EECS 598 Project Report

Optimizing Shots allocation in Variational Quantum Algorithms

Mohammad Aamir Sohail
mdaamir@umich.edu

Ph.D., Electrical and Computer Engineering
University of Michigan, Ann Arbor, US

1 Introduction

Variational Quantum Algorithms (VQAs) represent a class of hybrid quantum-classical algorithms aimed at solving optimization problems by using *parameterized quantum circuits* (PQCs) and leveraging *classical optimization scheme* to find optimal PQC parameters. VQAs include popular algorithms like the variational quantum eigensolver (VQE) [TCC+22] and the quantum approximate optimization algorithm (QAOA) [FGG14]. These algorithms utilize variational principles to iteratively update parameters in quantum circuits, aiming to minimize the loss function of the problem, such as finding the ground state energy of a molecule (see Figure 1).

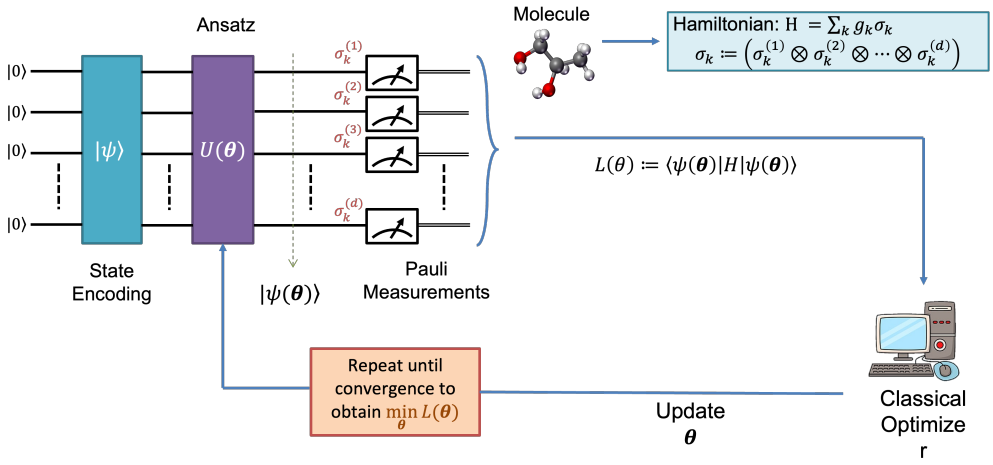


Figure 1: Variational Quantum Algorithms (VQA)

While VQAs offer significant potential for various applications, such as quantum chemistry and combinatorial optimizations, they also face several technical challenges. One primary challenge is the evaluation of the loss function itself, particularly for complex problems in quantum chemistry and materials science. The measurement process introduces noise that leads to errors in the loss function to be minimized. In VQAs, the accuracy of the loss function evaluation is crucial for the stability and speed of the optimization process. [KACC20] have shown that inaccurate evaluation of the loss function can lead to suboptimal solutions or slow convergence of the optimization algorithm. To reduce the error in the estimation of the loss function, a large number of measurements (shots) are needed, which can be costly, especially when the number of terms in the Hamiltonian is large. More formally, we need $O(1/\epsilon^2)$ number of shots to estimate the loss function with ϵ

error. Additionally, some quantum cloud platforms, such as Amazon's Braket service, charge by the number of shots, along with the number of different circuit runs. To reduce the overall run time and loss, we need an optimizer that uses a few VQAs iterations and a minimal number of shots.

Numerous techniques have emerged to boost the efficiency of VQE algorithms and reduce the required number of measurements. Some approaches involve restructuring Hamiltonian terms to streamline measurements [CYI22, CLI23, YGI23], and leveraging classical shadows to enhance both measurements [HKP20] and energy predictions [WSHY23]. To decrease shot counts, other methods concentrate on optimizing optimizers [LSDS24], implementing tiered shot allocations during optimization [PG23], adapting estimators based on historical data [SJM$^+$23], and assigning variable shot counts to different Hamiltonian terms [ACSC20, GLD$^+$21]. Notably, variance-minimization-based shot assignments have proven effective in refining the accuracy and efficiency of quantum measurements [YGI23, WHT15, ACSC20]. These techniques typically leverage prior knowledge of Hamiltonian terms, often obtained from classical simulations, to allocate measurement shots in a way that minimizes overall energy estimation variance.

2 Objective

The goal of the project is to implement strategies, namely, weighted random sampling (WRS) [ACSC20], variance0-minimization shots allocation (VMSA), and variance preserved shots reduction (VPSR) [ZLYL23] for optimizing measurement (shots) in variational quantum algorithms. These algorithms mainly discuss the shots minimization for the evaluation of the loss function. The WRS method [ACSC20] randomly selects the number of shots for the evaluation of the expectation of each Pauli string of the Hamiltonian. On the other side, the authors in [ZLYL23] introduce the variation-minimization (VM) method to select the number of shots to estimate each term of the Hamiltonian required in the evaluation of the loss function. Additionally, the project aims to develop a novel optimizer by simultaneously optimizing shot allocation for the evaluation of loss function [ZLYL23] and the estimation of each component of the gradient at each iteration of SGD [GLD$^+$21]. Further details are discussed in Sections 4 and 7.

