T') = \rho(0) - i \int_0^T [K(t'), \rho(t')] dt' + c \lambda \int_0^T \mathcal{L}(\rho(t')) dt' \quad (11)$$

$$\rho'_{\lambda}(T') = \rho_{c\lambda}(T) \quad (12)$$

Thus, rescaling the pulses is equivalent to rescaling the noise rate λ under the assumption that noise is invariant to such rescaling of time. However, Kandala et. al [10] demonstrated that T_1, T_2 times do fluctuate over time, challenging the assumption made earlier. To ensure its validity, all pulse stretching experiments should occur quickly in succession.

2. Gate Insertion

Gate Insertion relies on insertion of redundant gates in the circuit to amplify the noise. Recall that for any unitary gate U , $U^\dagger U$ constitutes the identity. Then if we insert n_i pairs of $U^\dagger U$ after the unitary U , we will obtain a total of $2n_i + 1$ gates. Assuming that each gate has some error rate λ , the addition of n_i such identities results in an overall error rate of $(2n_i + 1)\lambda$. Here, we have made another assumption: errors add up linearly with the number of gates. But if the errors are incoherent in nature, the errors would add up exponentially. So if error is of the form $e^{-\phi T}$ for a single qubit gate X , the total error due to $XX^\dagger X$ would be $e^{-3\phi T}$ (Here T is the time taken to implement the X gate). This inconsistency between the different types of errors presents a problem: how do the errors add up with gate insertion?

While there may many proposed models based on types and strength of noise sources, we generally want to avoid having any knowledge of the underlying noise model. In observing this notion, a popular approach constrains the magnitude of amplification. If $(2n_i + 1)\lambda \ll 1$ then a linear relationship between λ and n_i is a good approximation. Since λ is dependent on hardware, our only hopes lie in minimizing n_i . Indeed, even if we ignore this issue of error scaling, n_i must be small to ensure that the final circuit is within the coherence window. But even for moderately sized circuits, n_i values as small as 1, 2, 3 result in circuits exceeding T_1, T_2 times. Given that n_i can only take positive integer values, how then can we minimize the number of identities n_i in the quantum circuit?

A recent paper by He et al. [11] proposes to treat n_i as a Poisson random variable. In other words, we will choose n_i from the distribution $\text{Poisson}(\nu)$ for each gate in the circuit. The mean ν of the distribution will be chosen such that n_i is minimized and hence the stretch factor $c = 2n_i + 1$ is closer to 1. For example, suppose we want to amplify the overall noise rate by a factor of $c = 1.2$ in a circuit consisting of $M = 10$ gates. Since,

$$2n_i + 1 = c,$$

rearranging terms, we have

$$n_i = \frac{c - 1}{2} = 0.1$$

Since we can't directly insert a tenth of an identity after each gate, we sample $M n_i$'s from $\text{Poisson}(\nu = 0.1)$ distribution. In doing so, we might get a set of samples $\{0, 0, 0, 1, 0, 0, 0, 0, 0, 0\}$. While only one of the ten gates has an identity insertion, the average identity insertions per gate is 0.1, as desired for an overall amplification factor of $c = 1.2$. Thus, random identity insertion method (RIIM) allows us to do zero noise extrapolation with minimal number of gate insertions in our quantum circuit, ensuring that the overall circuit time is within the coherence budget.

However, there are still some issues with this noise amplification technique. First, at least on superconducting qubits, there is an order of magnitude difference between error rates of single and two qubit gates. So inserting $U_{CNOT}^\dagger U$ after a CNOT gate is different from inserting XX after an X gate. But RIIM method treats both cases as equivalent. A possible solution is to amplify only two qubit gates in a circuit. However, for problems like electronic structure which place tight bounds on accuracy, ignoring noise amplification of single qubit gates may screw up our results (Better verb needed). Thus, we need to assess the demands of the problem with hardware constraints before choosing this technique.

V. METHODS

To compare pulse stretching and gate insertions, we performed three sets of experiments on IBMQ's Casablanca device: two involving Randomized Benchmarking and the other concerning VQE.

