Measurement optimization of variational quantum simulation by classical shadow and derandomization

Kouhei Nakaji[1,4], Suguru Endo[2], Yuichiro Matsuzaki[1], and Hideaki Hakoshima[3]

[1]Device Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST),1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan.

[2]NTT Computer and Data Science laboratories, NTT corporation, Musashino, Tokyo 180-8585, Japan

[3]Center for Quantum Information and Quantum Biology, Osaka University, 1-2 Machikaneyama, Toyonaka, Osaka 560-0043, Japan.

[4]Current address: Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

arXiv:2208.13934v3 [quant-ph] 20 Apr 2023

Simulating large quantum systems is the ultimate goal of quantum computing. Variational quantum simulation (VQS) gives us a tool to achieve the goal in near-term devices by distributing the computation load to both classical and quantum computers. However, as the size of the quantum system becomes large, the execution of VQS becomes more and more challenging. One of the most severe challenges is the drastic increase in the number of measurements; for example, the number of measurements tends to increase by the fourth power of the number of qubits in a quantum simulation with a chemical Hamiltonian. This work aims to optimize measurements in VQS by recently proposed measurement optimization techniques such as classical shadow and derandomization. Even though previous literature shows that the measurement optimization techniques successfully reduce measurements in the variational quantum optimization (VQO), how to apply them to VQS was unclear due to the gap between VQO and VQS in measuring observables. In this paper, we bridge the gap by changing the way of measuring observables in VQS and propose an algorithm to optimize measurements in VQS. Our numerical experiment shows the validity of using our algorithm with quantum chemical systems. Importantly, we theoretically reveal the advantage of using shadow-based strategies, e.g., classical shadow and derandomization, not only in VQS but also in VQO by calculating the variance of an observable. To the best of our knowledge, the advantage of using shadow-based strategies in VQO was given only numerically in previous research; therefore, this paper also gives significant implications for measurement optimization in general variational quantum algorithms.

1 Introduction

With the advent of noisy intermediate scale quantum (NISQ) computers [1], the next milestone is to realize their practical applications. A class of quantum algorithms called hybrid quantum-classical algorithms (see [2, 3]) came to attention because a load of computation can be distributed between quantum computers and classical computers; accordingly, the depth of the quantum circuit may be reduced. Among hybrid algorithms, variational quantum algorithms (VQAs) [4–9] have been extensively studied for diverse applications such as quantum computational chemistry [5, 6], quantum machine learning [10], quantum metrology [11, 12], etc. In VQAs, parametrized quantum states are generated from shallow quantum circuits, and the parameters are iteratively updated with the optimization using classical computers. VQAs can be broadly classified into two types: variational quantum optimization (VQO) [4–8] and variational quantum simulation (VQS) [9]. VQO computes and optimizes a problem-specific cost function; for example, in the famous variational quantum eigensolver (VQE) [5, 6], the energy of parametrized quantum states is evaluated by measuring Pauli operators, which constitutes the problem Hamiltonian in quantum computers, and the minimization of the energy results in the ground state and its energy. Meanwhile, VQS simulates the dynamics of the quantum system, e.g., quantum real and imaginary time evolution [9, 13]

and open quantum system dynamics [14] by minimizing the distance between the variationally simulated evolution and the exact evolution. The update rule of the parameters in VQS is derived from variational principles for time evolution, such as McLachlan's variational principle [15]. In both VQO and VQS, updates of the parameters are performed according to the output from quantum computers, more specifically, expectation values of observables. Since the number of circuit runs is still restricted in the current NISQ era, it is crucial to reduce the effect of shot noise and save the measurement counts for estimating each observable.

However, with the larger scale of the problem, reducing the effect of shot noise becomes more and more challenging. That is because the number of measurements to realize a given shot noise linearly increases with the number of the terms in the observable (here, we call 'term' by an observable evaluatable with a quantum circuit). For example, given the number of qubits as n, the number of terms scales as $O(n^4)$ in the chemical Hamiltonian, and therefore, the number of measurements also scales as $O(n^4)$. To mitigate the challenge, several measurement optimization methods are proposed in the context of VQO [8, 16–33]. In general, there are two types of measurement optimization methods: the grouping strategy [8, 16, 18, 22–33] and the shadow-based strategy [17–21]. Previous results numerically show that the applications of grouping strategies [26, 34] and shadow-based strategies [17, 19, 35] to VQE of specific problems achieve faster or better convergence in terms of the number of measurements though the performance of the optimization depends on which we choose an optimizer. Let us briefly review those two types of strategies in the following.

On the one hand, the grouping strategy is a method to efficiently estimate the cost function of VQO, such as a molecular Hamiltonian, by grouping commuting terms in the Hamiltonian [8, 16, 18, 21–33]. We can reduce the number of measurements by simultaneously measuring all the terms in each group. On the other hand, the shadow-based strategies construct a classical representation of a quantum state with a small number of measurements [17–21]. The classical shadow, which is the original shadow-based strategy, predicts M observables with $O(\log M)$ measurements up to the additive error; the information-theoretic arguments show that the order of $O(\log M)$ measurements is asymptotically optimal [17]. However, the classical shadow is inefficient for estimating particular observables since it typically contains unnecessary measurement outcomes not used for estimation. Previous studies propose variations of classical shadow to predict specific observables, such as many-body Hamiltonians, including the sum of Pauli terms [18–21]. One of its variations, derandomization, can efficiently predict such observables [19]. Derandomization is a deterministic protocol to minimize the confidence bound of Pauli observables starting from completely randomized Pauli measurements, and it typically achieves better performance than the grouping strategy and original classical shadow. These measurement protocols are experiment-friendly even for NISQ devices, and a photonic system realized an experiment of these protocols [36].

Even though the above measurement optimization strategies mainly target VQO, measurement optimization is arguably more important in VQS than in VQO. In VQO, the evolution of parameters in each step does not need to be correct as far as the optimal value of the observable is found. Meanwhile, in the case of VQS, we expect that the evolution of parameters in each step correctly describes the time evolution, and therefore, we need to efficiently reduce the shot noise by optimizing the measurements. Still, there is no previous research that applies the above measurement optimization methods to VQS. As far as our knowledge, that is because the above measurement optimization methods are mainly constructed for the problem where their observables are written by the sum of Pauli terms. In VQS, we need to evaluate the real/imaginary part of the expectation of operators [9, 13, 14], and cannot be straightforwardly written in the form of estimating the sum of Pauli observables. Approaches to optimize measurements in VQS are highly demanded.

The aim of this paper is to show the way of applying measurement optimization techniques to VQS. We particularly focus on shadow-based strategies. Our contribution also exists in clarifying the advantage of utilizing the shadow-based methods both in VQS and VQO from the theoretical aspect. As far as our knowledge, the advantage of VQO was shown only numerically in previous research. In this paper, we show the advantage by evaluating the variance of an observable. Let us specify the contribution of our work is summarized as follows:

- We successfully simplify the quantum circuit for VQS by a circuit synthesis, so that measurement optimization techniques (shadow-based strategies and grouping strategies) are applicable,
- We theoretically show the advantage of using the shadow-based strategy in the VQS by evaluating

the variance of the observable; the analysis is also applicable for assessing the merit of using the shadow-based strategy in VQO,

- We show the validity of our algorithm and the theoretical discussion by numerical demonstration. We show that the derandomization [19], one of the shadow-based strategies, greatly reduces the number of measurements compared to the original approach.

