Quantum Circuit Ansatz Structures for Ising Model & A Comparative Analysis of Classical and Quantum Optimization Methods

Duc-Truyen Le*
Department of Physics,
National Tsing Hua University,
Hsinchu, Taiwan 300044, R.O.C.

Vu-Linh Nguyen
Department of Theoretical Physics, University of Science,
Vietnam National University, Ho Chi Minh City 70000, Vietnam

Minh- Triet Ha
Rhodes College, Memphis, TN 38112, USA

Cong-Ha Nguyen
Département de Physique de l'École normale supérieure, ENS-PSL

Quoc-Hung Nguyen
Nano and Energy Center, VNU University of Science,
Vietnam National University, Hanoi, Vietnam

Van-Duy Nguyen†
Phenikaa Institute for Advanced Study,
Phenikaa University, Hanoi 12116, Vietnam
(Dated: December 19, 2024)

Abstract: In this study, we delved into several optimization methods, both classical and quantum, and analyzed the quantum advantage that each of these methods offered, and then we proposed a new combinatorial optimization scheme, deemed as QN-SPSA+PSR which combines calculating approximately Fubini-study metric (QN-SPSA) and the exact evaluation of gradient by Parameter-Shift Rule (PSR). The QN-SPSA+PSR method integrates the QN-SPSA computational efficiency with the precise gradient computation of the PSR, improving both stability and convergence speed while maintaining low computational consumption. Our results provide a new potential quantum supremacy in the VQE's optimization subroutine and enhance viable paths toward efficient quantum simulations on Noisy Intermediate-Scale Quantum Computing (NISQ) devices. Additionally, we also conducted a detailed study of quantum circuit ansatz structures in order to find the one that would work best with the Ising model and NISQ, in which we utilized the symmetry of the investigated model.

Keyword: Ising Model, Variational Quantum Eigensolver, Quantum Optimization, Ansazt Construction

I. INTRODUCTION

From the early stage of studying quantum computers, that by its very quantum nature is promising for a bright new age of computation coming along with three crucial keys of quantum theory: quantum probabilistics, superposition, and entanglement. The distinct properties of quantum computerscomputers (QC) from classical computers make it be more unique in application, not only be an upgraded computational speed version of traditional computers. Currently, when the quantum hardware is rather limited on operational stats and noise resilience, despite of the ambiguity of the realization of such advantages, Variational Quantum Algorithms (VQAs) [1, 2] are thought to be the best at outperforming conventional computers, which are able to implement well on near-term quantum devices known as the so-called Noisy Intermediate-Scale Quantum (NISQ) computers.

Nowadays, when the quantum revolution is being on track, big tech companies like IBM, Google, D-Wave, etc. compete to build their own quantum computers to yield quantum supremacy, where the first-ever experimental demonstration was achieved by the Google AI Quantum team [3], but that is still far from what we expect the quantum computer could do, while coherence time, connectivity of qubit, and qubit number limitations keep us from being able to successfully run a long depth circuit and produce a significant result on the present noisy device, in addition to quantum gate implementation issues. Those restrictions would yet be challenging to hardware scientists in the near future, posing obstacles that computer scientists would need to weigh against trainability, precision, and efficiency, nevertheless, these NISQ devices are capable of exploitation [4]. Standing out among quantum algorithms envisioned to beat classical computers, which are mostly designed for the fault-

———

* leductruyenphys@gapp.nthu.edu.tw
† duy.nguyenvan@phenikaa-uni.edu.vn

tolerant quantum computer, VQAs turns up regarded as an appropriate candidate compatible with the current defective quantum device to address these constraints [2]. The first two prominent applications of VQAs coming up are the Variational Quantum Eigensolver (VQE) and the Quantum Approximate Optimization Algorithm (QAOA). These inherit the core scheme of a VQA, which is to utilize the hybrid routine, leverage the versatility and computational power of classical computers to handle the computation processing, and use quantum devices to execute the quantum circuit. Hence, in this paper's scope, we will travel through the standard procedures and investigate the properties of different methods implemented in a VQE application.

The structure of this paper is as follows, in Chapter 1, we provide an overview of the Variational Quantum Eigensolver (VQE) procedures. A brief introduction to the Transverse Ising Model and its symmetries is discussed in Chapter 2, which will be utilized to inform the ansatz structure. Subsequently, Chapter 3 delves into optimization methods, encompassing both classical and quantum approaches, where we will propose a new quantum optimization method. The numerical study and experiment are conducted in Chapter 4, and then we make the final conclusion in Chapter 5.

II. VARIATIONAL QUANTUM EIGENSOLVER (VQE)

VQE was initially proposed by Peruzzo et al. [5, 6] as an efficient alternative way to compute the ground state energy of many quantum chemistry systems [7, 8] or many-body systems, i.e. ground-state He-H$^+$ molecular energy, rather than Quantum Phase Estimation (QPE) [9] which demands impractically huge numbers of quantum gates. Consequently, VQE becomes a highly flexible, viable strategy with currently potential quantum resources that can outpace conventional computers. Following the success of running on the photonic quantum processor combined with the traditional devices, many research works are favored to develop various types of VQA reviewed in Ref. [4, 10]. In essence, what makes VQE (generally VQAs) engrossing is the trial state analysis step to choosing an ansatz adaptive to particular quantum device architecture and/or problems. Subsequently, the suitable ansatz is executed congruently on the NISQ device, the remainders are then handled by the classical computer with the hybrid loop to generate remarkable results with the aid of error mitigation techniques. Instead of diagonalizing a matrix representation of Hamiltonian $\hat{H}$ to find out eigenvalues that exponentially scale up the matrix size and calculation overhead in traditional computing as the problem grows up larger, VQE just focuses on catching up with the ground state

energy E_g basically based on the variational principle

$$E_g \leq E\left[\Psi(\boldsymbol{\theta})\right] = \frac{\langle\Psi(\boldsymbol{\theta})|\hat{H}|\Psi(\boldsymbol{\theta})\rangle}{\langle\Psi(\boldsymbol{\theta})|\Psi(\boldsymbol{\theta})\rangle} = \langle\hat{H}\rangle_{\hat{U}(\boldsymbol{\theta})}. \quad (1)$$

The arbitrary quantum state $|\Psi(\boldsymbol{\theta})\rangle \equiv \hat{U}(\boldsymbol{\theta})|\Psi_0\rangle$ is a trial solution, the so-called *ansatz* parametrized by a unitary operator $\hat{U}(\boldsymbol{\theta})$, so that when the parameter $\boldsymbol{\theta}$ varies, the ansatz $|\Psi(\boldsymbol{\theta})\rangle$ is readily capable to span on the relevant quantum space where the ground state located. By modifying $\boldsymbol{\theta}$ value systematically, the VQE task is to travel around spanned space to extract the ground state minimizing the energy function $E\left[\Psi(\boldsymbol{\theta})\right]$ or, in general VQA, the energy function described by $\mathcal{L}\left(\boldsymbol{\theta}, \langle\hat{H}\rangle_{\hat{U}(\boldsymbol{\theta})}\right)$, and the objective is

$$\min_{\boldsymbol{\theta}} \mathcal{L}\left(\boldsymbol{\theta}, \langle\hat{H}\rangle_{\hat{U}(\boldsymbol{\theta})}\right). \quad (2)$$

In fact, the optimal wave function is not necessarily required to close to the ground state solution, due to the error in energy being in the second order of the quantum state error. Formally, the classical approach for VQE insists on an analytical expression of the lost function $\mathcal{L}\left(\boldsymbol{\theta}, \langle\hat{H}\rangle_{\hat{U}(\boldsymbol{\theta})}\right)$, in particular, the wave function expression $|\Psi(\boldsymbol{\theta})\rangle$ that becomes eventually unachievable going along with more and more problem complexities. Nevertheless, this weakness is literally the quantum advantage we have in the VQE quantum computing approach. Quantum computers are adopted to simulate quantum states, and state expectation outcomes are subsequently measured to pipeline to calculate the lost function classically. And partly because of lost function precision greater quadratically than the state function, indeed, we need two states to sandwich operators, VQEs are hereby friendly with noisy devices, which can work out meaningful results even though the states are disturbed by noises. These motivations are robust enough to let us dive deeper into VQE operation, ideally with a noise-free assumption in this paper's scope though, the noisy one will be addressed in the future study.