3 Problem Setup

Notations: Let $[\Theta] := \{1, 2, \cdots, \Theta\}$.We use $\mathtt{nS}_i$ to denote the number of shots for computing the i^{th} component of the gradient and $s_k^{(i)}$ to denote the number of shots for estimating the k^{th} Pauli string of the Hamiltonian.

In this section, we will formalize the optimization problem in Variational Quantum Algorithms (VQAs). VQAs involve a Parameterized Quantum Circuit (PQC), also known as an *ansatz*, and measuring a problem-specific observable. These algorithms include a classical optimization step that determines the set of ansatz parameters that minimize the expected value of the problem-specific observable.

Consider a d qubit Hamiltonian, $H = \sum_{k=1}^{n} c_k P_k$, where $P_k := \bigotimes_{j=1}^{d} \boldsymbol{\sigma}_j^{(k)}$ are Pauli strings, $\boldsymbol{\sigma}_j^{(k)} \in \{X, Y, Z, I\}$ are Pauli operators acting on j^{th} qubit, n is the number of Pauli strings in the Hamiltonian, and $\{c_k\}$ represents the weight of Pauli strings. Consider an ansatz $U(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^d$ are ansatz parameters and c is the number of parameters. The ansatz is applied to the input quantum state $|\psi_0\rangle$ with the objective of minimizing the incurred loss (energy) function, expressed

as

$$\mathcal{L}(\boldsymbol{\theta}) := \langle\psi_0|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|\psi_0\rangle = \sum_{k=1}^{n} c_k \langle\psi_0|U^\dagger(\boldsymbol{\theta})P_kU(\boldsymbol{\theta})|\psi_0\rangle\,. \tag{1}$$

Here, note that the loss function is the expectation of an observable. The optimal value of the parameters $\boldsymbol{\theta}$, denoted as $\boldsymbol{\theta}^*$ is found by solving the optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}).$$

Here, quantum computers evaluate the energy function, and the optimization of parameters will be performed using a classical optimizer. A standard classical optimization approach is gradient descent, characterized by the following update rule: $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta\nabla\mathcal{L}(\boldsymbol{\theta})$. However, due to the quantum mechanical properties, such as state collapse and the stochastic nature of quantum measurements, we only have access to the estimates of the loss function. Therefore, under this setting, we use stochastic gradient descent.

Unbiased Estimators of the Gradient: To construct an unbiased estimator of gradient, we use the *parameter-shift rule* [MNKF18]. It states that the gradient of a function with respect to a parameter can be efficiently computed by the difference in function evaluations when perturbing the parameter values in opposite directions. Mathematically, this can be represented as

$$\partial_i\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2}\left[\mathcal{L}(\boldsymbol{\theta} + \frac{\pi}{2}e_i) - \mathcal{L}(\boldsymbol{\theta} - \frac{\pi}{2}e_i)\right], \tag{2}$$

where $\boldsymbol{\theta}$ denotes the parameter of interest and e_i is an unit vector along the i^{th} direction. Although we cannot evaluate the exact value of expectations required in (2), we can make use of a random variable $\mathbf{g}(\boldsymbol{\theta})$ that is an unbiased estimate of the gradient, i.e., $\mathbb{E}[\mathbf{g}(\boldsymbol{\theta})] = \nabla\mathcal{L}(\boldsymbol{\theta})$. We construct the i^{th} component of the estimate of the gradient as

$$\mathbf{g}_i(\boldsymbol{\theta}) := \frac{1}{2\cdot \mathtt{nS}_i}\sum_{j=1}^{\mathtt{nS}_i}(X_{j+}^{(i)} - X_{j-}^{(i)}),$$

where $X_{j\pm}^{(i)}$ is a single-shot unbiased estimator of the expectation $\mathcal{L}(\boldsymbol{\theta} \pm \frac{\pi}{2}e_i)$. In what follows, we compare various strategies for constructing $X_{j\pm}^{(i)}$, in terms of their variances, given a finite number of shots $\mathtt{nS}_i$ for the estimation of the i^{th} component of the gradient.

3.1 Variance of Estimators

In this subsection, we derive an expression for the variance of an unbiased estimator of the loss function. Let $\langle P_k\rangle$ denotes $\langle\psi(\boldsymbol{\theta})|P_k|\psi(\boldsymbol{\theta})\rangle$. Suppose $\hat{X}$ and $\hat{P}_k$ denote the estimators for $\mathcal{L}(\boldsymbol{\theta})$ and $\langle P_k\rangle$, respectively, where $|\psi(\boldsymbol{\theta})\rangle := U(\boldsymbol{\theta})\,|\psi_0\rangle$ is the parameterized quantum state. Thus, we have

$$\hat{X} = \sum_{k=1}^{n} c_k\hat{P}_k \text{ with } \hat{P}_k = \frac{1}{\mathbb{E}[s_k]}\sum_{j=1}^{s_k} Y_j^{(k)}. \tag{3}$$

Here, each $Y_j^{(k)}$ is an independent random variable associated measurement of Pauli string P_k. We now state the following proposition, which can be easily proved using the definition of expectation and variance of a random variable.