We note that even though we focus on the shadow-based methods in this paper, we can also utilize the grouping strategies for measurement optimization in VQS due to our circuit synthesis. Particularly, the recent result [31] shows that they can outperform derandomization in VQE by optimizing parameters in the grouping so that the variance of the observable is minimized. Also, in [32, 33], authors provide methods that may outperform the method in [31]. Thus, we expect better performance in VQS by using those grouping methods than shadow-based methods. It is still worth mentioning that those grouping methods need a reference quantum state for calculating the variance; the reference state is chosen to be a quantum state close to the ground state of the Hamiltonian in their work since they mainly target VQE. However, in most cases of VQS, our goal is not to find the ground state of the Hamiltonian, and we need to investigate the choice of a reference state. We leave combining our methods with those grouping strategies for future work.

The rest of the paper is organized as follows. In Section 2, we give a brief review of VQS. We also state the challenge of VQS in the same section. We review the shadow-based strategy in Section 3; particularly, we discuss the classical shadow [17] and the derandomization [19] in detail. Section 4 is dedicated to proposing our algorithm that applies the shadow-based strategy to VQS. The advantage of utilizing the shadow-based strategy is discussed in the same section. In Section 5, we numerically demonstrate the performance of our algorithm. Finally, we conclude our paper with some discussions in Section 6.

2 Variational quantum simulation without measurements optimization

In this section, we review the variational quantum simulation (VQS) proposed and studied in the previous literature [9, 13, 14] and state a challenge of the algorithm. In Section 2.1, we give a general description of the VQS based on [37]. In Section 2.2, we show the variational real and imaginary time evolution as examples. Section 2.3 is dedicated to stating a problem of the previous approach in the VQS.

2.1 General processes

We define the quantum time evolution for a state $|v(t)\rangle$ in this paper as the following differential equation,

$$B(t)\frac{d}{dt}|v(t)\rangle = A(t)|v(t)\rangle . \tag{2.1}$$

where $A(t)$ and $B(t)$ denote matrices. In the VQS, we parameterize $|v(t)\rangle$ as $\left|v(\vec{\theta}(t))\right\rangle$ by using a set of parameters. From Mclachlan's principle [15],

$$\delta \left|\left| B(t)\sum_{k=1}^{N_P}\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_k}\dot{\theta}_k - A(t)\left|v(\vec{\theta}(t))\right\rangle \right|\right| = 0, \tag{2.2}$$

where $|||\psi\rangle|| \equiv \langle\psi|\psi\rangle$, N_P is the number of parameters, and

$$\delta f \equiv \sum_{j=1}^{N_P}\frac{\partial f}{\partial\dot{\theta}_j}\delta\dot{\theta}_j. \tag{2.3}$$

We note that there are two other conventional variational principles than the Mclachlan's principle [15]: Dirac and Frenkel's variational principles [38, 39] and the time-dependent variational principle [40, 41]. A detailed comparison of different variational principals is made in [14]. The Dirac and Frenkel variational principle is not suitable for VQS since the equation of the parameters may involve complex solutions,

which contradicts the requirement that parameters are real. Even though the time-dependent variational principle gives a real solution, the time-dependent variational principle is shown to be more unstable and not applicable to the evolution of density matrices and imaginary time evolution. In contrast, McLachlan's principle generally produces stable solutions, and it is also applicable to all VQS problems.

Equation (2.2) is equivalent to

$$\sum_{\ell=1}^{N_P}\left[\frac{\partial\left\langle v(\vec{\theta}(t))\right|}{\partial\theta_k}B^\dagger(t)B(t)\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_\ell}+\frac{\partial\left\langle v(\vec{\theta}(t))\right|}{\partial\theta_\ell}B^\dagger(t)B(t)\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_k}\right]\dot{\theta}_\ell$$
$$-\left[\frac{\partial\left\langle v(\vec{\theta}(t))\right|}{\partial\theta_k}B^\dagger(t)A(t)\left|v(\vec{\theta}(t))\right\rangle+\left\langle v(\vec{\theta}(t))\right|A^\dagger(t)B(t)\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_k}\right]=0. \tag{2.4}$$

for all k. We can further simplify (2.4) as

$$\sum_{\ell=1}^{N_P}M_{k\ell}\dot{\theta}_\ell=V_k, \tag{2.5}$$

where

$$M_{k\ell}=\Re\left[\frac{\partial\left\langle v(\vec{\theta}(t))\right|}{\partial\theta_k}B^\dagger(t)B(t)\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_\ell}\right], \tag{2.6}$$

$$V_k=\Re\left[\frac{\partial\left\langle v(\vec{\theta}(t))\right|}{\partial\theta_k}B^\dagger(t)A(t)\left|v(\vec{\theta}(t))\right\rangle,\right] \tag{2.7}$$

with $\Re[w]$ as the real part of the complex number w.

Once we compute $M_{k\ell}$ and V_k, we can update the parameter by the Euler method with a small parameter δt as

$$\vec{\theta}(t+\delta t)=\vec{\theta}(t)+M^{-1}V\cdot\delta t, \tag{2.8}$$

where M (V) is the matrix (vector) representation of $M_{k\ell}$ (V_k). Next, we show how to compute $M_{k\ell}$ and V_k in the following.

Estimation of $M_{k\ell}$ and V_k using quantum circuits

Let R be the unitary operator corresponding to the parameterized quantum circuit (PQC); namely

$$\left|v(\vec{\theta}(t))\right\rangle=R|v_{\text{ref}}\rangle, \tag{2.9}$$

with $|v_{\text{ref}}\rangle$ as a reference state. We often call R the *ansatz*. We expand R as

$$R\equiv R_{N_P}(\theta_{N_P})R_{N_P-1}(\theta_{N_P-1})\cdots R_1(\theta_1), \tag{2.10}$$

where R_k is the k-th unitary operator applied to the reference state and θ_k is the parameter of R_k. Let us describe

$$\frac{\partial R_k(\theta_k)}{\partial\theta_k}=\sum_{i=1}^{N_g}g_{k,i}R_k(\theta_k)P_{k,i}, \tag{2.11}$$

where N_g is the number of Pauli observables when expanding the derivative, $g_{k,i}$ is a coefficient, and $P_{k,i}$ is a tensor product of Pauli operators and identity: $P_{k,i}\in\{X,Y,Z,I\}^{\otimes n}$. Then

$$\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_k}=\sum_{i=1}^{N_g}g_{k,i}R_{k,i}|v_{\text{ref}}\rangle, \tag{2.12}$$

where

$$R_{k,i} = R_{N_P}(\theta_{N_P}) R_{N_P-1}(\theta_{N_P-1}) \cdots R_k(\theta_k) P_{k,i} R_{k-1}(\theta_{k-1}) \cdots R_1(\theta_1). \tag{2.13}$$

In typical settings, $N_g = O(1)$ is satisfied. For example, if we embed each parameter as $\exp(-i\theta_k P/2)$, where P denotes a tensor products of Pauli operators, $N_g = 1$ and $g_{k,1} = -i/2$. Also, let us describe

$$B^\dagger(t)B(t) = \sum_{q=1}^{N_{\mathrm{BB}}} \beta_q P_q, \tag{2.14}$$

$$B^\dagger(t)A(t) = \sum_{r=1}^{N_{\mathrm{BA}}} \alpha_r P_r, \tag{2.15}$$

where P_q and P_r are tensor products of Pauli operators and identity, and where N_{BB} and N_{BA} are the numbers of terms when decomposing $B^\dagger(t)B(t)$ and $B^\dagger(t)A(t)$ as the sums of Pauli operators.