The structure of a VQE algorithm is depicted schematically as Fig. 1 within four basic steps: *Hamiltonian construction, Ansatz preparation, Measurement strategy* and *Optimization*.

Hamiltonian construction: Starting from a given abstract problem, eg. molecule ground state energy, shortest path, transshipment problems, an initially mathematical form called the Hamiltonian $\hat{H}$ is modeled. A mapping scheme is needed to convert $\hat{H}$ into a new form made out of operators that are able to execute on a quantum computer. Invoking prior knowledge of the physical system, particular symmetries are drawn on to simplify the Hamiltonian, such information is usefully reused to guide the erection of other VQE subroutines.

Ansatz preparation: To pick out a good ansatz, an essential condition that must be satisfied is that the space spanned by the ansatz contains the desired state to extremize the objective function. An obvious one is a
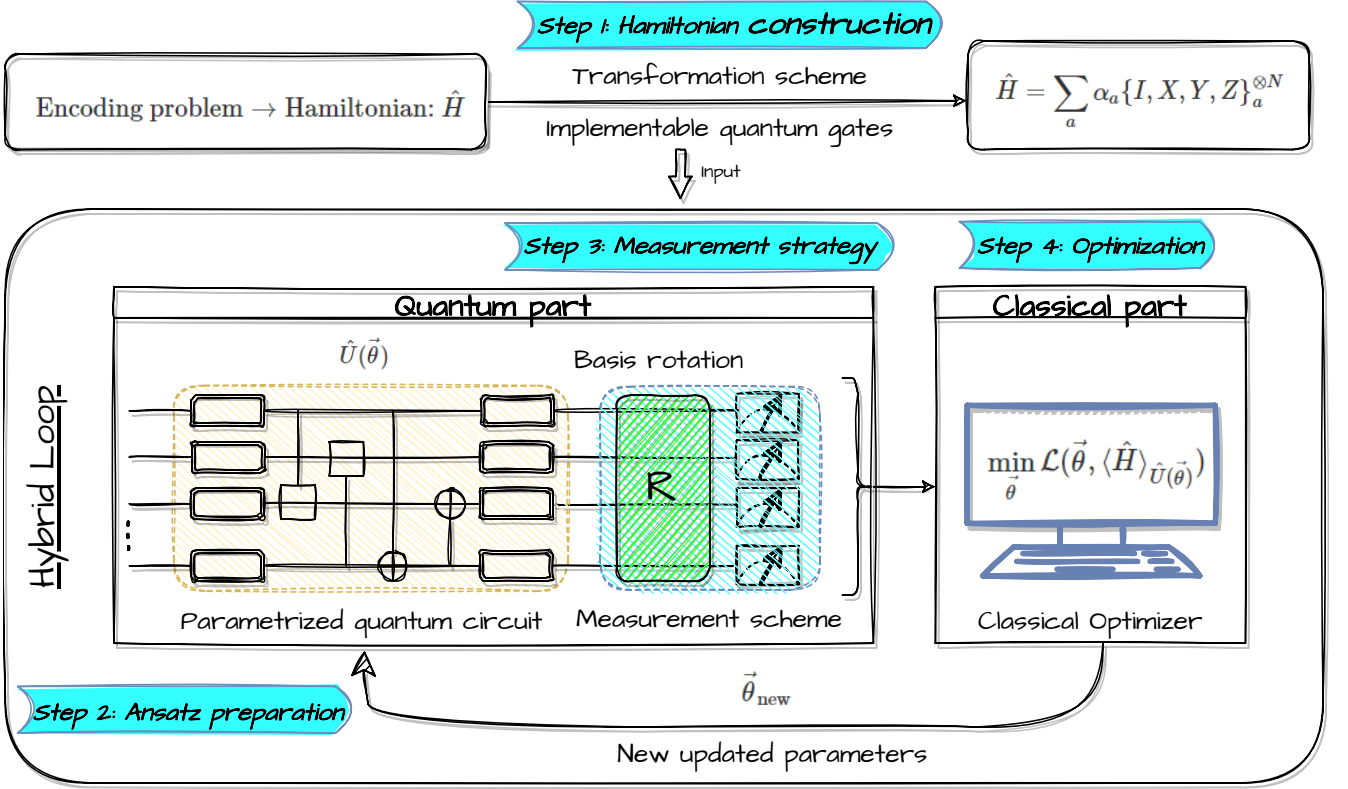
FIG. 1: VQE Architecture.

generic ansatz spread over all Hilbert space, a quantum circuit for that kind of ansatz is usually generated by multi-control-$U_3(\boldsymbol{\theta})$ gates so that all problems could be ultimately settled without ansatz concern. Nevertheless, that is not the end of the story, the computation cost of such a multi-control-$U_3(\boldsymbol{\theta})$ gate is extremely high and even unfeasible when deployed on the NISQ device. As a consequence, an initial crucial step in VQE is to find an ansatz that is expressive enough to capture the important features of the wavefunction but also efficient enough to be optimized using optimization techniques. In this article, we employ the approach that excavates up symmetries and physical properties of the physical system to guide the design of the ansatz respecting those symmetries, which helps us considerably reduce the number of training parameters and circuit depth needed in a theoretically speculative way, thereby improves the efficiency of the optimization. There are also other methods that use a hierarchical ansatz, where the circuit is divided into layers and each layer is optimized independently, and also use machine learning to automatically generate the quantum circuit [11, 12].

Measurement: To read off information from the wave function to estimate the expectation value of the Hamiltonian, apart from the standard basis measurement where each qubit in the output state is measured in the computational basis, the measurement scheme is manifestly came up with is a unitary transformation to the di-

agonal basis of observable operators in the Hamiltonian, typically defined by tensor products of Pauli operators, such as X, Y, and Z. Almost Hamiltonian of interest are hermitian and be able to decompose in terms of Pauli strings

$$\hat{H} = \sum_i \otimes_{j=1}^N \sigma_j^i, \qquad (3)$$
$$\sigma_j^i \in \{I, X, Y, Z\},$$

each Pauli string term $P^i = \otimes_{j=1}^N \sigma_j^i$ has likely overlapped the others in somewhere qubit sites, which means $\sigma_j^m \equiv \sigma_j^k$ being able to measure simultaneously, the grouping methods skillfully make use of this information to lower the number of measurements, the symmetry structures of the problem are able to exploit as well, and other efficient measurement methods are reviewed in [13]. Those on-research topics are highly crucial to the current NISQ device when a minimum measurement performance is required upon a given restricted number of shots. Furthermore, for a given ϵ accuracy, VQE demands a trade-off between the $\mathcal{O}(\frac{1}{\epsilon^2})$ number of circuit samples and $\mathcal{O}(1)$ circuit depth with other finding ground state energy methods such as QPE, which consumes $\mathcal{O}(1)$ circuit measurement repetitions with depth $\mathcal{O}(\frac{1}{\epsilon})$. A better approach is essentially proposed to reconcile disadvantages between the two above methods, which is deemed a generalized VQE algorithm, α-VQE, when outrunning VQE by just $\mathcal{O}(\frac{1}{\epsilon^{2(1-\alpha)}})$ circuit samples and lower circuit depth $\mathcal{O}(\frac{1}{\epsilon^\alpha})$ than QPE, where $\alpha \in [0, 1]$ [14].