In randomized benchmarking, we generate random sequences of either single qubit or two qubit clifford gates, adding up to identity. Ideally, we should measure the initial state, either $|0\rangle$ or $|00\rangle$, with 100% probability. However, due to coherent and incoherent errors, we are less likely to measure qubits in their initial state in larger circuits.

Given their simplicity and generality, randomized benchmarking protocols provide a good starting point for analysis of noise amplification techniques. We then proceed to the VQE for H_2 molecule. In principle, we would extrapolate the zero noise energy in every iteration of VQE algorithm. However, this procedure warrants analysis of different optimization algorithms which is outside the scope of our work. Hence, we run the VQE algorithm on a noisy simulator and compute the optimal parameters beforehand. With these parameters in hand, we can prepare the optimal UCCSD state and perform noise amplification.

Pulse Stretching is relatively straightforward, consisting of conversion of circuit to pulse form and rescaling of the Gaussian pulses. For gate insertion, we perform random identity insertion on each of the twenty gates in the circuit. Yet, twenty events will not be sufficient to suppress stochastic fluctuations. For example, our intended stretch factor may be $c = 1.5$ but after applying RIIM method, we find our stretch factor to be $\tilde{c} = 1.9$. One way of ensuring that $\tilde{c} \approx c$ is by increasing the number of events i.e. the number of gates acted upon by RIIM. To do so, we create $M = 40$ copies of the optimal circuit and apply RIIM independently on each one. Measuring each noise amplified circuit 8192 times and adding up the results, we have effectively acted RIIM on $20M$ gates and hence suppressed some stochastic fluctuations.

In addition to noise amplification techniques, we also apply measurement error mitigation scheme because it is reasonable to claim that neither noise amplification strategies will increase readout errors. Let's represent readout errors by a matrix R such that

$$C_{noisy} = RC_{ideal}$$

where C_{ideal} are the true counts and C_{noisy} are the measured counts from a circuit. For example, a two qubit circuit will have a C vector with entries for counts corresponding to the states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. Our job is to compute the matrix R and invert it such that

$$R^{-1}C_{noisy} = C_{ideal}$$

. The matrix R holds readout probabilities. Each entry R_{ij} corresponds to the probability that $|j\rangle$ is going to be read as $|i\rangle$. So diagonal entries correspond to proper reading of states and off diagonal elements correspond to errors. If there weren't readout errors, then R would be equal to the identity matrix since .

For two qubits, computing these entries requires measurements from four trivial circuits. In each circuit, you prepare a basis state and measure the outcomes. For example, in the second circuit, you may prepare $|01\rangle$ by acting on the second qubit with an X gate and measure the two qubits. You may get