By substituting (2.12) and (2.14) into (2.6), we obtain

$$M_{k\ell} = \sum_{i,i'}^{N_g} \sum_{q=1}^{N_{\mathrm{BB}}} \Re \left[g^*_{k,i} g_{\ell,i'} \beta_q \langle v_{\mathrm{ref}}| R^\dagger_{k,i} P_q R_{\ell,i'} |v_{\mathrm{ref}}\rangle \right]. \tag{2.16}$$

Similarly, by substituting (2.12) and (2.15) into (2.7), we obtain

$$V_k = \sum_{i=1}^{N_g} \sum_{r=1}^{N_{\mathrm{BA}}} \Re \left[g^*_{k,i} \alpha_r \langle v_{\mathrm{ref}}| R^\dagger_{k,i} P_r R |v_{\mathrm{ref}}\rangle \right]. \tag{2.17}$$

For computing (2.16), we need to evaluate each $\Re \left[e^{i\phi} \langle v_{\mathrm{ref}}| R^\dagger_{k,i} P_q R_{\ell,i'} |v_{\mathrm{ref}}\rangle \right]$ with $e^{i\phi}$ as a phase, which is possible by using the Hadamard test. In Fig. 1, we show the quantum circuit to compute the value for $k \leq \ell$. Note that since $M_{k\ell} = M_{\ell k}$, we do not need to consider the case when $k > \ell$. To compute (2.17), we need to evaluate $\Re \left[e^{i\phi} \langle v_{\mathrm{ref}}| R^\dagger_{k,i} P_r R |v_{\mathrm{ref}}\rangle \right]$, which is possible by using the quantum circuit in Fig. 2.

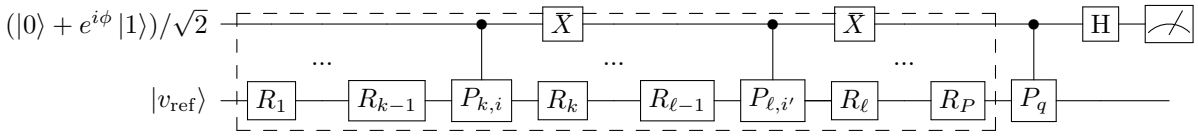


Figure 1: A quantum circuit to compute $\Re \left[e^{i\phi} \langle v_{\mathrm{ref}}| R^\dagger_{k,i} P_q R_{\ell,i'} |v_{\mathrm{ref}}\rangle \right]$ for $k \leq \ell$. We use the circuit enclosed by the dotted rectangle in Section 4. We note that the output quantum state after applying the gates in the dotted box is $|\psi^\phi_{k,i;\ell,i'}\rangle = 1/\sqrt{2} \left(R_{\ell,i'}|0\rangle|v_{\mathrm{ref}}\rangle + e^{i\phi} R_{k,i}|1\rangle|v_{\mathrm{ref}}\rangle \right)$.

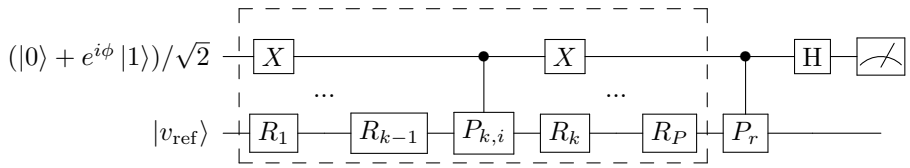


Figure 2: A quantum circuit to compute $\Re \left[e^{i\phi} \langle v_{\mathrm{ref}}| R^\dagger_{k,i} P_r R |v_{\mathrm{ref}}\rangle \right]$. As in Figure 1, we use the circuit enclosed by the dotted rectangle in Section 4. We note that the output quantum state after applying the gates in the dotted box is $|\psi^\phi_{k,i}\rangle = 1/\sqrt{2}(|0\rangle R_{k,i}|v_{\mathrm{ref}}\rangle + |1\rangle R|v_{\mathrm{ref}}\rangle)$.

2.2 Variational real and imaginary time evolution

Here we show $M_{k\ell}$ and V_k in real-time evolution (RTE) and imaginary time evolution (ITE) as examples of the VQS. The real-time evolution is an algorithm to simulate the Schrödinger equation under a time-evolution Hamiltonian H:

$$\frac{\partial|v(t)\rangle}{\partial t} = -iH|v(t)\rangle. \tag{2.18}$$

Thus, $A(t) = -iH$ and $B(t) = 1$ and we obtain

$$M_{k\ell} = \sum_{i,i'}^{N_g} \Re\left[g^*_{k,i} g_{\ell,i'} \langle v_{\rm ref}|R^\dagger_{k,i} R_{\ell,i'}|v_{\rm ref}\rangle\right], \tag{2.19}$$

$$V_k = \sum_{i=1}^{N_g}\sum_{r=1}^{N_{\rm BA}} \Re\left[-ig^*_{k,i}\alpha_r \langle v_{\rm ref}| R^\dagger_{k,i} P_r R |v_{\rm ref}\rangle\right], \tag{2.20}$$

where we expand the Hamiltonian as $H = \sum_{j=1}^{N_{\rm BA}} \alpha_r P_r$ with α_r as a real coefficient. Note that we see $N_{\rm BB} = 1$ from (2.19).

The imaginary time evolution is the operation obtained by replacing it to τ and H to $H - \langle H\rangle$ in (2.18), where $\langle H\rangle = \langle v(t)|H|v(t)\rangle$. Namely,

$$\frac{\partial\,|v(\tau)\rangle}{\partial\tau} = -(H - \langle H\rangle)\,|v(\tau)\rangle\,. \tag{2.21}$$

Note that the term $\langle H\rangle$ is added in order to preserve the norm of the quantum state. Importantly, we can show that, in a large τ limit, $|v(\tau)\rangle$ becomes the ground state of the Hamiltonian H. Thus, the imaginary time evolution is often used as a way of searching for the ground state. The functions $A(t)$ and $B(t)$ are given by $A(t) = -(H - \langle H\rangle)$ and $B(t) = 1$. The term $\langle H\rangle$ does not affect the computation of V_k, which can be checked by substituting the values of $A(t)$ and $B(t)$ to (2.7) as

$$V_k = \Re\left[\left\langle v(\vec{\theta}(t))\right| - (H - \langle H\rangle)\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_k}\right], \tag{2.22}$$

$$= -\Re\left[\left\langle v(\vec{\theta}(t))\right| H\frac{\partial\left|v(\vec{\theta}(t))\right\rangle}{\partial\theta_k}\right] + \frac{1}{2}\langle H\rangle\frac{\partial\left(\left\langle v(\vec{\theta}(t))\right| v(\vec{\theta}(t))\right\rangle\right)}{\partial t}. \tag{2.23}$$