Optimization: Last but not least, the last piece in VQE subroutine is to vary the parameter $\boldsymbol{\theta}$ to catch up with the global extreme point (but in many cases, the local one is preferred) in the function of interest described by Eq. (2). The criteria should be pondered to problem-friendly implement, which are able to low-cost function evaluation, fast convergence with a given quantum of resources, and silence to noise. Based on different kinds of utilizing objective function information, people classify it into two main optimization approaches: gradient-based and derivative-free. Meanwhile, the gradient-based employs the gradient information in mostly first order, and second order to travel along the steepest direction of the cost function landscape. The derivative-free rather directly uses responses of the objective function at different points in the parameter space to iteratively improve the parameter values, and as usual, these utilize the subroutine classical algorithm and just consume a limited quantum evaluation expense than gradient-based, these are thus seemingly more congruous with NISQ. In the context of VQAs, gradient-free approaches can be used to optimize quantum circuit parameters when computing the gradient is either impractical or excessively costly, for example, when simulating the behavior of a quantum system becomes computationally expensive, or when the objective function cannot be expressed analytically. Nevertheless, the gradient-based methods are arguably more reliable and have faster convergence as the more complicated structure of the objective function, henceforth, the total trade-off would be shoulder-by-shoulder when deliberately adopting between these two. Furthermore, the step size problem in gradient-based and applying machine learning in derivative-free are active topics of research at this stage, see [4] section II.D for more reviews. In the context of this paper, we investigated several optimization methods and instead labeled them again into two regimes, classical and quantum optimization.

III.   ANSATZ FOR TRANSVERSE ISING MODEL

A.   Transverse Ising model (TIM)

The Transverse Ising model was introduced first in 1963 by de Gennes when he built a model for describing the low-frequency collective modes of protons in the ferroelectric phase of ferroelectric crystals KH_2PO_4 [15]. From then until now, the Transverse Ising Model has been used to represent enormous simple-to-complex systems such as brain science and information science and technology.

We consider the 1D Transverse Ising, describing the nearest-neighbor interactions of the spin projection along the z-axis and uniform external magnetic field along the x-axis (in principle, perpendicular to the z-axis). Defined by the two-spin exchange interaction factor J and representative strength of the external field h as described below

$$\hat{H}_{\text{TIM}} = -J \sum_{n=1}^{N-1} \sigma_{n-1}^z \sigma_n^z - h \sum_{n=0}^{N-1} \sigma_n^x. \qquad (4)$$

The Hamiltonian in Eq. (4) possesses a $\mathbf{Z}_2$ symmetry remaining invariant under spin-flipping action. When the coupling constant is $h < 1$, the system exhibits a ferromagnetic phase, and the spins favorably align along the z-direction. Conversely, for $h > 1$, the system transitions to a disordered paramagnetic phase. The σ_x terms take a dominant role when $h \to \infty$, leading the ground state to align predominantly in the $|+\rangle^{\otimes N}$ state. At the critical point, $h = 1$ renders the gapless property in the thermodynamic limit.

One of the key differences between the classical and quantum versions of the Ising model is that the expectation value of two terms in the Eq. (4) must be measured separately due to the non-commutative feature of these elements. However, the commutation property allows us to perform only two observations to extract the Hamiltonian's expectation. The first term $\sum_{n=1}^{N-1} \sigma_{n-1}^z \sigma_n^z$ can be cumulatively achieved by measuring directly on the computational basis as follows:

$$\langle \sigma_m^z \sigma_n^z \rangle = \sum_{\{q_0, \cdots, q_{N-1}\} = \{0,1\}} |a_{q_0 \cdots q_m \cdots q_n \cdots q_{N-1}}|^2 (-1)^{q_m + q_n},$$
$$(5)$$

$$\langle \sigma_n^z \rangle = \sum_{\{q_0, \cdots, q_{N-1}\} = \{0,1\}} |a_{q_0 \cdots q_n \cdots q_{N-1}}|^2 (-1)^{q_n}, \qquad (6)$$

with $|a_{q_0 \cdots q_m \cdots q_n \cdots q_{N-1}}|^2$ and $|a_{q_0 \cdots q_n \cdots q_{N-1}}|^2$ are the probabilities of being in the computational basis state $|q_0 \cdots q_m \cdots q_n \cdots q_{N-1}\rangle$ and $|q_0 \cdots q_n \cdots q_{N-1}\rangle$ respectively. Similarly, $\sum_{n=0}^{N-1} \sigma_n^x$ is obtained by applying the Hadamard gate or the $R_y \left(-\dfrac{\pi}{2} \right)$ gate in advance to transform to the Z-basis, or computational basis, and then taking the measurements following Eq. (6).

B.   Ansatz Construction

To construct an ansatz in terms of a parameterized quantum circuit (PQC), as mentioned above, we need the $U(\boldsymbol{\theta})$ operator to be able to span all quantum Hilbert space so that we can use it to solve all issues. Indeed, we do not, in practice though, a general $U(\theta)$ operator that is hard to implement experimentally because of the noisy and weighty composite $C-U_3$ gate and actually does not acquire overall good operation even in an ideal simulator. Hence, the ansatz implementation for a particular problem's purpose is a serious studying process of the VQE working roadmap. In addition to the compatibility

with the current quantum device, tapping into symmetries and properties of our problem could be leveraged to shrink the parameter space, which are the important criteria for adopting an ansatz.

1. Symmetry of the Transverse Ising model

At the scope of this article, we consider three properties of TIM so that we take its suggesting information into account to decrease the size of the ansatz

- *Real representation.* Using the eigenstates of the σ_Z (Pauli-z) operator as the elementary binary computational basis, in terms of which, σ^z, σ^x are real matrices, due to that, we can represent the TIM Hamiltonian in real form, the eigenstates of the TIM Hamiltonian can thus be chosen to be real for conventional purposes. Namely, considering $|\Psi\rangle = \sum_n C_n |n\rangle$ is an eigenstate of H_{TIM} expanding in the computational basis $|n\rangle$. The real form of H_{TIM} means the Hermitian real element matrix, that induces coefficients satisfied $C_n^* C_m = C_n C_m^*$ $\forall m, n \in [0, 2^N - 1]$, then, in generality

$$C_n = r_n e^{i(c + k_n \pi)}, \quad c \text{ is a constant}, \tag{7}$$

or, in other words, the angles in the complex plane of coefficients C_n differ from every other by a factor $k\pi$, where k and k_n are integers. Then, the complex angle can be shifted to the real coefficient $C_n^* = C_n$ by according to the quantum global phase principle.

- *Local interaction.* The first term $\hat{H}_{\text{TIM}}^{\text{LI}}$ in the TIM Hamiltonian describes the kind of neighboring spin interaction along the z-axis

$$\hat{H}_{\text{TIM}}^{\text{LI}} = \sum_{n=1}^{N-1} \sigma_{n-1}^z \sigma_{n+1}^z. \tag{8}$$

The interaction of two local spins formulated by this term causes a sort of entanglement structure between every two local spins of the ground state energy. Nevertheless, in the case of the order phase, when the external magnetic field is dominant, this interaction can be broken down, each spin will be free interacting and aligns in the same direction as the magnetic field.

- *Total spin-flip symmetry .* The most featured symmetry of the Ising model, can be referred to as the classical counterpart, Time-reversal symmetry (generally, $\mathcal{Z}_2$ symmetry). Under the total spin-flip transformation in the z direction $(\sigma^x)^{\otimes N}$

$$\left[(\sigma^x)^{\otimes N}, \hat{H}_{\text{TIM}} \right] = 0, \tag{9}$$

which implies the TIM Hamiltonian remains unchanged. This one leads to the eigenstate $|\Psi\rangle$ and

$(\sigma^x)^{\otimes N} |\Psi\rangle = |\tilde{\Psi}\rangle$ has the same energy, the relation between them is thus put in two cases

$$\langle \Psi | \tilde{\Psi} \rangle = \begin{cases} 0 & \text{Degeneracy} \\ \pm 1 & \text{Non-degeneracy} \end{cases}, \tag{10}$$

where, in the degenerate case, it $g = 0$ is trivial, and in the non-degenerate $g > 0$ case, the expanded coefficients in the z-direction basis representation have a structure

$$C_n = \pm C_{2^N - 1 - n}. \tag{11}$$

2. Ansatz selection

For the chosen ansatz, the conventional real coefficients of eigenstates tell us that it is enough to span in real quantum parameter space for finding the ground state energy, and the linear entanglement mapping comes from the information of the local interaction term in the Hamiltonian. Moreover, to accommodate the device constraints, in a speculative way, the available common-use RealAmplitudes ansatz is our good candidate for the implementation, which uses only the quantum gate $R_y(\theta)$ for the parametrized quantum circuit and the entanglement scheme as in Fig. 2.
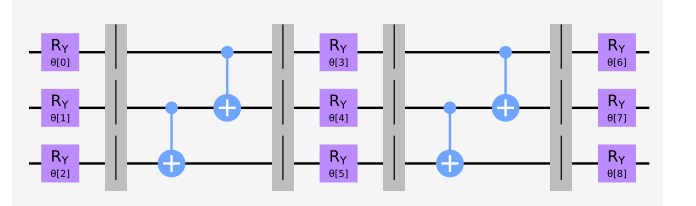

FIG. 2: A RealAmplitudes circuit with 2 layers on 3 qubits and the linear entanglement scheme.