Proposition 1. *$\hat{X}$ is an unbiased estimator of $\mathcal{L}(\boldsymbol{\theta})$.*

Proposition 2. *The variance of $\hat{L}$ is*

$$\mathrm{Var}(\hat{X}) = \sum_{k=1}^{M} \frac{c_i^2}{\mathbb{E}[s_k]}\sigma_k^2 + \sum_{k_1,k_2} \frac{c_{k_1}c_{k_2}\langle P_{k_1}\rangle\langle P_{k_2}\rangle}{\mathbb{E}[s_{k_1}]\mathbb{E}[s_{k_2}]}\mathrm{Cov}(s_{k_1}, s_{k_2}), \tag{4}$$

where $\sigma_k^2 = 1 - \langle P_k\rangle^2$ is the variance of Pauli string P_k with respect to $|\psi(\boldsymbol{\theta})\rangle$.

3.1.1 Grouping Commuting Pauli Terms into Cliques for Simultaneous Measurements

A useful strategy to minimize the number of shots required for measuring individual Pauli terms of the Hamiltonian is to group them into cliques. Each clique contains terms that commute with each other, allowing simultaneous measurement of all terms within the clique. This way, the same measurement result can be used to calculate the energy contribution of each term within the clique, reducing the overall number of shots needed for energy estimation.

A Hamiltonian clique refers to a subset of terms in the Hamiltonian that can be measured simultaneously. In the case of H_2 (illustrated in Section 5), for instance, the Hamiltonian consists of six terms that can be consolidated into three Hamiltonian cliques because the first three non-identity terms commute with each other, enabling simultaneous measurements on both qubits. However, identifying these cliques is a challenging problem, as determining the partition with the smallest number of cliques is known to be NP-hard.[WSHY23].

We now establish a connection between notations and Hamiltonian cliques. Suppose a Hamiltonian H has n Pauli terms, which can be grouped into M Hamiltonian cliques. Let $X_{\pm}^{(i)}$ denotes the estimator for $\mathcal{L}(\boldsymbol{\theta} \pm \frac{\pi}{2}e_i)$. Thus, $X_{\pm}^{(i)}$ can be expressed as

$$X_{\pm}^{(i)} = \sum_{k=1}^{M} c_k \left(\frac{1}{\mathbb{E}[s_k^{(i)}]} \sum_{j=1}^{s_k^{(i)}} Y_j^{(k)} \right) \mathbb{1}_{\left\{\sum_{k=1}^{M} s_k^{(i)} = \mathtt{nS}_i\right\}},$$

where $\mathtt{nS}_i$ is the number of shots allocated for evaluating the i^{th} component of the gradient and $s_k^{(i)}$ is the number of shots allocated to estimate k^{th} Hamiltonian clique.

4 Methods

In this section, we provide the background and complete description of the following algorithms: (*i*) Uniform Deterministic Sampling (UDS), (*ii*) Weighted Random Sampling (WRS) [ACSC20], (*iii*) Variance Minimization Shots Allocation (VMSA) [ZLYL23], (*iv*) Variance Preserved Shots Reduction (VPSR) [ZLYL23]. For readability, we present the results of all the above-mentioned algorithms in unified notations.

To begin understanding the theory behind these algorithms, we will first provide a brief overview in a unified manner. All of the algorithms mentioned earlier assume that $\mathtt{nS}_i = \mathtt{nS}$ for all $i \in [c]$, meaning that $\mathtt{nS}$ number of shots are assigned to evaluate each element of the gradient vector (Fig. 7). The optimization process then involves determining how to distribute the $\mathtt{nS}$ among the M Hamiltonian cliques. UDS distributes the $\mathtt{nS}$ uniformly among the M Hamiltonian cliques, WRS distributes $\mathtt{nS}$ among M Hamiltonian cliques using a multinomial distribution, and VMSA allocates shots to each Hamiltonian based on the variance contribution of the respective cliques. VPSR is

essentially similar to VMSA, except that it allows for a small error of 1 kcal/mole in the ground state energy. This, in turn, results in a reduction in the number of shots required for evaluating each component of the gradient. The new number of shots required would be $\mathtt{nS}_{\text{VPSR}} = \eta \cdot \mathtt{nS}$ for some $\eta \in (0, 1)$.

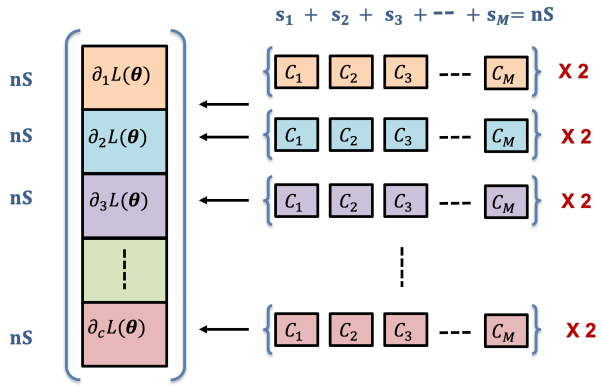


Figure 2: Shots allocation in VQA. Here, a factor of 2 denotes two function evaluations are needed to compute the gradient using the parameter-shift rule.