$$\{00 : 5, 01 : 85, 10 : 5, 11 : 5\}$$

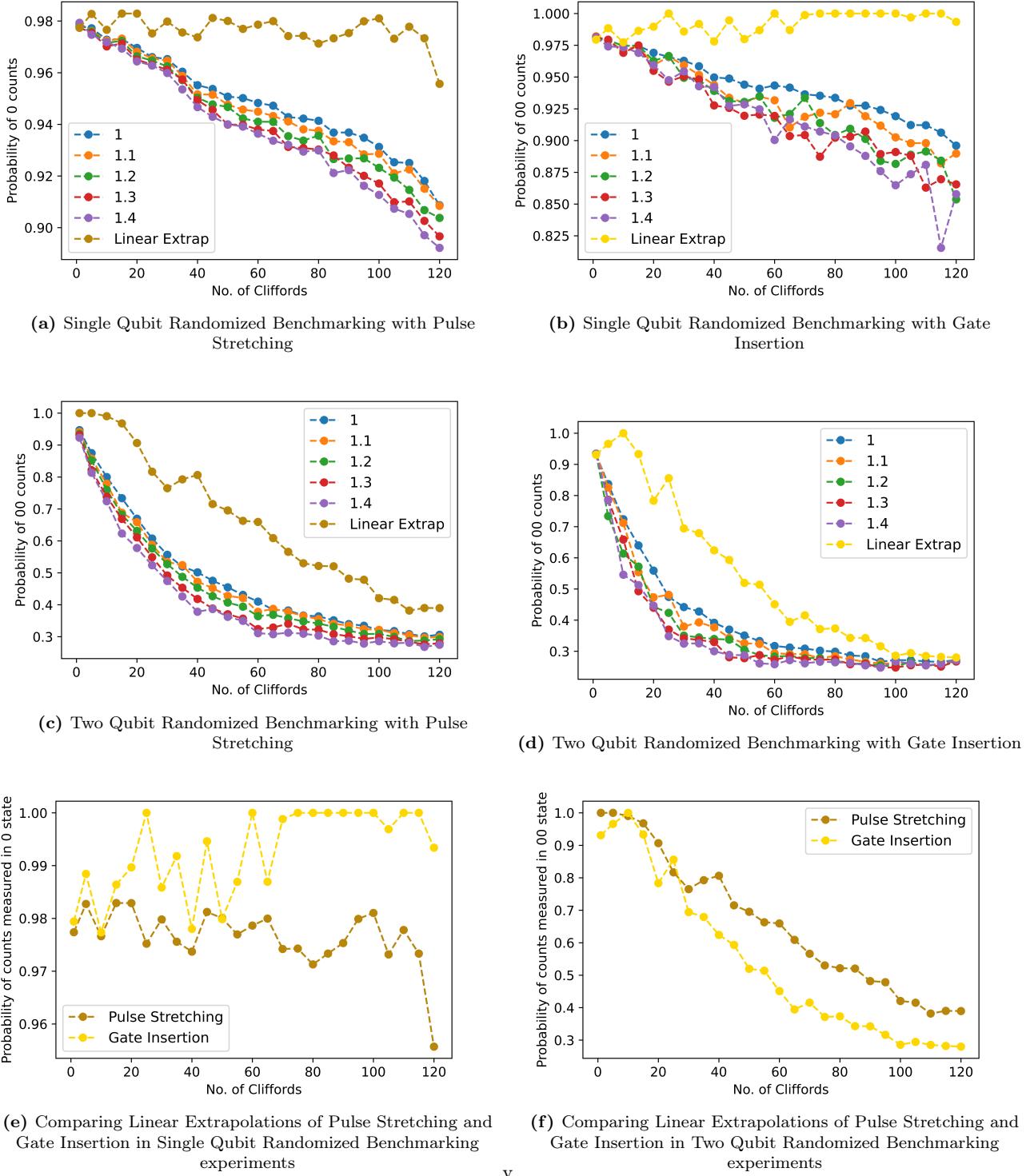


Figure 3: Comparing Pulse Stretching and Gate Insertion on Randomized Benchmarking (RB) experiments with single qubit (subfigures a,b) and two qubits (subfigures c,d). These experiments were conducted on IBMQ Casablanca device with 8192 samples for each circuit measurement.