The additional reason for choosing such a Hardware-efficient ansatz architecture [16] comes from the efficient implementation of quantum optimization to attain our full quantum algorithm purpose, which will be addressed in the next subsection IV B.

To efficiently travel around full quantum Hilbert space, we need to tune $2^{N+1} - 2$ degrees of freedom, $+1$ inside of the exponent indicates the complex number condition, and -2 turns up from normalization and global phase conditions. After bringing TIM's properties we investigated from section III B 1 to dissect, *Real representation* Eq. (7) helps us eliminate $+1$ within the exponent and *Total spin-flip symmetry* Eq. (11) reduces the number of parameters we have to search the ground state by half. Finally,

$$2^{N+1} - 2 \rightarrow 2^{N-1} - 1. \tag{12}$$

The number of layers L, which changes directly ansatz number of parameters p according to $p = N(L+1)$. Following Eq. (12), we roughly evaluate a necessary number of layer

$$L \geq \frac{2^{N-1} - 1}{N} - 1, \qquad (13)$$

notes that this is not a rigorous condition because ansatz's entanglement mapping scheme used also contributes, which is able to change the bound as well, it is still a good estimation for surface analysis nonetheless. Yet, we can realize that, not only TIM, but Eq. (12) also holds for all symmetrically real Hamiltonians being symmetric under $(\sigma^x)^{\otimes N}$ the operator. Therefore, there are some hidden TIM properties that we are unable to analytically incorporate to give an exact boundary value of L. Through conducting the experimental survey though, we are able to choose a reliable value, as we will see in chapter V B, it works well even for choosing $L = 2$ for all qubit scaling.

IV. OPTIMIZATION

The next subroutine in the VQE work map, which is a purely technical perspective we investigate to study the performance of different optimization methods, provides us a larger view of the way to reaching the faster and more precise convergence of our target value. In TIM, we try to minimize the cost function defined as the expectation value of Hamiltonian $\langle \hat{H}_{\text{TIM}} \rangle$ using common methods of optimization categorized into two class operations.

A. Classical Operation

Classical operation means regardless of the analysis of the quantum structure of the cost function, in particular ansatz structure, we are able to find the optimal parameters minimizing the objective function. Normally, in the sense of operation, we can do the classical optimization process independently from other VQE parts, and even with your problem-specific concerns, and appears that you can develop your optimization algorithm for general variational issues.

Constrained Optimization BY Linear Approximation (COBYLA). One of the most powerful derivative-free methods, which has been favored by many users in lately decades, COBYLA makes use of linear interpolation of the objective function at each iteration by a unique linear polynomial function at the vertices for finding an optimal vector parameter within the trust region, then feeding optimal point evaluated to the objective function to get the value improving the next iteration of approximation [17–20]. In VQE, COBYLA's linear interpolation at each step is a classical subroutine running on the CPU using only one objective value from the QPU evaluation running. And so because of ignored derivative information, it would beneficially avoid several problems in analysis optimizations, especially Barren plateaus landscape, in return less accuracy for more parameters ($p > 9$) [20]. For a bird's-eye view, new updates to COBYLA versions like UOBYQA, NEWUOA, and BOBYQA count the objective function's curvature information to increase convergence [21–23].

Finite Difference (FD). Let's dawn on one of the primary first-order derivative-based approaches, where the parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ is updated at the iteration k-th by

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta_k \nabla f(\boldsymbol{\theta}_k), \qquad (14)$$

which we usually name the Gradient Descent (GD) method. Finite Difference is a numerical technique to calculate the gradient vector $\nabla f(\boldsymbol{\theta}_k)$ without the analytical function's texture, which is the basic one used in the gradient descent method. We use the well-known central difference formula,

$$\nabla f(\boldsymbol{\theta}_k)_i \simeq \frac{f(\boldsymbol{\theta}_k + \epsilon \vec{i}) - f(\boldsymbol{\theta}_k - \epsilon \vec{i})}{2\epsilon}, \qquad (15)$$

where $\vec{i} \in \mathbb{R}^p$ is i-th the unit vector and ϵ is an infinitesimal change, with the error proportional to $\mathcal{O}(\epsilon^2)$. The smaller ϵ we use, the more exact result we get. However, due to the restricted accuracy of a classical computer, we are unable to achieve a value of ϵ that is too small, which even becomes greater when implemented on a quantum device, where at least the sample error enters the picture.

Simultaneous Perturbation Stochastic Approximation (SPSA). To surmount some of obstacles that emerge from gradient descent operation on the near-term devices, the idea of a perturbing stochastic approximation method is a good choice. SPSA generates an unbiased estimator $\tilde{f}(\boldsymbol{\theta}_k)$ of the gradient by simultaneously randomly perturbating the gradient direction of parameters [24], we replace the ordinary gradient vector $\nabla f(\boldsymbol{\theta}_k)$ by

$$\nabla f(\boldsymbol{\theta}_k) \rightarrow \nabla \tilde{f}(\boldsymbol{\theta}_k) = \frac{f(\boldsymbol{\theta}_k + s_k \vec{\Delta}_k) - f(\boldsymbol{\theta}_k - s_k \vec{\Delta}_k)}{2s_k} \vec{\Delta}_k, \qquad (16)$$

where $\vec{\Delta}_k$ is a random perturbed vector sampled from the zero-mean distribution, usually, the Bernoulli distribution is used. We can see that all parameters are simultaneously shifted by a random amount ($\pm s_k$), our computation therefore only requires two objective function evaluations per iteration. Whereas standard gradient computation time is scaled along to the number of parameters, the SPSA optimizer is, however, independent, which saves a lot of time as we work on a higher parameter regime. Besides that, noise from quantum circuit executions computing the objective value be regarded as a stochastic perturbation part absorbed into the SPSA procedure. These advantages promote SPSA and its other versions to be efficient techniques in the NISQ era.

B. Quantum Operation

As opposed to the classical ones, the quantum operations require the quantum information extracted from the structure of the trial wave function. This sort of behavior inevitably entails the optimization process in whole problem analysis, especially related closely to Ansatz Construction.

Parameter-shift rules (PSR). Inspired by the classical shift rules in the exact derivative computation of some special function, we are successfully able to figure out the analytical value of the objective function's derivatives using quantum devices. For any kind of quantum gate has the form

$$\hat{\mathcal{G}}(\theta) = e^{-i\theta\hat{G}} \tag{17}$$

generated by the Hermitian operator $\hat{G}$. Supposed that we have an expectation value of $\hat{H}$, which is our objective function $f(\theta) = \langle\psi(\theta)|\hat{H}|\psi(\theta)\rangle$ parametrized by θ. The ansatz wave function $|\psi(\theta)\rangle = \hat{U}(\theta)|\psi\rangle_I$ is made up of $\hat{U}(\theta) = \hat{A}\hat{\mathcal{G}}(\theta)\hat{B}$, with $\hat{A}$, $\hat{B}$ are arbitrarily other operators. As a result, the partial derivative of the $f(\theta)$ with respect to θ is thus obtained via the parameter-shift rules [25]

$$\partial_\theta f(\theta) = s\left[f\left(\theta + \frac{\pi}{4s}\right) - f\left(\theta - \frac{\pi}{4s}\right)\right]. \tag{18}$$

The partial derivative $\partial_\theta = \frac{\partial}{\partial\theta}$ implies we can generalize to a set of multiple parameters $\{\theta_i\}$. The evaluation of $f(\theta \pm \frac{\pi}{4s})$ can be easy to run on quantum computers, and then the exact amplitude of the vector gradient is derived. Within the aim of this paper, the given parametric ansatz is made out of Pauli rotations, and $s = \frac{1}{2}$ is selected accordingly. The state-of-the-art formalism for more generic PQCs is also able to be derived [26, 27].