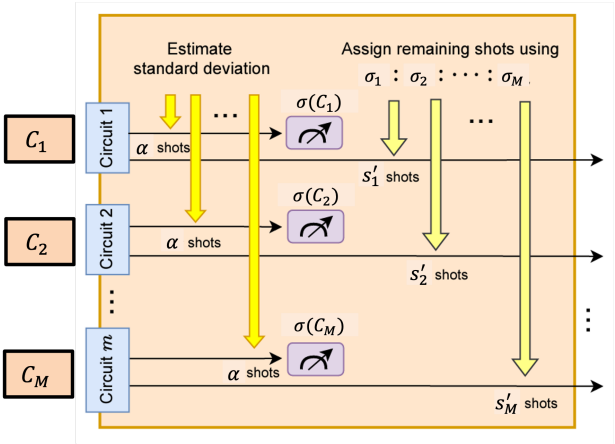


Figure 3: VMSA and VPSR algorithms

4.1 Uniform Deterministic Sampling

Given a budget $\mathtt{nS}$ to estimate $\mathcal{L}(\boldsymbol{\theta} \pm \frac{\pi}{2} e_i)$, the fundamental approach is to simply divide the number of shots equally among the M Hamiltonian clique. Thus, it will lead us to $s_k^{(i)} = \mathtt{nS}/M$. In general, when $\mathtt{nS} \neq \alpha M$, for some $\alpha \in \mathbb{N}$, the shots for each Pauli-strings can be given as $s_k^{(i)} = \lfloor \mathtt{nS}/M \rfloor$. Here, note that $\mathbb{E}[s_k^{(i)}] = \mathtt{nS}/M$, for all $k \in [M]$ and $\text{Cov}(s_{k_1}^{(i)}, s_{k_2}^{(i)}) = 0$, for all $k_1, k_2 \in [M]$. Moreover, from (4), we have

$$\text{Var}(\langle H \rangle) = \frac{M}{\mathtt{nS}} \sum_{i=1}^{M} |c_k|^2 \sigma_k^2.$$

4.2 Weighted Random Operator Sampling

The variance of $\mathcal{L}(\boldsymbol{\theta} \pm \frac{\pi}{2} e_i)$ using the above deterministic approach is sub-optimal because it is desirable to invest more shots in estimating the highly weighted operators, i.e., the Pauli strings corresponding to the largest $|c_k|$. To address this issue, the WRS method is introduced. This method introduces randomness to distribute the shots more effectively. It does so by randomly selecting $s_k^{(i)}$ from a multinomial distribution where the probability p_k of measuring k^{th} Hamiltonian clique is proportional to the sum of the absolute value of the weight of Pauli terms in that clique, i.e., $p_k \propto \sum_{k \in \text{H.Clique}} |c_k|$. Thus, mathematically, it can be expressed as:

$$s_k^{(i)} \sim \mathtt{Multi}\left(\mathtt{nS}, p_k := \frac{\sum_{k \in \text{H.Clique}} |c_k|}{\sum_{i=1}^{M} |c_i|}\right).$$

Here, note that, from the definition of multinomial distribution $\mathbb{E}[s_k^{(i)}] = p_k \cdot \mathtt{nS}$. Moreover, using (4) and variance and covariance of the multinomial distribution, the variance of the loss function

estimator can be expressed as:

$$\mathrm{Var}(\langle H\rangle) = \frac{\left(\sum_{i=1}^{M}|c_i|\right)}{\mathtt{nS}}\sum_{k=1}^{M}|c_k|\sigma_k^2 + \frac{\left(\sum_{i=1}^{M}|c_i|\right)}{\mathtt{nS}}\sum_{k=1}^{M}|c_k|\langle P_k\rangle^2 - \frac{\langle H\rangle^2}{\mathtt{nS}_i}.$$

Here, σ_k^2 denotes the variance k^{th} Hamiltonian clique, and $\langle P_k\rangle^2$ denotes the square of the expectation of all Pauli strings present in the k^{th} Hamiltonian clique. When $\mathtt{nS}_i$ is large, this method is similar to the above deterministic method. However, in general, this method has been shown to outperform the deterministic method for VQAs [ACSC20].

4.3 Variance Minimization Shots Allocation

This method aims to reduce the variance of the estimator (4), while maintaining a constant total number of shots. This is achieved by minimizing the following expression:

$$\min\left\{\sum_{k=1}^{M}\frac{1}{s_k^{(i)}}\sigma_k^2 : \sum_{k=1}^{M}s_k^{(i)} = \mathtt{nS}\right\} \quad \forall\ i \in [c]. \tag{5}$$

Note that we are assuming that $\mathtt{nS}_i = \mathtt{nS}\ \forall\ i \in [c]$. For the sake of brevity, we will drop the superscript from $s_k^{(i)}$. According to the Cauchy-Schwarz inequality, the objective function in (5) is minimized if and only if the shot allocation ratios are identical to the ratios of the standard deviations associated with the measurements of different Hamiltonian cliques, i.e.,

$$s_1 : s_2 : \cdots : s_M = \sigma_1 : \sigma_2 : \cdots : \sigma_M. \tag{6}$$

In order to calculate the ratio given by equation (6), we need to determine the standard deviation σ_k for each Hamiltonian clique. This can be done by allocating α shots from the total budget of $\mathtt{nS}$. The remaining $\mathtt{nS} - \alpha M$ shots are then distributed according to equation (6). Therefore, the number of shots assigned to the k^{th} clique is given by:

$$s_k = \alpha + s_k^{'} = \alpha + \frac{\sigma_k}{\sum_{j=1}^{M}\sigma_j}(\mathtt{nS} - \alpha M),$$

where the first term (α) accounts for the shots used to estimate the standard deviation, and the second term $s_k^{'}$ represents the additional shots assigned based on the ratio of standard deviations (see Fig. 3).