The second term, which corresponds to $\langle H\rangle$, vanishes since the norm of $\left|v(\vec{\theta}(t))\right\rangle$ preserves under the transformation of $\vec{\theta}(t)$, and therefore, we obtain $M_{k\ell}$ and V_k by setting $B(t) = 1$ and $A(t) = -H$ when computing (2.16) and (2.17). Thus, in the variational imaginary time evolution, we obtain

$$M_{k\ell} = \sum_{i,i'}^{N_g} \Re\left[g^*_{k,i} g_{\ell,i'} \langle v_{\rm ref}|R^\dagger_{k,i} R_{\ell,i'}|v_{\rm ref}\rangle\right], \tag{2.24}$$

$$V_k = -\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\rm BA}} \Re\left[g^*_{k,i}\alpha_r \langle v_{\rm ref}| R^\dagger_{k,i} P_r R |v_{\rm ref}\rangle\right]. \tag{2.25}$$

The circuit used for evaluating each term of V_k in real and imaginary time evolution is Fig. 2, which we introduced in Section 2.1. On the other hand, a simpler circuit is used for evaluating $M_{k\ell}$ in real and imaginary time evolution; we show the circuit in Fig. 3. We can readily check that the output of the circuit returns $\Re\left[e^{i\phi}\langle v_{\rm ref}| R^\dagger_{k,i} R_{\ell,i'} |v_{\rm ref}\rangle\right]$.

2.3 Problems of the current approach

The approach mentioned above to performing the VQS is challenging to implement because the number of measurements linearly increases as the number of terms denoted by $N_{\rm BB}$ and $N_{\rm BA}$ increases. That

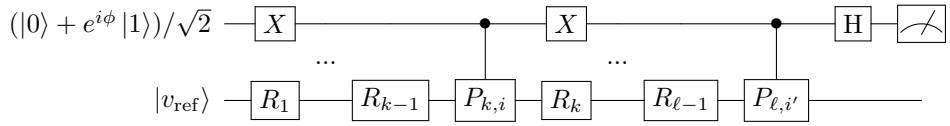


Figure 3: A quantum circuit to compute $\Re\left[e^{i\phi}\langle v_{\mathrm{ref}}|R_{k,i}^{\dagger}R_{\ell,i'}|v_{\mathrm{ref}}\rangle\right]$ for $k \leq \ell$.

is because we need to evaluate the circuit corresponding to each term inside the summations of (2.16) and (2.17) one by one (the number of measurements also linearly increases according to the value of N_g, but it is not problematic since N_g is typically taken to be small). For example, in RTE and ITE, the number of measurements for evaluating V_k linearly scales with the number of terms in H, which could be impractical when H has many terms. Suppose that we use a molecular Hamiltonian as the time-evolution Hamiltonian. In that case, the number of terms in the Hamiltonian is $O(n^4)$ with n as the number of qubits, and therefore, the number of measurements also increases in proportion to $O(n^4)$.

The challenge of a large number of Pauli observables in Hamiltonian is also addressed in the variational quantum eigensolver (VQE), and many solutions have been proposed [6, 16–22, 26, 27, 30, 42–47]. However, to our best knowledge, no previous work has explored this issue in the context of the VQS. Importantly, there is a crucial difference between VQS and VQE; namely, to evaluate each term in the summation in (2.14) and (2.15), we need the controlled operations associated with each Pauli observable to evaluate each term in the summation of (2.14) and (2.15) in VQS, while the VQE only requires estimating the expectation value of each Pauli observable. Thus, it is not straightforward to apply the methods discussed in [6, 16–22, 26, 27, 30, 42–47] to VQS. In the rest of the paper, we show a way to overcome the difficulties and discuss how to apply the proposed methods for the VQE to the VQS. Before discussing the main point, we will review the proposed methods for the VQE in the next section.

It should be noted that in the case of ITE, the parameter shift-rule [48] is applicable to calculate V_k under a suitable condition (e.g., $N_g = 1$ and $g_{k,1} \neq 0$ is real), meaning that the evaluation of V_k boils down to the evaluation of the expectation value of $H = \sum_r \alpha_r P_r$ with respect to a quantum state generated by a PQC. In that case, we can reduce the number of measurements by utilizing the techniques proposed in VQE. Still, in most of the applications of VQS, the parameter-shift rule is not applicable; therefore, we need to construct a measurement optimization procedure tailored for VQS.

3 Measurements optimization for sum of Pauli observables

In this section, we review strategies to reduce the number of measurements in the VQE, when the Hamiltonian has many terms such as $\mathcal{H} = \sum_{j=1}^{K} a_j P_j$ for $K \gg 1$. In the VQE, we need to estimate the value $\langle \mathcal{H} \rangle = \sum_{j=1}^{K} a_j \mathrm{Tr}(P_j \rho)$ where $\{a_j\}$ are real coefficients, $\{P_j\}$ are Pauli observables ($P_j \in \{X, Y, Z, I\}^{\otimes n}$), and ρ is a n-qubit quantum state generated by a quantum circuit. When K increases, the number of measurements to achieve a given precision linearly increases as far as we naively evaluate each term.

For improving the scaling of the number of measurements, several solutions have been proposed [6, 16–22, 26, 27, 30, 42–47]. The most promising solutions are arguably shadow-based strategies [17–21], which we review in this section. In Section 3.1, we first give a general discussion about shadow-based strategies [17–21]. Next, in Section 3.2, we show the detail of some of the shadow-based strategies that we will examine in our numerical experiment in Section 5.

3.1 General discussion for shadow-based strategies

Shadow-based strategies for estimating $\mathcal{H}$ [17–21] are mostly composed of three steps. The first step of the estimation process takes the number of measurements N_{shot}, $\{a_j\}_{j=1}^{K}$ and $\{P_j\}_{j=1}^{K}$ as its inputs. Let us denote a measurement basis by one of the Pauli observables $\{X, Y, Z\}^{\otimes n}$, where each measurement is performed on the basis where the Pauli observable is diagonalized. For example, when $n = 3$, the measurement basis can be expressed as XZX, meaning that the quantum state ρ is measured in the computational

basis after we operate the Hadamard gates to the first and the third qubits. The first step returns the array of measurement basis as $\{M_r\}_{r=1}^{N_{\rm shot}}$; how to select those basis depends on the algorithm. We define the function corresponding to the first step as **buildMeasurements**$(\{a_j\},\{P_j\},N_{\rm shot})$. The second step takes $\{M_r\}_{r=1}^{N_{\rm shot}}$ and ρ as its inputs and returns the measurement results of ρ as $\{b_r\}_{r=1}^{N_{\rm shot}}$, where $b_r \in \{1,-1\}^{\otimes n}$. We denote the function corresponding to the second step as **getMeasurementResults**$(\{M_r\}_{r=1}^{N_{\rm shot}},\rho)$. The third step takes $\{M_r\}_{r=1}^{N_{\rm shot}}$, $\{b_r\}_{r=1}^{N_{\rm shot}}$, $\{P_j\}_{j=1}^{K}$, and $\{a_j\}_{j=1}^{K}$ as its inputs, and returns the following ν:

$$\nu = \frac{1}{N_{\rm shot}} \sum_{r=1}^{N_{\rm shot}} \nu_r, \tag{3.1}$$

where

$$\nu_r = \sum_{j=1}^{K} a_j \frac{1}{q(P_j)} f(P_j, M_r)\mu(P_j, b_r). \tag{3.2}$$

Here, $f(P,M)$ denotes a covering function, $q(P_j)$ denotes a covering probability function, and $\mu(P,b)$ denotes an estimation function. These are defined so that ν should be an unbiased estimator of $\mathrm{Tr}(\mathcal{H}\rho)$. We show the detailed definitions as follows.