Quantum Natural Gradient Descent (QNG). How to improve the convergence? The global fixed learning rate η_k is apparently not a good choice, one does not be good sensitive to the model information with respect to parameter changes. Initially, many attempts try to tune the learning rate η_k such as learning rate schedulers that vary η_k after an iteration, latterly, adaptive methods that count previous iteration values, or capturing the curvature of the objective function using the diagonal approximation of the Hessian (HES), where each η_k is the inverse Hessian's diagonal element instead of being equally fixed. Those methods educe that we can embed further information rather than only just gradient vector, to make the optimal step size for our variational quantum algorithms. Likewise, the classical counterpart using the Fisher Information Matrix (FIM), Quantum Natural Gradient Descent invokes the quantum geometry of the wave function, which can make us get the optimal spot faster. Namely, we transform to the Riemann parameter space using the metric $g \in \mathbb{R}^{p\times p}$ of projected Hilbert space $\mathcal{PH}$ instead of the current Euclidean space $g = \mathbb{1}_{p\times p}$. The update

Eq. (14) turns to

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta_k g^+(\boldsymbol{\theta}_k)\nabla f(\boldsymbol{\theta}_k), \tag{19}$$

$g^+(\boldsymbol{\theta}_k)$ means it is a pseudo-inverse local metric. Riemann parameter space metric g is generally defined

$$ds^2 = g_{ij}d\theta_i d\theta_j. \tag{20}$$

The Hilbert space $\mathcal{H}$ of "bare" quantum states reduces to the $\mathcal{PH} = \mathcal{H}/U(1)$ space as we ignore the local phase $U(1)$ of quantum states [28], the quantum distance in $\mathcal{PH}$ is then based on to calculate the Riemann metric g

$$ds^2 = 1 - |\langle\psi_\theta, \psi_{\theta+d\theta}\rangle|^2$$
$$= 1 - \left|1 + \frac{1}{2}\langle\psi_\theta|\partial_i\partial_j\psi_\theta\rangle d\theta_i d\theta_j + \langle\psi_\theta|\partial_i\psi_\theta\rangle d\theta_i\right|^2$$
$$= \text{Re}[G_{ij}]d\theta_i d\theta_j. \tag{21}$$

We expand $|\psi_{\theta+d\theta}\rangle = |\psi_\theta\rangle + |\partial_i\psi_\theta\rangle d\theta_i + \frac{1}{2}|\partial_i\partial_j\psi_\theta\rangle d\theta_i d\theta_j$ upto second order in $d\theta$, where $\partial_i = \frac{\partial}{\partial\theta_i}$. The Quantum Geometry tensor (QGT) $G_{ij} \in \mathbb{R}^{p\times p}$ has two parts: the anti-symmetric imaginary part $\sigma_{ij} = -\sigma_{ji}$ related to the gauge field of $U(1)$ eventually vanishes, the only contribution comes from the real symmetric part $g_{ij} \equiv g_{ij}(\boldsymbol{\theta})$ reflects the quantum distance in $\mathcal{PH}$ space, where its formula is

$$g_{ij}(\boldsymbol{\theta}) = \text{Re}\left[\langle\partial_i\psi_\theta|\partial_j\psi_\theta\rangle - \langle\partial_i\psi_\theta|\psi_\theta\rangle\langle\psi_\theta|\partial_j\psi_\theta\rangle\right]. \tag{22}$$

The Fubini-Study metric tensor $g_{ij}(\boldsymbol{\theta})$ is a quantum analogue of the Fisher information matrix (QFIM). Given that it is proportional to p^2, the complicated computing and computational cost of $g(\boldsymbol{\theta})$ is high and incompatible with the short-term quantum device. To solve this problem, an approximation strategy is required.

- *QNG - Block Diagonal Approximation (QN-BDA).* Moreover, more than Diagonal Approximation, where we just count p elements in the diagonal line, the QN-BDA employing the Parametric Family circuit is conveniently deployed entirely in the computation of the approximated metric tensor on the quantum computer. The parametric unitary operator $\hat{U}(\boldsymbol{\theta})$ acts on the initial state $|\psi\rangle_I$ entailing L layers

$$\hat{U}(\boldsymbol{\theta}) = S_L P_L(\theta^L)\ldots S_1 P_1(\theta^1), \tag{23}$$

where:

- S_l are the static parts, which usually denote to the entanglement layer.
- $P_l(\theta^l)$ are the parametric parts, which can be decomposed to single qubit gates $P_l(\theta^l) = \bigotimes_{i=1}^N R_i(\theta_i^l)$.
- The single qubit gate $R_i(\theta_i^l) = \exp\{[i\theta_i^l K_i]\}$ is constructed from Hermitian generator K_i with the parameter $\theta_i^l \in \theta^l = \{\theta_1^l, \ldots, \theta_N^l\}$.

Such a type of parametrized circuit whose nice properties we can prospect to yield a block diagonal form of QGT running completely on the quantum processor, each block corresponding to each layered vector parameter $\theta^l \in \boldsymbol{\theta} = \theta^1 \oplus \ldots \oplus \theta^L$. We denote

$$\hat{U}_n^m = S_m P_m(\theta^m) \ldots S_n P_n(\theta^n), \qquad (24)$$

then $\hat{U}(\boldsymbol{\theta}) \equiv \hat{U}_1^L = \hat{U}_{l+1}^L S_l P_l(\theta^l) \hat{U}_1^{l-1}$. The partial derivative state can be written in the form

$$|\partial_i \psi_\theta\rangle = \partial_i \hat{U}(\boldsymbol{\theta}) |\psi\rangle_I \qquad (25)$$
$$= \hat{U}_{l+1}^L S_l \partial_i P_l(\theta^l) \hat{U}_1^{l-1} |\psi\rangle_I \qquad (26)$$
$$= \hat{U}_l^L (iK_i) \hat{U}_1^{l-1} |\psi\rangle_I \qquad (27)$$
$$= \hat{U}_l^L (iK_i) |\psi_\theta^{(l-1)}\rangle. \qquad (28)$$

Note that $[K_i, K_j] = 0$ in each layer, the unitarity of $\hat{U}_n^m$ and overlap of partial derivative states tell us the block matrix element

$$g_{ij}^{(l)}(\boldsymbol{\theta}) = \langle \psi_\theta^{(l-1)} | K_i K_j | \psi_\theta^{(l-1)} \rangle$$
$$- \langle \psi_\theta^{(l-1)} | K_i | \psi_\theta^{(l-1)} \rangle \langle \psi_\theta^{(l-1)} | K_j | \psi_\theta^{(l-1)} \rangle \qquad (29)$$

of the block-diagonal form of the Fubini-Study metric [29]

$$g_{ij}(\boldsymbol{\theta}) = \begin{matrix} & \begin{matrix} \theta^1 & \theta^2 & \ldots & \theta^L \end{matrix} \\ \begin{matrix} \theta^1 \\ \theta^2 \\ \vdots \\ \theta^L \end{matrix} & \begin{pmatrix} g^{(1)} & 0 & \ldots & 0 \\ 0 & g^{(2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & g^{(L)} \end{pmatrix} \end{matrix}. \qquad (20)$$

QN-BDA approximation involves L quantum evaluations and works well on models with weak correlation.