4.4 Variance Preserved Shots Reduction

Although minimizing variance is an effective way to optimize shot allocation for faster VQE convergence, it does not reduce the number of measurements when the variance falls below a threshold limit. To address this issue, the VPSR method has been proposed, which aims to find the optimal shot usage efficiency while ensuring that the variance of the estimator of the loss function remains within a threshold range. The objective is to solve an optimization problem that involves determining the optimal efficiency in shot usage given a threshold value δ required for VQE convergence.

$$\min\left\{\sum_{k=1}^{M}s_k^{(i)} : \sum_{k=1}^{M}\frac{1}{s_k^{(i)}}\sigma_k^2 \leq \delta\right\} \quad \forall\ i \in [c]. \tag{7}$$

The optimization problem (7) can be solved by writing the solutions as

$$s_k = \eta \cdot \frac{\sigma_k}{\sum_{j=1}^{M} \sigma_j} \mathtt{nS}, \tag{8}$$

which is fully derived in [ZLYL23, Appendix B]. Here,

$$\eta = \frac{1}{M} \frac{\left(\sum_{k=1}^{M} \sigma_k\right)^2}{\sum_{k=1}^{M} \sigma_k^2} \leq 1.$$

Similar to VMSA, we need to obtain standard deviation σ_k and variance σ_k^2 to obtain the number of shots s_k. To obtain these values, we follow the same approach as in VMSA, wherein we allocate α shots from the total budget of $\mathtt{nS}$. The remaining $(\mathtt{nS} - \alpha M)$ shots are then distributed according to equation (8). Therefore, in VPSR, the number of shots assigned to the i^{th} Hamiltonian clique is given by:

$$s_k = \alpha + s_k^{'} = \alpha + \eta \frac{\sigma_k}{\sum_{j=1}^{M} \sigma_j} (\mathtt{nS} - \alpha M).$$

We must note that $\eta \leq 1$, which ensures that $\sum_{k=1}^{M} s_k = \eta \cdot \mathtt{nS} \leq \mathtt{nS}$.

5 Experiments and Results

We describe our plan of step-by-step experiments in the subsections below.

5.1 Experiment setup

All experiments were implemented with Pennylane. We utilize a stochastic gradient descent (SGD) based optimizer with a learning rate of 0.01. The gradient of the loss function is computed using the parameter shift rule.

6 H_2 molecules with two qubits

We use the truncated version of the Unitary Coupled-Cluster Single and Double (UCCSD) ansatz for quantum simulation of H_2 molecule due to its representation as unitary transformations, enabling straightforward circuit implementation [BGS+18, HMR+18]. For our initial experiment, we will be using the UCCSD wavefunction ansatz and the minimal STO-6G basis to simulate an H_2 molecule with a bond length of 1.75A°. This wavefunction can be effectively represented by two qubits. The unitary transformation $U(\theta)$ contains only one parameter θ and can be written as $U(\theta) = \exp(-i\theta X_0 X_1)$, where $|\psi_0\rangle = |01\rangle$. This represents the Hartree–Fock reference of the H_2 system. Each qubit represents a molecular spin-orbital, where $|0\rangle$ and $|1\rangle$ denote unoccupied and occupied orbitals, respectively. The quantum circuit for the UCCSD state preparation is shown in Figure 4. The Bravyi-Kitaev transformed Hamiltonian [OBK+16] with spin symmetries using two qubits for an H_2 molecule is given as:

$$H = c_0 I + c_1 Z_0 \otimes I_1 + c_2 I_0 \otimes Z_1 + c_3 Z_0 \otimes Z_1 + c_4 Y_0 \otimes Y_1 + c_5 X_0 \otimes X_1,$$

The values $\{c_k\}$ are obtained from a Hartree-Fock calculation that is based on the bond distances and the type of atomic basis function. The coefficients $\{c_k\}$ are: c_0 = -0.5597, c_1 = 0.1615, c_2 = -0.0166, c_3 = 0.4148, c_4 = 0.1226, and c_5 = 0.1226. We now group the Pauli terms of H_2 into three

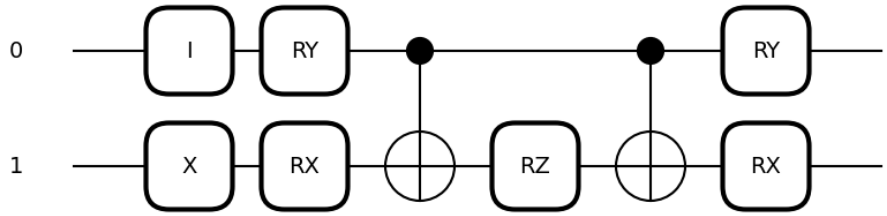


Figure 4: The quantum circuit implementation of the UCCSD wavefunction ansatz $\exp(-i\theta X_0 X_1)$ for an H_2 molecule with a minimal STO-6G basis.

cliques. As the first three non-identity terms (c_{1-3}) commute with each other, we group them into a Hamiltonian clique.