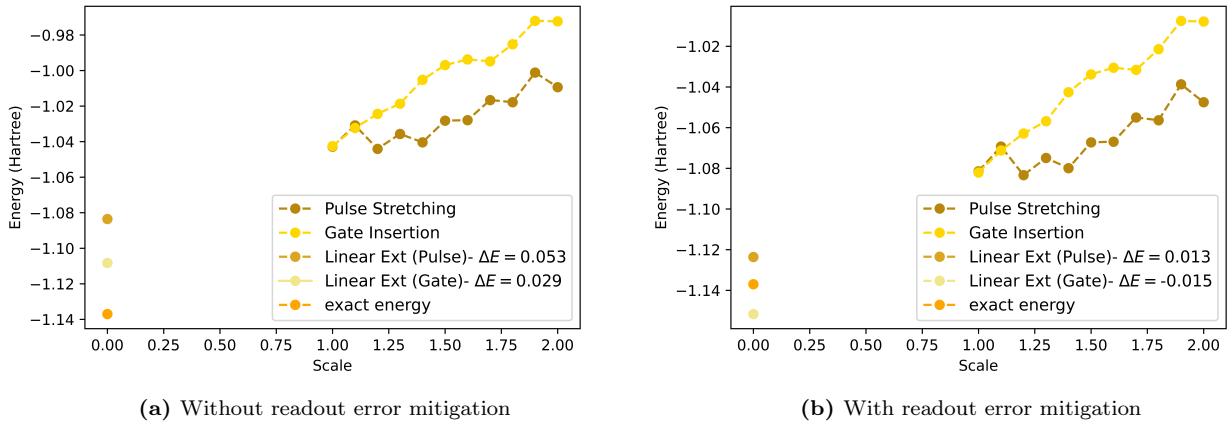


Figure 4: Preparing the UCCSD circuit for hydrogen with ideal parameters $\{0.0026304, -0.0025734, -0.11250428\}$ (optimized on noise model), we plot measured expectation value for different noise amplification factors. Then, employing linear extrapolation, we compare the resultant zero noise energies with the exact energy (ΔE). This procedure was repeated $M = 40$ times for both noise amplification techniques. Further, each circuit measurement on IBMQ Casablanca device consisted of 8192 samples.

Converting these counts to probabilities, we now have the entries for the second column of matrix R . Repeating this procedure for all the basis states and computing $R^{-1}C_{noisy}$, we will obtain the measurement of readout error mitigated outcomes.

VI. RESULTS AND DISCUSSION

Results from randomized benchmarking experiments suggest that gate insertion suffers more from sampling errors than pulse stretching. The problem lies in sampling the number of identities from the poisson distribution. As discussed in the previous section, this process requires circuits to have a large number of gates. However, larger circuit depth invalidates our assumption that noise parameter λ is weak and constant in time. Of course, we could still use higher order polynomial extrapolations if differences between $c = 1$ and $c > 1$ show nonlinear trends as in figures ??, ?. However, this led to overfitting and unstable extrapolation results.

If we adopt the procedure of repeatedly applying gate insertion for a given circuit to curb sampling errors (as discussed in the previous section), then gate insertion amplifies more errors than pulse stretching. This trend is evident in the steeper yellow curves in figure ???. In adding more redundant gates to the circuit, we increase incoherent errors by increasing the time taken by the quantum computer to run the circuit. In addition, we increase coherent errors (imperfections in qubit rotation) since the circuit now contains more noisy gates. However, in pulse stretching, the only amplified errors

are incoherent. It is unclear whether rescaling the time needed to implement each gate would increase the rotation errors.

Further, neither technique reasonably affects readout errors. Thus using the readout error mitigation technique in conjunction with zero noise extrapolation clearly improved the zero noise outcomes in figure ???. However, these extrapolated results still differ from exact energy by more than an order of magnitude larger than the threshold of 10^{-3} Hartree to be chemically accurate.

VII. FUTURE WORK

VIII. APPENDIX

A. Exponentiation of Operators

Suppose we want to act on a two qubit system, initially in state $|\psi\rangle = |01\rangle$ with the transformation

$$|\psi'\rangle = e^{-i(Z_1 \otimes Z_2)t} |\psi\rangle$$

From section 4.2 in [2], we can express the exponential as

$$\begin{aligned} |\psi'\rangle &= \cos t |01\rangle - i \sin t Z_1 Z_2 |01\rangle \\ |\psi'\rangle &= \cos t |01\rangle - i \sin t (Z|0\rangle \otimes Z|1\rangle) \\ |\psi'\rangle &= \cos t |01\rangle - i \sin t ((1)|0\rangle \otimes (-1)|1\rangle) \\ |\psi'\rangle &= (\cos t + i \sin t) |01\rangle \\ |\psi'\rangle &= e^{it} |01\rangle \end{aligned}$$

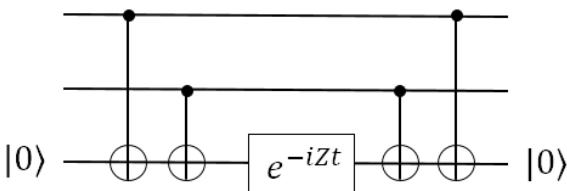


Figure 5: 2 qubit system

Here, the state $|01\rangle$ incurs a phase factor of e^{it} because the *parity* of the state was odd ($0+1=1$). If the parity were even, then the state would incur a phase factor of e^{-it} . More generally, for a system of n qubits,

$$e^{-i(Z_1 \otimes \dots \otimes Z_n)t} |\psi\rangle = \begin{cases} e^{it} |\psi\rangle & \text{even parity} \\ e^{-it} |\psi\rangle & \text{odd parity} \end{cases} \quad (13)$$

Computing the parity of qubits can be made easier if we introduce an ancilla qubit and act on it with successive CNOT operations as shown in figure ?? [2]. Then we apply e^{-iZt} on the ancilla qubit to achieve the same effect as equation ?? . The latter set of CNOT gates then uncompute the parity and effectively remove ancilla qubit from the circuit [2].