Covering function $f(P,M)$

The covering function $f(P,M)$ takes a Pauli obserable P and a measurement basis M as its inputs, and returns outputs as follows

$$f(P,M) = \prod_{i=1}^{n} f_{\rm local}(P[i],M[i]), \tag{3.3}$$

$$f_{\rm local}(P[i],M[i]) = \begin{cases} 1 & \text{if } P[i]=M[i] \text{ or } P[i]=I \\ 0 & \text{otherwise} \end{cases}, \tag{3.4}$$

where $P[i]$ and $M[i]$ is the Pauli observable corresponds to the i-th qubit of P and M. For example, $f(XYI,XYX)=1$ and $f(ZIX,ZXY)=0$. We call M *covers* P if and only if $f(P,M)=1$ as in [19].

Covering probability function

We define the covering probability function $q(P)$ depending on whether we generate the measurement basis probabilistically or not. When they are probabilistically generated according to a probability distribution $q_{\rm m}(M)$, $q(P)$ is the probability that the Pauli observable P is covered, namely

$$q(P) = \sum_{M\in\{X,Y,Z\}^{\otimes n}} q_{\rm m}(M) f(P,M). \tag{3.5}$$

When measurements are chosen deterministically, we define the covering probability function by the ratio that P is covered when measurements are $\{M_{r'}\}_{r'}^{N_{\rm shot}}$, namely,

$$q(P) = \frac{1}{N_{\rm shot}} \sum_{r'=1}^{N_{\rm shot}} f(P,M_{r'}). \tag{3.6}$$

Note that in this deterministic case, $q(P)$ depends on the chosen measurement basis $\{M_{r'}\}_{r'=1}^{N_{\rm shot}}$, but we omit the corresponding index from $q(P)$ for simplicity.

Estimation function

The second function is the estimation function $\mu(P, b)$ that takes the Pauli observable and the bit array $b \in \{1, -1\}^{\otimes n}$ as its inputs and operates as follows:

$$\mu(P, b) = \prod_{i=1}^{n} \mu_{\text{local}}(P[i], b[i]), \tag{3.7}$$

$$\mu_{\text{local}}(P[i], b[i]) = \begin{cases} 1 & P[i] = I \text{ or } b[i] = 1 \\ -1 & \text{otherwise} \end{cases}, \tag{3.8}$$

where $b[i]$ is the i-th bit value. Let $p(\rho, M; b)$ be the probability that n-bit bitarray b is obtained as a result of measuring ρ in a basis M covering P. Then, we can readily show

$$\sum_{b} p(\rho, M; b)\mu(P, b) = \text{Tr}(P\rho). \tag{3.9}$$

As we show in Appendix A, ν becomes an unbiased estimator of $\text{Tr}(\mathcal{H}\rho)$, i.e., $\mathbf{E}[\nu] = \text{Tr}(\mathcal{H}\rho)$. Also, the variance of ν is given by

$$\text{Var}(\nu) = \mathbf{E}[\nu^2] - (\mathbf{E}[\nu])^2 = \frac{1}{N_{\text{shot}}} \left[\sum_{j=1}^{K} \frac{a_j^2}{q(P_j)} + \sum_{j=1}^{K} \sum_{k \neq j} a_j a_\ell g(P_j, P_\ell) \text{Tr}(P_j P_\ell \rho) - [\text{Tr}(\mathcal{H}\rho)]^2 \right], \tag{3.10}$$

where if M_r is sampled from $q_{\text{m}}(M_r)$,

$$g(P_j, P_\ell) = \sum_{M_r \in \{X,Y,Z\}^{\otimes n}} \frac{f(P_j, M_r)}{q(P_j)} \frac{f(P_\ell, M_r)}{q(P_\ell)} q_{\text{m}}(M_r), \tag{3.11}$$

and if $\{M_r\}_{r=1}^{N_{\text{shot}}}$ is deterministically chosen,

$$g(P_j, P_\ell) = \frac{1}{N_{\text{shot}}} \sum_{r=1}^{N_{\text{shot}}} \frac{f(P_j, M_r)}{q(P_j)} \frac{f(P_\ell, M_r)}{q(P_\ell)}. \tag{3.12}$$

We write the three steps of obtaining ν as **estimateNu**$(\rho, \{a_j\}_{j=1}^K, \{P_j\}_{j=1}^K, N_{\text{shot}})$ and summarize it in **Algorithm 1**. It should be noted that the variance $\text{Var}(\nu)$ determines the number of measurements N_{shot} to achieve a given precision ϵ; namely, N_{shot} scales as $N_{\text{shot}} = O(\text{Var}(\nu)/\epsilon^2)$.

Algorithm 1 estimateNu$(\rho, \{a_j\}, \{P_j\}, N_{\text{shot}})$

Input: $\rho, \{a_j\}, \{P_j\}, N_{\text{shot}}$

1: $\{M_r\}_{r=1}^{N_{\text{shot}}} =$ **buildMeasurements**$(\{a_j\}, \{P_j\}, N_{\text{shot}})$

2: Set $\{b_r\}_{r=1}^{N_{\text{shot}}} =$ **getMeasurementResults**$(\{M_r\}, \rho)$

3: **for** $r = 1...N_{\text{shot}}$ **do**

4:

$$\nu_r = \sum_{j=1}^{K} a_j \frac{1}{q(P_j)} f(P_j, M_r)\mu(P_j, b_r).$$

5: **end for**

return $\nu = \frac{1}{N_{\text{shot}}} \sum_{r=1}^{N_{\text{shot}}} \nu_r$.

3.2 Examples of shadow-based strategies

The shadow-based strategy is originally proposed by the literature [17] and various methods motivated by the technique have been proposed [17–21], which mainly target to reduce the value of $\text{Var}(\nu)$ in (3.10). Among them, we review the classical shadow and the derandomization in the following discussion.

Algorithm 2 buildMeasurements$(\{a_j\}, \{P_j\}, N_{\rm shot})$ in Classical shadow

Input: $\{a_j\}, \{P_j\}, N_{\rm shot}$
SET n as the number of qubits of the operators $\{P_j\}$.
SET results $= []$.
for $r = 1...N_{\rm shot}$ **do**
  SET $M = []$.
  for $i = 1...n$ **do**
    Choose one of $\{X, Y, Z\}$ with equal probability and set it to $M[i]$.
  end for
  SET results$[r]$ $= M$.
end for
return results

Classical shadow

In the classical shadow, we randomly choose the measurement basis. Namely, each $M_r[i]$ (i-th Pauli matrix in r-th measurement basis) is chosen from $\{X, Y, Z\}$ with equal probability, i.e., $q_{\rm m}(M) = 1/3^n$. We show the pseudo-code for **buildMeasurements**$(N_{\rm shot})$ in classical shadow in Algorithm 2. We can readily show that the covering probability is given by

$$q(P_j) = \sum_{M \in \{X,Y,Z\}^{\otimes n}} \frac{f(P_j, M)}{3^n} = \frac{1}{3^{\text{locality}(P_j)}}, \tag{3.13}$$

where locality(O) is the number of non-identity Pauli matrix in the Pauli observable O. As we show in Appendix B, we obtain

$$g(P_j, P_\ell) = g_{\rm cs}(P_j, P_\ell) \equiv \prod_{i=1}^{n} g_{\rm local}(P_j[i], P_\ell[i]), \tag{3.14}$$

where

$$g_{\rm local}(P_j[i], P_\ell[i]) = \begin{cases} 1 & P_j[i] = I \text{ or } P_\ell[i] = I \\ 3 & P_j[i] = P_\ell[i] \neq I \\ 0 & otherwise \end{cases}. \tag{3.15}$$