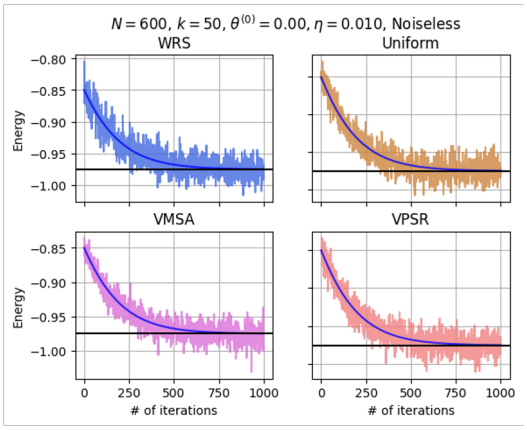
$$H_{1-3}(\theta) = c_1 \langle\psi(\theta)| Z_0 \otimes I_1 |\psi(\theta)\rangle + c_2 \langle\psi(\theta)| I_0 \otimes Z_1 |\psi(\theta)\rangle + c_3 \langle\psi(\theta)| Z_1 \otimes Z_0 |\psi(\theta)\rangle$$

where the same set of measurement results $|\langle 00|\psi\rangle|^2, |\langle 01|\psi\rangle|^2, |\langle 10|\psi\rangle|^2$, and $|\langle 11|\psi\rangle|^2$ can be used as follows:

$$\begin{aligned}
c_1\langle P_1\rangle := c_1 \langle\psi(\theta)| Z_0 \otimes I_1 |\psi(\theta)\rangle &= c_1 \langle\psi|(|0\rangle\langle 0| - |1\rangle\langle 1|) \otimes (|0\rangle\langle 0| + |1\rangle\langle 1|)|\psi\rangle \\
&= c_1 \left(|\langle 00|\psi\rangle|^2 + |\langle 01|\psi\rangle|^2 - |\langle 10|\psi\rangle|^2 - |\langle 11|\psi\rangle|^2\right), \\
c_2\langle P_2\rangle := c_2 \langle\psi(\theta)| I_0 \otimes Z_1 |\psi(\theta)\rangle &= c_2 \langle\psi|(|0\rangle\langle 0| + |1\rangle\langle 1|) \otimes (|0\rangle\langle 0| - |1\rangle\langle 1|)|\psi\rangle \\
&= c_2 \left(|\langle 00|\psi\rangle|^2 - |\langle 01|\psi\rangle|^2 + |\langle 10|\psi\rangle|^2 - |\langle 11|\psi\rangle|^2\right), \\
c_3\langle P_3\rangle := c_3 \langle\psi(\theta)| Z_0 \otimes Z_1 |\psi(\theta)\rangle &= c_2 \langle\psi|(|0\rangle\langle 0| - |1\rangle\langle 1|) \otimes (|0\rangle\langle 0| - |1\rangle\langle 1|)|\psi\rangle \\
&= c_3 \left(|\langle 00|\psi\rangle|^2 - |\langle 01|\psi\rangle|^2 - |\langle 10|\psi\rangle|^2 + |\langle 11|\psi\rangle|^2\right).
\end{aligned}$$

The other two cliques H_4 and H_5 are $c_4 \langle\psi(\theta)| Y_1 \otimes Y_0 |\psi(\theta)\rangle$ and $c_5 \langle\psi(\theta)| X_1 \otimes X_0 |\psi(\theta)\rangle$, respectively. While H_4 and H_5 require separate measurements, their measurement process can be streamlined through the use of appropriate transformation unitaries, which rotate them into the computational basis. For H_4, as $\left(HS^\dagger\right)^\dagger Z \left(HS^\dagger\right) = Y$, we can apply $HS^\dagger$ to each qubit before $|\psi(\theta)\rangle$ and then measure it in the computational basis. Similarly, for H_5, as $HZH = X$, we can apply H to each qubit before measuring it. Here, H and S denote the Hadamard gate and Phase gate, respectively.

To illustrate the efficacy of the VPSR approach, we conducted several experiments with a total shots budget of $\texttt{nS} = 600$ and $\texttt{nS} = 1200$. In these experiments, we estimate the loss function $\mathcal{L}(\theta)$ for $\texttt{nS} = 600$ and 1200 starting from three initial states: $\theta_0 = 0$, representing the Hartree–Fock reference, $\theta_0 = 2$, and $\theta_0 = -2.5$. Figures 5, 6, and 7 showcase a comparison of the VPSR strategy with the VMSA approach and non-variance-based methods, including WRS and UDS, using two different shots budgets: $\texttt{nS} = 600$ (first column) and $\texttt{nS} = 1200$ (second column). Benchmark tests reveal that VPSR efficiently reduces the number of shots required while achieving a convergence profile comparable to other approaches. For shot budgets of $\texttt{nS} = 600$ and $\texttt{nS} = 1200$, the VPSR approach achieves a reduction in the number of measurements needed for convergence (to within 1 kcal/mol from the ground state) by approximately 10-25%, relative to the other three strategies, which is a significant reduction while having the same convergence rate.