ACKNOWLEDGEMENTS

REFERENCES

- [1] Szabó Attila and Neil S. Ostlund. *Modern quantum chemistry*. Dover, 1989.
- [2] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th. USA: Cambridge University Press, 2011. ISBN: 1107002176.
- [3] Jacob T. Seeley, Martin J. Richard, and Peter J. Love. “The Bravyi-Kitaev transformation for quantum computation of electronic structure”. In: *The Journal of Chemical Physics* 137.22 (Dec. 2012), p. 224109. ISSN: 1089-7690. DOI: 10.1063/1.4768229. URL: <http://dx.doi.org/10.1063/1.4768229>.

- [4] Jarrod R. McClean et al. “Exploiting Locality in Quantum Computation for Quantum Chemistry”. In: *The Journal of Physical Chemistry Letters* 5.24 (Dec. 2014), pp. 4368–4380. ISSN: 1948-7185. DOI: 10.1021/jz501649m. URL: <http://dx.doi.org/10.1021/jz501649m>.
- [5] Alberto Peruzzo et al. “A variational eigenvalue solver on a photonic quantum processor”. In: *Nature Communications* 5.1 (July 2014). ISSN: 2041-1723. DOI: 10.1038/ncomms5213. URL: <http://dx.doi.org/10.1038/ncomms5213>.
- [6] Abhinav Kandala et al. “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets”. In: *Nature* 549.7671 (Sept. 2017), pp. 242–246. ISSN: 1476-4687. DOI: 10.1038/nature23879. URL: <http://dx.doi.org/10.1038/nature23879>.
- [7] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error Mitigation for Short-Depth Quantum Circuits”. In: *Physical Review Letters* 119.18 (Nov. 2017). ISSN: 1079-7114. DOI: 10.1103/physrevlett.119.180509. URL: <http://dx.doi.org/10.1103/PhysRevLett.119.180509>.
- [8] Jarrod R. McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9.1 (Nov. 2018). ISSN: 2041-1723. DOI: 10.1038/s41467-018-07090-4. URL: <http://dx.doi.org/10.1038/s41467-018-07090-4>.
- [9] Jonathan Romero et al. *Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz*. 2018. arXiv: 1701.02691 [quant-ph].
- [10] Abhinav Kandala et al. “Error mitigation extends the computational reach of a noisy quantum processor”. In: *Nature* 567.7749 (Mar. 2019), pp. 491–495. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1040-7. URL: <http://dx.doi.org/10.1038/s41586-019-1040-7>.
- [11] Andre He et al. “Zero-noise extrapolation for quantum-gate error mitigation with identity insertions”. In: *Physical Review A* 102.1 (July 2020). ISSN: 2469-9934. DOI: 10.1103/physreva.102.012426. URL: <http://dx.doi.org/10.1103/PhysRevA.102.012426>.
- [12] Rebecca Hicks, Christian W. Bauer, and Benjamin Nachman. *Readout Rebalancing for Near Term Quantum Computers*. 2020. arXiv: 2010.07496 [quant-ph].
- [13] Wim Lavrijsen et al. *Classical Optimizers for Noisy Intermediate-Scale Quantum Devices*. 2020. arXiv: 2004.03004 [quant-ph].
- [14] Sam McArdle et al. “Quantum computational chemistry”. In: *Reviews of Modern Physics* 92.1 (Mar. 2020). ISSN: 1539-0756. DOI: 10.1103/revmodphys.92.015003. URL: <http://dx.doi.org/10.1103/RevModPhys.92.015003>.