Therefore,

$$\text{Var}(\nu) = \frac{1}{N_{\rm shot}} \left[\sum_{j=1}^{K} a_j^2 3^{\text{locality}(P_j)} + \sum_{j \neq \ell} a_j a_\ell g_{\rm cs}(P_j, P_\ell) \text{Tr}(P_j P_\ell \rho) - [\text{Tr}(\mathcal{H}\rho)]^2 \right]. \tag{3.16}$$

Derandomization

As we see in (3.10), the value of $\text{Var}(\nu)$ tends to be small if $q(P_j)$ is large. From this point of view, the random choice of measurements in the classical shadow is not always efficient; particularly, when the locality of a Pauli term P_j is large, the value of $q(P_j)$ is exponentially suppressed, as we see in (3.13). Thus, strategies to choose the measurement basis set that more efficiently cover Pauli observables are proposed as consecutive works [18–21].

One of such strategies is the derandomization [19]. Unlike the classical shadow, a pre-defined measurement basis set is used in the derandomization (thus, it corresponds to the deterministic case). For obtaining the pre-defined measurement basis $\{M_r\}_{r=1}^{N_{\rm shot}}$, the procedure of the derandomization iteratively determines them in the following order:

$$M_1[1], M_1[2], \cdots M_1[n], M_2[1], M_2[2], \cdots M_2[n], \cdots M_{N_{\rm shot}}[1], M_{N_{\rm shot}}[2], \cdots M_{N_{\rm shot}}[n]. \tag{3.17}$$

Suppose that $\{M_r\}_{r=1}^{r_0}$ and $\{M_{r_0+1}[i]\}_{i=1}^{i_0}$ are already determined and let $\mathbf{M}$ be the concatinate of $\{M_r\}_{r=1}^{r_0}$ and $\{M_{r_0+1}[i]\}_{i=1}^{i_0}$. Then, $M_{r_0+1}[i_0+1] \in \{X, Y, Z\}$ is determined so that a score function $C(\mathbf{M}, M_{r_0+1}[i_0+$

1]) is maximized (here we assume $i_0 < n$ for simplicity, but we can give the same argument by replacing $M_{r_0+1}[i_0+1]$ with $M_{r_0+2}[1]$ even when $i_0 = n$).

The score function C is constructed so that the increase of C leads to the increase of $\{q(P_j)\}_{j=1}^K$ in our notation. Let

$$V(P_j, \mathbf{M}, M_{r_0+1}[i_0+1]) = \frac{\eta}{2}\sum_{r=1}^{r_0} f(P_j, M_r) - \log\left(1 - \beta_{i_0+1}\prod_{k=1}^{i_0+1} f_{\mathrm{local}}(P_j[k], M_{r_0+1}[k])\right) \tag{3.18}$$

with

$$\beta_{i_0+1} = \gamma \prod_{j=i_0+2}^{K} \frac{1}{3^{\mathrm{locality}(P_j)}}, \tag{3.19}$$

where γ, η are hyperparameters. In [19], they give some options for the choice of the score function C; among them, the one used in their numerical experiment is

$$C(\mathbf{M}, M_{r_0+1}[i_0+1]) = \sum_{j=1}^{K} \exp\left(-V\left(P_j, \mathbf{M}, M_{r_0+1}[i_0+1]\right)/w_j\right), \tag{3.20}$$

where

$$w_j = \frac{|a_j|}{\max_j |a_j|}. \tag{3.21}$$

For example, suppose that the observable is

$$H = XXXZ + XXII + IIXZ + YYZX + YYII + IIZX \tag{3.22}$$

and $N_{\mathrm{shot}} = 10$. By minimizing (3.20) step by step, $\{M_r\}_{r=1}^{10}$ includes five $XXXZ$ and $YYZX$, which leads to

$$\frac{1}{q(P_r)} = 2. \tag{3.23}$$

In the classical shadow,

$$\frac{1}{q(P_r)} \geq 3^2. \tag{3.24}$$

Thus, we see that the covering probability improves in the derandomization compared to the classical shadow.

4 Measurements optimization in the VQS

In this section, we propose the VQS with shadow-based strategies reviewed in the previous section. In Section 4.1, we synthesize the circuits in Fig. 1 and Fig. 2 and remove unnecessary two-qubit operations. Based on the result of the circuit synthesis, we propose an algorithm applying shadow-based strategies to the VQS in Section 4.2. In Section 4.3, we discuss the advantage of using our proposed algorithm. Note that we generally discuss our proposed algorithm in this section and demonstrate its performance by using RTE and ITE in the numerical demonstration in Section 5. Section 4.4 is dedicated to apply our implication obtained in Section 4.3 to the general shadow-based strategies discussed in Section 3

4.1 Circuit synthesis

Let us define $\rho^{\phi}_{k,i;\ell,i'}$ as the quantum state generated by operating the partial circuits enclosed by the dotted rectangle in Fig. 1 to the reference state: $(|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2} \otimes |v_{\mathrm{ref}}\rangle$. The specific form of $\rho^{\phi}_{k,i;\ell,i'}$ is given by

$$\rho^{\phi}_{k,i;\ell,i'} = |\psi^{\phi}_{k,i,\ell;i'}\rangle\langle\psi^{\phi}_{k,i;\ell,i'}|, \tag{4.1}$$

$$|\psi^{\phi}_{k,i;\ell,i'}\rangle = \frac{1}{\sqrt{2}}\left(R_{\ell,i'}|0\rangle|v_{\mathrm{ref}}\rangle + e^{i\phi}R_{k,i}|1\rangle|v_{\mathrm{ref}}\rangle\right). \tag{4.2}$$

Let CP_q also be the controlled-P_q operation and I_n be the n-qubit identity operator. Then, for each q, k, and ℓ, the output of the circuit shown in Fig. 1 has the following form,

$$\Re\left[e^{i\phi}\left\langle v_{\mathrm{ref}}\right| R_{k,i}^{\dagger} P_q R_{\ell,i'}\left|v_{\mathrm{ref}}\right\rangle\right] = \mathrm{Tr}\left[(X\otimes I_n)(\mathrm{CP}_q)\rho_{k,i;\ell,i'}^{\phi}(\mathrm{CP}_q^{\dagger})\right] \tag{4.3}$$

$$= \mathrm{Tr}\left(O_q \rho_{k,i;\ell,i'}^{\phi}\right), \tag{4.4}$$

with

$$O_q \equiv (\mathrm{CP}_q^{\dagger})(X\otimes I_n)(\mathrm{CP}_q). \tag{4.5}$$

Similarly, let $\rho_{k,i}^{\phi}$ be the quantum state generated by operating the partial circuits enclosed by the dotted rectangle in Fig. 2 to the reference state. The specific form of $\rho_{k,i}^{\phi}$ is given by

$$\rho_{k,i}^{\phi} = |\psi_{k,i}^{\phi}\rangle\langle\psi_{k,i}^{\phi}| \tag{4.6}$$

$$|\psi_{k,i}^{\phi}\rangle = \frac{1}{\sqrt{2}}(|0\rangle R_{k,i}|v_{\mathrm{ref}}\rangle + |1\rangle R|v_{\mathrm{ref}}\rangle). \tag{4.7}$$