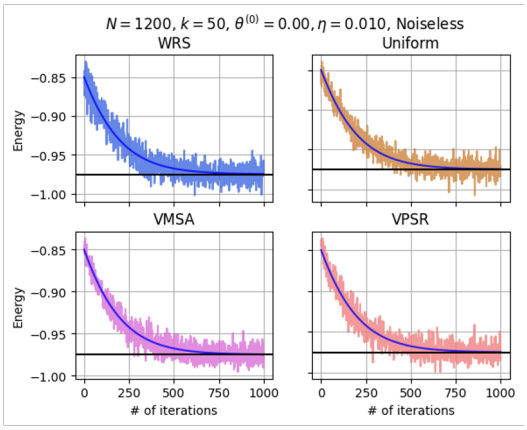
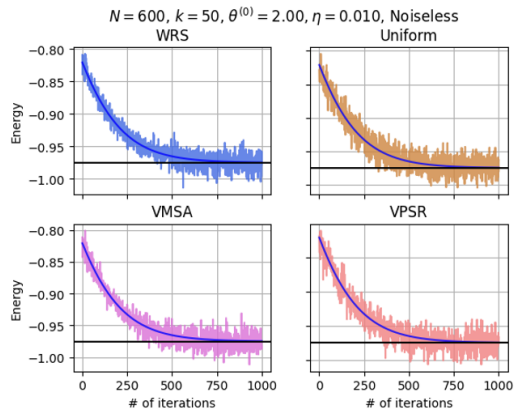


Figure 5: Comparison of different shot assignment strategies for the VQE optimization for H_2 molecule with initial parameter $\theta_0 = 0$. For $\mathtt{nS} = 600$ and $\mathtt{nS} = 1200$, the average number of VPSR shots $= 2 \times 522$ and 2×1017, respectively. Thus, % shots reduction in VPSR are 12.94% and 15.21%, respectively.



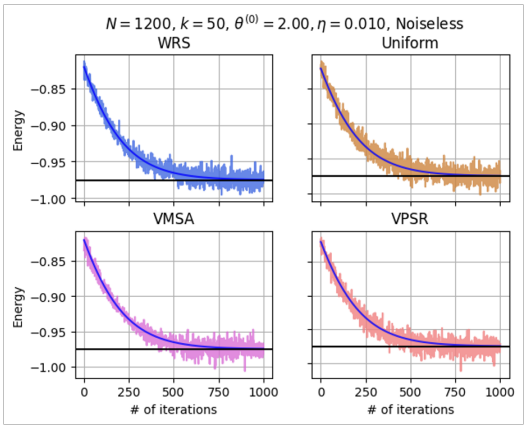


Figure 6: Comparison of different shot assignment strategies for the VQE optimization for H_2 molecule with initial parameter $\theta_0 = 2$. For $\mathtt{nS} = 600$ and $\mathtt{nS} = 1200$, the average number of VPSR shots $= 2 \times 540$ and 2×1063, respectively. Thus, % shots reduction in VPSR are 10% and 11.43%, respectively.

7 Future Work

7.1 Evaluating the efficacy of VPSR for LiH molecule

To assess the performance of the VPSR method further for larger molecules, we are working on simulating the VQE for LiH molecule using 4 qubits. The simplified expression of the Hamiltonian will consist of 27 Pauli strings, which can be grouped into 9 Hamiltonian cliques.

7.2 Developing a Novel Adaptive Optimizer

In [GLD+21], the authors introduced a globally coupled adaptive number of shots (gCANS), which dynamically selects the number of shots required to estimate each component of the gradient at every iteration of stochastic gradient descent (SGD). In the gCANS method, authors performed an analysis to derive optimal shot allocation for each component of the gradient by maximizing

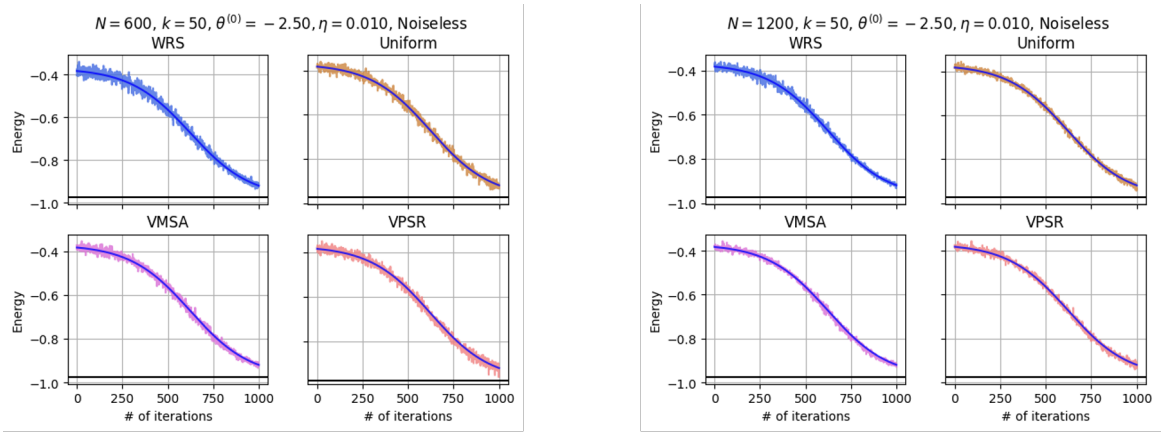


Figure 7: Comparison of different shot assignment strategies for the VQE optimization for H_2 molecule with initial parameter $\theta_0 = -2.5$. For `nS` $= 600$ and `nS` $= 1200$, the average number of VPSR shots $= 2 \times 466$ and 2×891, respectively. Thus, % shots reduction in VPSR are 22.26% and 25.77%, respectively.

the ratio of expected gain at each iteration and a total number of shots for estimating gradient vector, i.e., $\mathbb{E}[\mathcal{L}(\boldsymbol{\theta}_t) - \mathcal{L}(\boldsymbol{\theta}_{t+1})]/(\sum_{i=1}^{c} s_i)$, where s_i is the number of shots for estimating $\nabla_i \mathcal{L}(\boldsymbol{\theta})$. To estimate each component of the gradient, the authors used the parameter-shift rule [MNKF18], which requires the evaluation of the energy function at two shifted parameters. Here, for the evaluation of the energy function, authors used the WRS method [ACSC20], which distributes the shots allocated to estimate $\nabla_i \mathcal{L}(\boldsymbol{\theta})$, i.e., s_i, among the M terms of the Hamiltonian.