Quantum Natural-Simultaneous Perturbation Stochastic Approximation (QN-SPSA). Inheriting the SPSA idea to calculate the gradient, it is capable to put up generating the Hessian matrix with fewer evaluations, the so-call 2-SPSA. Normally, the HES or FIM matrix consumes $\mathcal{O}(p^2)$ quantum expectation, QN-BDA enables us to turn to relying on the number of layers $\mathcal{O}(L)$, saving us lots of computational sources. Even further, QN-SPSA just only executes four quantum runs to obtain the full second-order derivative matrix. Let's consider the Hessian of the Fubini-Study metric,

$$H_{ij}(\boldsymbol{\theta}) \equiv g_{ij}(\boldsymbol{\theta}) = - \frac{1}{2} \partial_i \partial_j |\langle \psi_\theta, \psi_{\tilde{\theta}} \rangle|^2 \Big|_{\tilde{\theta}=\theta}, \qquad (30)$$

which is just the Hessian form of Eq. (22) so that we can deploy the 2-SPSA method to generate the QN-SPSA matrix. We can see the equivalence

$$- \frac{1}{2} \partial_i \partial_j |\langle \psi_\theta, \psi_{\tilde{\theta}} \rangle|^2 \Big|_{\tilde{\theta}=\theta} = -\partial_i \mathrm{Re} \left\{ \langle \psi_\theta, \psi_{\tilde{\theta}} \rangle \langle \psi_{\tilde{\theta}}, \partial_j \psi_\theta \rangle \right\} \big|_{\tilde{\theta}=\theta}$$
$$= -\mathrm{Re} \left\{ \langle \psi_\theta, \partial_i \partial_j \psi_\theta \rangle + \langle \partial_i \psi_\theta, \psi_\theta \rangle \langle \psi_\theta, \partial_j \psi_\theta \rangle \right\}$$
$$= -\mathrm{Re} \left\{ -\langle \partial_i \psi_\theta, \partial_j \psi_\theta \rangle + \langle \partial_i \psi_\theta, \psi_\theta \rangle \langle \psi_\theta, \partial_j \psi_\theta \rangle \right\}, \qquad (31)$$

which is exactly the same as Eq. (22). Applying the 2-SPSA approach to compute the second-order derivative of the function $F(\boldsymbol{\theta}, \tilde{\theta}) = -\frac{1}{2} |\langle \psi_\theta, \psi_{\tilde{\theta}} \rangle|^2$ instead of our loss function $f(\boldsymbol{\theta})$ as being in Newton method. The core estimator in second-order SPSA is perturbed by two random vectors $\vec{\Delta}_k^1, \vec{\Delta}_k^2 \in \mathcal{U}^p\{-1, 1\}$ at step k-th

$$\triangle F = \frac{-1}{2} \Big[F(\boldsymbol{\theta}_k + s_k \vec{\Delta}_k^1 + s_k \vec{\Delta}_k^2, \boldsymbol{\theta}_k)$$
$$- F(\boldsymbol{\theta}_k + s_k \vec{\Delta}_k^1, \boldsymbol{\theta}_k) + F(\boldsymbol{\theta}_k - s_k \vec{\Delta}_k^1, \boldsymbol{\theta}_k)$$
$$- F(\boldsymbol{\theta}_k - s_k \vec{\Delta}_k^1 + s_k \vec{\Delta}_k^2, \boldsymbol{\theta}_k) \Big], \qquad (32)$$

which is composed of four terms respective to four quantum expectations we run on the quantum processor. Then, the Fubini-Study metric Eq. (30) is replaced by the QN-SPSA metric at k-th iteration

$$g^k(\boldsymbol{\theta}) \to \bar{H}^k(\boldsymbol{\theta}) = \frac{\triangle F}{4s_k^2} \left(\vec{\Delta}_k^1 \vec{\Delta}_k^{2\mathrm{T}} + \vec{\Delta}_k^2 \vec{\Delta}_k^{1\mathrm{T}} \right), \qquad (33)$$

where $\bar{H}^k(\boldsymbol{\theta}) \in \mathbb{R}^{p \times p}$ and s_k is a small positive hyperparameter being able to be tuned. The metric is, however, still too much stochastic, some helpful techniques are invoked to resolve that hitch [30, 31], say, counting information from previous updates to smooth the estimator

$$\tilde{H}^k = \frac{k}{k+1} \tilde{H}^{k-1} + \frac{1}{k+1} \bar{H}^k, \qquad (34)$$

eventually, because the estimator is still ill-conditioned and unstable. To satisfy the positive semi-definite warranting the local convex analysis and invertibility condition of $\tilde{H}^k$, the below replacement is needed accordingly,

$$\tilde{H}^k \to \sqrt{\tilde{H}^k \tilde{H}^k} + \beta \mathbb{1}, \qquad (35)$$

where the second term $\beta \mathbb{1} \in \mathbb{R}^{p \times p}$ corresponds to the invertibility condition. The effect from geometry metric is suppressed as the positive regulator gets a huge value $\beta \gg 0$, which indeed reduces to the standard gradient descent scenario and is more unstable. Therefore, that, in turn, causes a larger deviation in the average sample result when $\beta \to 0$. That makes a constant regulator β is thus a trade-off between Quantum Natural information and numerical instability. Another regularization approach, called "half-inversion", generalizes the power of the second-order derivative matrix to n instead of the usual ones -1 corresponding to standard natural gradient and 0 standing for original gradient, given by [32].

Quantum Natural-Simultaneous Perturbation Stochastic Approximation with Parameter-Shift Rule (QN-SPSA+PSR) is our extension for the QN-SPSA+SPSA method. Although QN-SPSA+SPSA has smaller complexity when compared with other methods, this stochastic method seems so randomized and made it harder to converge and less stable when it comes to a bigger cost function landscape. Here, We propose the QN-SPSA+PSR algorithm, which is inherent in approximating the Fubini-Study metric properties in a way of saving computational cost. However, instead of using the SPSA method to approximate the gradient of the cost function, we use the analytical method to calculate the gradient by PSR, which would compensate for the instability of the stochastic approximation in the Fubini-Study metric and drive itself in the gradient direction. This is expected for more stable and fast convergence compared to the former method QN-SPSA+SPSA.

V. RESULTS

A. Complexity estimation

Derivative-based method efficiency			
$\triangledown f(\boldsymbol{\theta})$	PSR	FD	SPSA
Comp. cost	2p	2p	2
$g(\boldsymbol{\theta})$	HES-PSR	QN-BDA	QN-SPSA
Comp. cost	3p.p	L	4

TABLE I: Showing comparison among combinations of gradient and adaptive learning rate methods. Where p is the number of parameters, L is the number of PQC layers of the same rotation generator.

In summary of the derivatives-based optimization framework, we tell apart two subroutines: gradient calculation and $g(\boldsymbol{\theta})$ matrix evaluation. For gradient calculation, we got three approaches: PSR offering exact gradients of quantum ansatz structure, FD-a classical numerical method that approximates gradients, and SPSA-a stochastic approximation technique requiring only two function evaluations per iteration. For the $g(\boldsymbol{\theta})$ matrix calculation, we generally have four methods: Const—an identity matrix; HES-PSR, which provides an exact second-order Hessian matrix by the PSR method of the quantum structure objective; QN-BDA approximates the Fubini-Study metric using the block diagonal structure; and QN-SPSA, which stochastically approximates the FIM with only four evaluations per iteration. A method is a combination of the gradient and $g(\boldsymbol{\theta})$, such as QN-BDA+PSR, which implies the application of the QN-BDA technique and PSR for the estimation of the gradient.
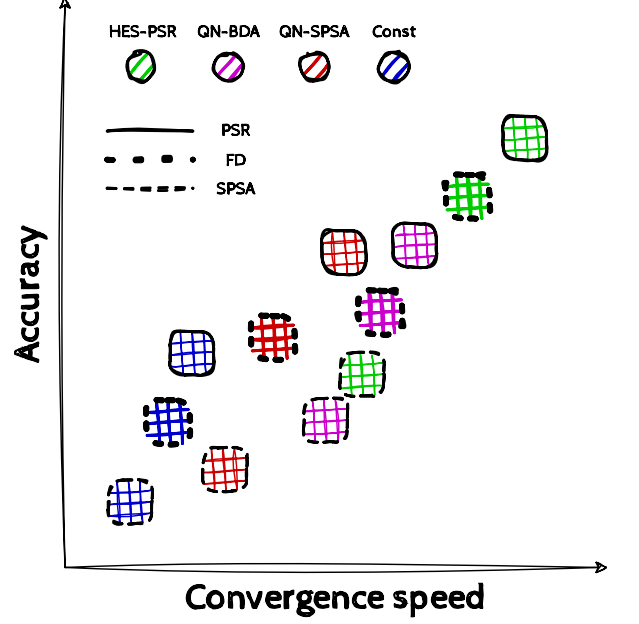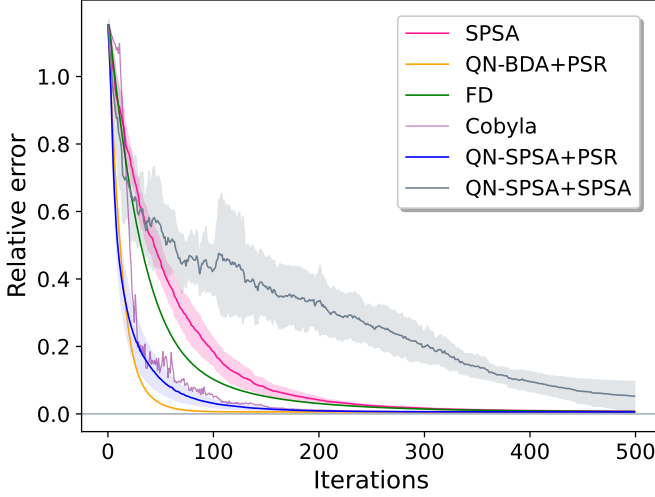