The output of the circuit shown in Fig. 2 has the following form,

$$\Re\left[e^{i\phi}\left\langle v_{\mathrm{ref}}\right| R_{k,i}^{\dagger} P_r R\left|v_{\mathrm{ref}}\right\rangle\right] = \mathrm{Tr}\left[(X\otimes I_n)(\mathrm{CP}_r)\rho_{k,i}^{\phi}(\mathrm{CP}_r^{\dagger})\right] \tag{4.8}$$

$$= \mathrm{Tr}\left(O_r \rho_{k,i}^{\phi}\right). \tag{4.9}$$

In general, given an n-qubit unitary operator as U, estimating an observable written by $\tilde{O} \equiv U(X\otimes I_n)U^{\dagger}$ requires many two-qubit gates. In fact, previous literature utilizes many two-qubit gates to estimate O_q in (4.5) as we see in Fig. 1 and Fig. 2. However, in our case, we can write O_q in (4.5) by only one Pauli observable. That is because CP_q is a Clifford operator, and $X\otimes I_n$ is a Pauli observable; a Clifford operator generally maps one Pauli observable to another Pauli observable. Direct calculation gives the explicit form of O_q as

$$\begin{aligned} O_q = \mathrm{CP}_q^{\dagger}(X\otimes I_n)\mathrm{CP}_q &= (|0\rangle\langle 0|\otimes I_n + |1\rangle\langle 1|\otimes P_q)(X\otimes I_n)(|0\rangle\langle 0|\otimes I_n + |1\rangle\langle 1|\otimes P_q) \\ &= (|0\rangle\langle 1|\otimes I_n + |1\rangle\langle 0|\otimes P_q)(|0\rangle\langle 0|\otimes I_n + |1\rangle\langle 1|\otimes P_q) \\ &= |0\rangle\langle 1|\otimes P_q + |1\rangle\langle 0|\otimes P_q \\ &= X\otimes P_q. \end{aligned} \tag{4.10}$$

To summarize, we do not need the two-qubit operations in Fig. 1 and Fig. 2 and we can compute each summand of (2.16) and (2.17) as

$$\Re\left[e^{i\phi}\left\langle v_{\mathrm{ref}}\right| R_{k,i}^{\dagger} P_q R_{\ell,i'}\left|v_{\mathrm{ref}}\right\rangle\right] = \mathrm{Tr}\left[(X\otimes P_q)\rho_{k,i;\ell,i'}^{\phi}\right], \tag{4.11}$$

$$\Re\left[e^{i\phi}\left\langle v_{\mathrm{ref}}\right| R_{k,i}^{\dagger} P_r R\left|v_{\mathrm{ref}}\right\rangle\right] = \mathrm{Tr}\left[(X\otimes P_r)\rho_{k,i}^{\phi}\right]. \tag{4.12}$$

4.2 VQS with shadow-based strategies

The benefit of writing each O_q and O_r as a Pauli observable is not limited to the reduction of the two-qubit gates. Although we consider a more general case later, we begin by describing a simplified scenario that each phase of the summand in (2.16) does not depend on the index q and that in (2.17) does not depend on the index r; namely

$$g_{k,i}^{*} g_{\ell,i'}\beta_q = G_{k,i,\ell,i',q} e^{i\phi_{k,i,\ell,i'}} \quad (\forall q), \tag{4.13}$$

$$g_{k,i}^{*}\alpha_r = G_{k,i,r} e^{i\phi_{k,i}} \quad (\forall r), \tag{4.14}$$

where $G_{k,i,\ell,i',q}$ and $G_{k,i,r}$ are real values, and $e^{i\phi_{k,i,\ell,i'}}$ and $e^{i\phi_{k,i}}$ are phases independent of indices q and r. Then, we obtain

$$M_{k\ell} = \sum_{i,i'}^{N_g} M_{k\ell}^{i,i'}, \; V_k = \sum_{i=1}^{N_g} V_k^{i}, \tag{4.15}$$

with $M_{k\ell}^{i,i'} \equiv \sum_{q=1}^{N_{\mathrm{BB}}} G_{k,i,\ell,i',q} \mathrm{Tr}\left[(X \otimes P_q)\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}\right]$ and $V_k^i = \sum_{r=1}^{N_{\mathrm{BA}}} G_{k,i,r} \mathrm{Tr}\left[(X \otimes P_r)\rho_{k,i}^{\phi_{k,i}}\right]$. Note that $\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}$ is the density operator $\rho_{k,i;\ell,i'}^{\phi}$ with $\phi = \phi_{k,i,\ell,i'}$ and $\rho_{k,i}^{\phi_{k,i}}$ is the density operator $\rho_{k,i}^{\phi}$ with $\phi = \phi_{k,i}$.

The notable point here is that $M_{k\ell}^{i,i'}$ and V_k^i have the form of $\langle \mathcal{H} \rangle = \sum_{j=1}^{K} a_j \mathrm{Tr}(P_j \rho)$ with $\{a_j\}$ as real coefficients, $\{P_j\}$ as Pauli observables, and ρ as a quantum state. Therefore, we can use shadow-based strategies to efficiently compute the summation inside the brackets. Namely, by our estimating $M_{k\ell}^{i,i'}$ and V_k^i as

$$\begin{aligned} \nu_{k,i;\ell,i'} &= \mathbf{estimateNu}(\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}, \{G_{k,i,\ell,i',q}\}_{q=1}^{N_{\mathrm{BB}}}, \{X \otimes P_q\}_{q=1}^{N_{\mathrm{BB}}}, N_{\mathrm{shot}}), \\ \nu_{k,i} &= \mathbf{estimateNu}(\rho_{k,i}^{\phi_{k,i}}, \{G_{k,i,q}\}_{q=1}^{N_{\mathrm{BA}}}, \{X \otimes P_q\}_{q=1}^{N_{\mathrm{BA}}}, N_{\mathrm{shot}}), \end{aligned} \tag{4.16}$$

the following values

$$\tilde{M}_{k\ell} = \sum_{i,i'}^{N_g} \nu_{k,i;\ell,i'}, \ \tilde{V}_k = \sum_{i=1}^{N_g} \nu_{k,i}, \tag{4.17}$$

become unbiased estimator of $M_{k\ell}$ and V_k. We summarize our proposed estimation process of $M_{k\ell}$ and V_k in Algorithm 3 and Algorithm 4. It should be noted that our circuit synthesis is also useful for grouping strategies [8, 16, 18, 22–33], mainly targeting VQO problems, where we need to calculate $M_{k\ell}^{i,i'}$ and V_k^i.