If we observe, the authors do not simultaneously optimize the shot allocation for each term of the Hamiltonian and each component of the gradient. Moreover, WRS does not leverage the commutativity of the Pauli strings, making it suboptimal. In contrast, in [ZLYL23], the authors introduce a variance-minimization strategy to allocate the shots for estimating the expectation of each Pauli string, considering commutativity among Pauli strings to reduce the number of shots for energy function evaluation. With this motivation and idea, we aim to integrate both strategies to develop an optimizer that optimizes shot allocation for each component of the gradient and each term in the Hamiltonian simultaneously.

Moreover, while gCANS maximized the expected gain achieved per iteration, classical optimization literature has explored optimizing the expected improvement in energy value relative to the optimal value [GKR22]. We plan to investigate the impact of considering the expected distance between the energy value at each iteration and the optimal value on shot allocation. Ultimately, our goal is to consolidate the findings from this project into a publication demonstrating the performance enhancements achieved by the novel optimizer.

References

[ACSC20] Andrew Arrasmith, Lukasz Cincio, Rolando D Somma, and Patrick J Coles. Operator sampling for shot-frugal optimization in variational algorithms. *arXiv preprint arXiv:2004.06252*, 2020.

[BGS+18] Panagiotis Kl Barkoutsos, Jerome F Gonthier, Igor Sokolov, Nikolaj Moll, Gian Salis, Andreas Fuhrer, Marc Ganzhorn, Daniel J Egger, Matthias Troyer, Antonio Mezzacapo,

et al. Quantum algorithms for electronic structure calculations: Particle-hole hamiltonian and optimized wave-function expansions. *Physical Review A*, 98(2):022322, 2018.

[CLI23] Seonghoon Choi, Ignacio Loaiza, and Artur F Izmaylov. Fluid fermionic fragments for optimizing quantum measurements of electronic hamiltonians in the variational quantum eigensolver. *Quantum*, 7:889, 2023.

[CYI22] Seonghoon Choi, Tzu-Ching Yen, and Artur F Izmaylov. Improving quantum measurements by introducing "ghost" pauli products. *Journal of Chemical Theory and Computation*, 18(12):7394–7402, 2022.

[FGG14] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

[GKR22] Zhan Gao, Alec Koppel, and Alejandro Ribeiro. Balancing rates and variance via adaptive batch-size for stochastic optimization problems. *IEEE Transactions on Signal Processing*, 70:3693–3708, 2022.

[GLD+21] Andi Gu, Angus Lowe, Pavel A Dub, Patrick J Coles, and Andrew Arrasmith. Adaptive shot allocation for fast convergence in variational quantum algorithms. *arXiv preprint arXiv:2108.10434*, 2021.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[HMR+18] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P Lanyon, Peter Love, Ryan Babbush, et al. Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X*, 8(3):031022, 2018.

[KACC20] Jonas M Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J Coles. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum*, 4:263, 2020.

[LSDS24] Di Luo, Jiayu Shen, Rumen Dangovski, and Marin Soljacic. Quack: Accelerating gradient-based quantum optimization with koopman operator learning. *Advances in Neural Information Processing Systems*, 36, 2024.

[MNKF18] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.

[OBK+16] Peter JJ O'Malley, Ryan Babbush, Ian D Kivlichan, Jonathan Romero, Jarrod R McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, Nan Ding, et al. Scalable quantum simulation of molecular energies. *Physical Review X*, 6(3):031007, 2016.

[PG23] Koustubh Phalak and Swaroop Ghosh. Shot optimization in quantum machine learning architectures to accelerate training. *IEEE Access*, 2023.

[SJM+23] Ariel Shlosberg, Andrew J Jena, Priyanka Mukhopadhyay, Jan F Haase, Felix Leditzky, and Luca Dellantonio. Adaptive estimation of quantum observables. *Quantum*, 7:906, 2023.

[TCC+22] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H Booth, et al. The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986:1–128, 2022.

[WHT15] Dave Wecker, Matthew B Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92(4):042303, 2015.

[WSHY23] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. Overlapped grouping measurement: A unified framework for measuring quantum states. *Quantum*, 7:896, 2023.

[YGI23] Tzu-Ching Yen, Aadithya Ganeshram, and Artur F Izmaylov. Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates. *npj Quantum Information*, 9(1):14, 2023.

[ZLYL23] Linghua Zhu, Senwei Liang, Chao Yang, and Xiaosong Li. Optimizing shot assignment in variational quantum eigensolver measurement. *arXiv preprint arXiv:2307.06504*, 2023.