FIG. 3: The illustratively comparative graphics among combinations of gradient and adaptive learning rate methods.
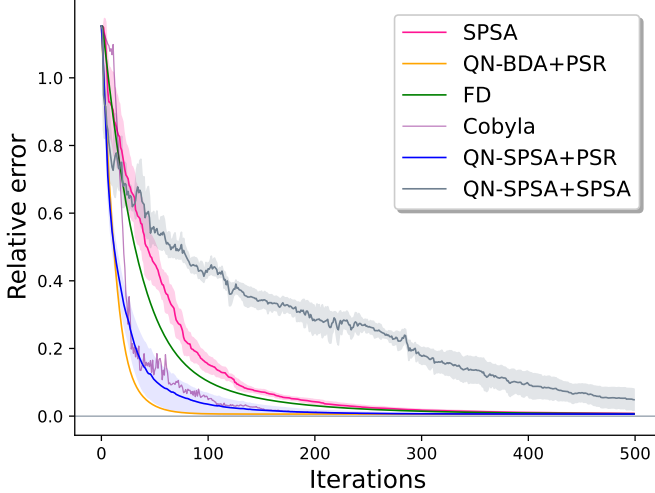
We provide a rough assessment of the optimization methods, grounded in theoretical judgment, where you are able to read off their information from Tab. I and Fig. 3 for more illustrative purpose, demonstrating the trade-offs in performance and resource requirements among methods, e.g., computational cost, accuracy, and convergence speed. Tab. I is derived from the underlying theories of methods discussed above. Besides that, Fig. 3 provides an intuitive estimation based on theory-driven and bias-informed conjectures. In this estimation, the methods HES-PSR, QN-BDA, QN-SPSA, and Const are ranked in descending order of convergence, accuracy, and computational cost. Similarly, for gradient methods, SPSA, FD, and PSR follow an ascending order of accuracy, convergence, and computational cost. Consequently, one can infer, for instance, that HES-PSR+PSR should achieve the best performance at the expense of the intractably highest computational cost, while QN-BDA+PSR consumes much quantum overhead but is expected to outperform QN-SPSA+SPSA overall.

From this framework, a key motivation of the QN-SPSA+PSR algorithm is to merge the computational resource efficiency of QN-SPSA with the precision of the PSR. The computational cost of QN-SPSA, at four quantum evaluations per iteration, demonstrates remarkable scalability compared to traditional Hessian-based approaches like HES-PSR or QN-BDA, which scale quadratically and linearly with the number of parame-

ters and layers, while the PSR would keep the method at good stability and convergence. This refinement enables faster convergence and greater stability while maintaining a low computational overhead, highlighting its potential for practical application on NISQ devices. Notably, these are only theory-based rough estimations. In practice, the complexity of objective functions and the influence of noise would lead to variations in the performance of these methods, and approximate methods, such as methods using SPSA, would have advantages as noises come about. The experimentally numerical results, however, demonstrate an even better performance of QN-SPSA+PSR than predicted by the estimation, as we will present in the next section.

B. Simulation results

We performed a detailed investigation of the VQE algorithm applied to the TIM model, testing several prominent optimization methods, ansatz structures, and entanglement configurations. Our simulations revealed several important insights into the performance of optimization methods and ansatz structures, highlighting the practical advantages of the proposed QN-SPSA+PSR algorithm. Remember that, if not specified, the external field and coupling constants were set to the default values $h = 2$ and $J = 1$, respectively. For the stochastic methods, which are evaluated based on seven samples. The source code can be found at GitHub[33].

(a) Linear entanglement mapping.

(b) Full entanglement mapping.

FIG. 4: Comparing multiple optimizations of TIM's 12 spins with the RealAmplitudes Ansatz.
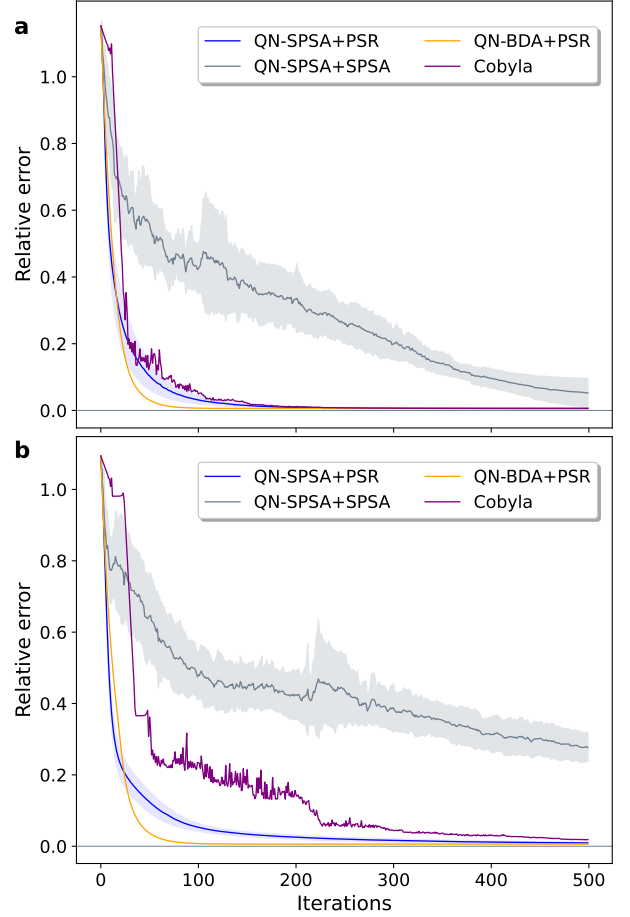
FIG. 5: Comparing multiple optimizations of TIM's 12 spins with a) the RealAmplitudes Ansatz and b) the EfficientSU2 Ansatz.

In Fig. 4, we use the RealAmplitudes ansatz, linear and full entanglement with 2 layers, to compare six different optimizations: SPSA, QN-BDA+PSR, FD, Cobyla, QN-SPSA+PSR, and QN-SPSA_SPSA. Noticeably, we observed that QN-SPSA+PSR outperforms Cobyla and Finite Difference methods. While the QN-BDA+PSR method did its job, as the theoretical prediction, showing the fastest convergence by direct use of the Fubini-
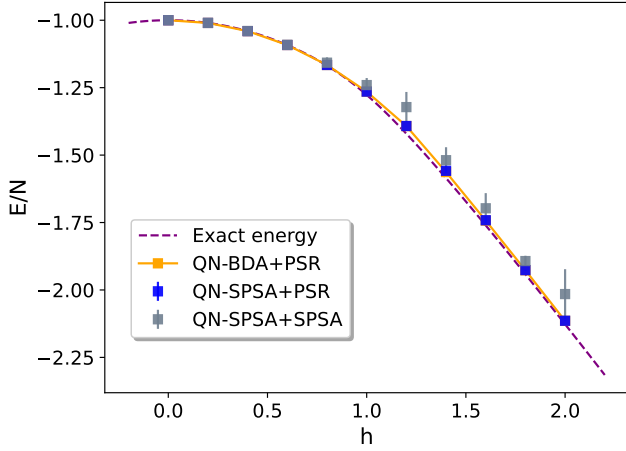
FIG. 6: Estimate the average ground state energy with QN-SPSA+SPSA, QN-SPSA+PSR, and QN-BDA+PSR for TIM of 12 spins and different external fields with the RealAmplitudes ansatz.
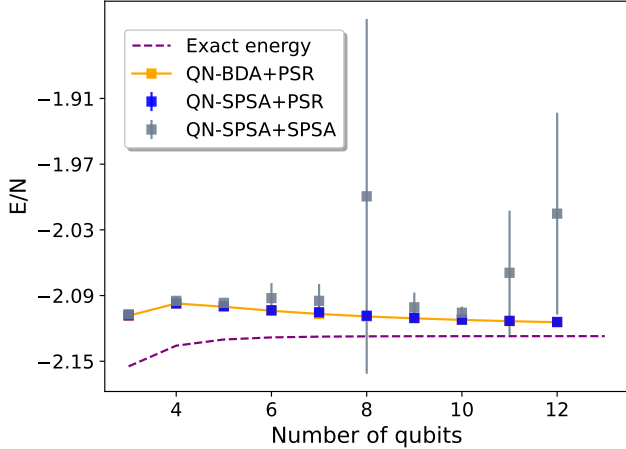

FIG. 7: Estimate the average ground state energy with QN-SPSA+SPSA, QN-SPSA+PSR, and QN-BDA+PSR for TIM of different numbers of spins with the RealAmplitudes ansatz.

Study metric tensor, QN-SPSA+PSR achieved comparable results with much lower computational overhead. By a small change in the gradient part from stochastic estimation to exact, we improved the performance of the inspired QN-SPSA+SPSA method, which made QN-SPSA+PSR more stable and faster convergence. Additionally, the classical derivative-free Cobyla exhibits a striking achievement, measuring up to QN-BDA+PSR and surpassing several algorithms while needing the least

quantum overhead of cost function evaluation, one per iteration. That makes it a considerable method to implement in the hybrid algorithm.

The robustness of the QN-SPSA+PSR method further manifests using the more complex EfficientSU2 ansatz, which introduces a highly intricate cost function, shown in Fig. 5. Although EfficientSU2 ansatz would create a more complex cost function, QN-SPSA+PSR still shows stable and fast convergence properties and achieves results comparable to QN-BDA+PSR optimization. Whereas, QN-SPSA+SPSA and Cobyla showed signs of stagnation and requires more iterations to approach optimal solutions.

Furthermore, Figs. 4 and 5, which compared two types of ansatz and two types of entanglement schemes, turned out to show that linear entanglement is not much different than the full entanglement, and the RealAmp performed better in convergence than Efficient SU2. That proved the initial arguments aforementioned in section III B 2, the RealAmplitude with linear entanglement, therefore, is good enough to survey the Ising model. Finally, the TIM energy estimates are obtained by varying external fields and numbers of spins, as illustrated in Figs. 6 and 7. In both cases, the QN-SPSA+PSR consistently delivered reliable and accurate results for the average ground state energy, closely aligning with the QN-BDA+PSR.

VI. CONCLUSION

In this study, the research delved into the review of optimization algorithms, comparing quantum and classical algorithms to provide a bird's eye view. Moreover, leveraging the model's properties and symmetries, the analysis of the Ising Model was conducted to conjecture an ansatz form.

The results of the running indicated that the chosen ansatz form, based on the Hamiltonian's properties, performed reasonably well, despite estimating the number of layers in Eq. 13 not working in this case. Nevertheless, it is believed to serve as a reasonably good initial estimation as the system scales up.

Lastly, our exploration of optimization methods has yielded exciting outcomes. The newly proposed QN-SPSA+PSR algorithm exhibited unexpected results of computational efficiency, as it demonstrates potential not only in reducing computational overhead, as it does not scale up significantly with qubits like traditional Quantum Natural methods, but also in achieving more accurate optimal solutions. This work suggests promising prospects for further exploration and application of such an algorithm for future research endeavors in VQAs and NISQ devices.

[1] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, Nature

Reviews Physics **3**, 625 (2021).

[2] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[3] 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 **574**, 505 (2019).

[4] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, *et al.*, Noisy intermediate-scale quantum (nisq) algorithms, arXiv preprint arXiv:2101.08448 (2021).

[5] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nature communications **5**, 1 (2014).

[6] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New Journal of Physics **18**, 023023 (2016).

[7] C. Feniou, O. Adjoua, B. Claudon, J. Zylberman, E. Giner, and J.-P. Piquemal, Sparse quantum state preparation for strongly correlated systems, The Journal of Physical Chemistry Letters **15**, 3197 (2024), pMID: 38483286, https://doi.org/10.1021/acs.jpclett.3c03159.

[8] C. Cao, J. Hu, W. Zhang, X. Xu, D. Chen, F. Yu, J. Li, H.-S. Hu, D. Lv, and M.-H. Yung, Progress toward larger molecular simulation on a quantum computer: Simulating a system with up to 28 qubits accelerated by point-group symmetry, Physical Review A **105**, 10.1103/physreva.105.062452 (2022).

[9] I. L. C. Michael A. Nielsen, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, New York, 2010).

[10] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, *et al.*, The variational quantum eigensolver: a review of methods and best practices, arXiv preprint arXiv:2111.05176 (2021).

[11] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, Hierarchical quantum classifiers, npj Quantum Information **4**, 10.1038/s41534-018-0116-9 (2018).

[12] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017), https://www.science.org/doi/pdf/10.1126/science.aag2302.

[13] X. Bonet-Monroig, R. Babbush, and T. E. O'Brien, Nearly optimal measurement scheduling for partial tomography of quantum states, Phys. Rev. X **10**, 031064 (2020).

[14] D. Wang, O. Higgott, and S. Brierley, Accelerated variational quantum eigensolver, Physical Review Letters **122**, 10.1103/physrevlett.122.140504 (2019).

[15] P. de Gennes, Collective motions of hydrogen bonds, Solid State Communications **1**, 132 (1963).

[16] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[17] M. J. D. Powell, Direct search algorithms for optimization calculations, Acta Numerica **7**, 287–336 (1998).

[18] M. J. Powell, A view of algorithms for optimization without derivatives, Mathematics Today-Bulletin of the Institute of Mathematics and its Applications **43**, 170 (2007).

[19] A. R. Conn, K. Scheinberg, and P. L. Toint, On the convergence of derivative-free methods for unconstrained optimization, Approximation theory and optimization: tributes to MJD Powell , 83 (1997).

[20] M. J. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, in *Advances in optimization and numerical analysis* (Springer, 1994) pp. 51–67.

[21] M. J. Powell, Uobyqa: unconstrained optimization by quadratic approximation, Mathematical Programming **92**, 555 (2002).

[22] M. J. Powell, The newuoa software for unconstrained optimization without derivatives, in *Large-scale nonlinear optimization* (Springer, 2006) pp. 255–297.

[23] M. J. Powell, The bobyqa algorithm for bound constrained optimization without derivatives, Cambridge NA Report NA2009/06, University of Cambridge, Cambridge **26** (2009).

[24] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, Johns Hopkins apl technical digest **19**, 482 (1998).

[25] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Physical Review A **99**, 10.1103/physreva.99.032331 (2019).

[26] L. Banchi and G. E. Crooks, Measuring Analytic Gradients of General Quantum Evolution with the Stochastic Parameter Shift Rule, Quantum **5**, 386 (2021).

[27] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, Quantum **6**, 677 (2022).

[28] R. Cheng, Quantum geometric tensor (fubini-study metric) in simple quantum system: A pedagogical introduction (2010).

[29] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).

[30] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Simultaneous perturbation stochastic approximation of the quantum fisher information, Quantum **5**, 567 (2021).

[31] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 012405 (2021).

[32] T. Haug and M. S. Kim, Optimal training of variational quantum algorithms without barren plateaus (2021), arXiv:2104.14543 [quant-ph].

[33] The code in this work can be accessed at electronic address: https://github.com/linhhacoV-I-P/Variational-Quantum-EigeinSolver.