The variances of $\tilde{M}_{k\ell}$ and $\tilde{V}_k$ are given by

$$\mathrm{Var}(\tilde{M}_{k\ell}) = \sum_{i,i'}^{N_g} \mathrm{Var}\left(\nu_{k,i;\ell,i'}\right), \ \mathrm{Var}(\tilde{V}_k) = \sum_{i=1}^{N_g} \mathrm{Var}(\nu_{k,i}). \tag{4.18}$$

By using (3.10) with $\{a_j\}_{j=1}^{K} = \{G_{k,i,\ell,i',q}\}_{q=1}^{N_{\mathrm{BB}}}$ and $\rho = \rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}$, we obtain

$$\begin{aligned} \mathrm{Var}(\tilde{M}_{k\ell}) &= \frac{(\Delta M_{k\ell})^2}{N_{\mathrm{shot}}}, \\ \Delta M_{k\ell} &\equiv \sqrt{\sum_{i,i'}^{N_g} \left(\sum_{q=1}^{N_{\mathrm{BB}}} \frac{(G_{k,i,\ell,i',q})^2}{q(P_q)} + \sum_{q=1}^{N_{\mathrm{BB}}} \sum_{q' \neq q} G_{k,i,\ell,i',q} G_{k,i,\ell,i',q'} g(\tilde{P}_q, \tilde{P}_{q'}) \mathrm{Tr}(\tilde{P}_q \tilde{P}_{q'} \rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}) \right) - M_{k\ell}^2}, \end{aligned} \tag{4.19}$$

where $\tilde{P}_q = X \otimes P_q$, and by using (3.10) with $\{a_j\}_{j=1}^{K} = \{G_{k,i,r}\}_{q=1}^{N_{\mathrm{BA}}}$ and $\rho = \rho_{k,i}^{\phi_{k,i}}$, we obtain

$$\begin{aligned} \mathrm{Var}(\tilde{V}_k) &= \frac{(\Delta V_k)^2}{N_{\mathrm{shot}}}, \\ \Delta V_k &\equiv \sqrt{\sum_{i}^{N_g} \left(\sum_{r=1}^{N_{\mathrm{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)} + \sum_{r=1}^{N_{\mathrm{BA}}} \sum_{r' \neq r} G_{k,i,r} G_{k,i,r'} g(\tilde{P}_r, \tilde{P}_{r'}) \mathrm{Tr}(\tilde{P}_r \tilde{P}_{r'} \rho_{k,i}^{\phi_{k,i}}) \right) - V_k^2}, \end{aligned} \tag{4.20}$$

where $\tilde{P}_r = X \otimes P_r$.

The above computed variances when using the shadow-based strategy are different from the ones when we naively estimate each summand of $M_{k\ell}^{i,i'} \equiv \sum_{q=1}^{N_{\mathrm{BB}}} G_{k,i,\ell,i',q} \mathrm{Tr}\left[(X \otimes P_q)\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}\right]$ and $V_k^i = \sum_{r=1}^{N_{\mathrm{BA}}} G_{k,i,r} \mathrm{Tr}\left[(X \otimes P_r)\rho_{k,i}^{\phi_{k,i}}\right]$, which is done by the previous literature (we simply call the strategy as the naive strategy in the rest of the paper). Let N_{naive} be the number of measurements for each circuit in the naive strategy, and let the corresponding estimators of $M_{k\ell}$ and V_k be $\tilde{M}_{k\ell}^{\mathrm{naive}}$ and $\tilde{V}_k^{\mathrm{naive}}$. Then the

variances of $\tilde{M}_{k\ell}^{\rm naive}$ and $\tilde{V}_k^{\rm naive}$ are given by

$$\begin{aligned}
{\rm Var}(\tilde{M}_{k\ell}^{\rm naive}) &= \frac{(\Delta M_{k\ell}^{\rm naive})^2}{N_{\rm naive}},\ {\rm Var}(\tilde{V}_k^{\rm naive}) = \frac{(\Delta V_k^{\rm naive})^2}{N_{\rm naive}}, \\
\Delta M_{k\ell}^{\rm naive} &\equiv \sqrt{\sum_{i,i'}^{N_g}\sum_{q=1}^{N_{\rm BB}}\left[(G_{k,i,\ell,i',q})^2 - \left(G_{k,i,\ell,i',q}{\rm Tr}\left(\tilde{P}_q\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}\right)\right)^2\right]}, \\
\Delta V_k^{\rm naive} &\equiv \sqrt{\sum_{i}^{N_g}\sum_{r=1}^{N_{\rm BA}}\left[(G_{k,i,r})^2 - \left(G_{k,i,r}{\rm Tr}\left(\tilde{P}_r\rho_{k,i}^{\phi_{k,i}}\right)\right)^2\right]}.
\end{aligned} \tag{4.21}$$

The difference in the variances between the naive strategy and the shadow-based strategy results in the difference in the infidelity due to the shot noise for a given number of measurements. In the following subsection, we quantitatively evaluate how the variance affects the infidelity and clarifies the benefit of using the shadow-based strategy in the VQS.

Note that in the above introduction of our proposed algorithm, we require that each phase of the summand in (2.16) does not depend on the index q and that in (2.17) does not depend on the index r. However, we can remove the requirement as follows. Instead of (2.14) , we can expand $B^\dagger(t)B(t)$ as

$$B^\dagger(t)B(t) = \sum_{q=1}^{N_{\rm BB}}\Re(\beta_q)P_q + \sum_{q=1}^{N_{\rm BB}}\Im(\beta_q)P_q, \tag{4.22}$$

where the symbol $\Im(z)$ is the imaginary part of z. Then, we can evaluate $\sum_{q=1}^{N_{\rm BB}}\Re(\beta_q)P_q$ and $\sum_{q=1}^{N_{\rm BB}}\Im(\beta_q)P_q$ independently by the shadow-based strategy though the number of measurements increase. That is also true for the computation of $B^\dagger(t)A(t)$.

Hybrid strategy in RTE and ITE

In RTE and ITE, $N_{\rm BB} = 1$ is set to calculate $M_{k\ell}$ and there are no advantages for using a shadow-based strategy for the computation of $M_{k\ell}$. Therefore, in RTE and ITE, we take the hybrid strategy, meaning that we estimate V_k by the shadow-based strategy and $M_{k\ell}$ by the naive strategy. At the end of Section 4.3, we discuss the advantage of the hybrid strategy.

Algorithm 3 Estimation of $M_{k\ell}$ with a shadow-based strategy.

Input: $\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}$, $\{G_{k,i,\ell,i',q}\}_{q=1}^{N_{\rm BB}}$, $\{P_q\}_{q=1}^{N_{\rm BB}}$, $N_{\rm shot}$, N_g
1: SET result = 0.
2: **for** $i = 1...N_g$ **do**
3: **for** $i' = 1...N_g$ **do**
4: SET $\nu_{k,i;\ell,i'}$ = **estimateNu**($\rho_{k,i;\ell,i'}^{\phi_{k,i,\ell,i'}}$, $\{G_{k,i,\ell,i',q}\}_{q=1}^{N_{\rm BB}}$, $\{X\otimes P_q\}_{q=1}^{N_{\rm BB}}$, $N_{\rm shot}$).
5: SET result = result + $\nu_{k,i;\ell,i'}$.
6: **end for**
7: **end for**
return result

4.3 Advantage of using shadow-based strategy

To clarify the advantage of using the shadow-based strategy over the naive strategy, we compare the infidelity due to the shot noise after the duration T. The time derivative of the parameter vector can be ideally computed as

$$\frac{\partial\vec{\theta}_{\rm ideal}(t)}{\partial t} = M^{-1}\vec{V}. \tag{4.23}$$

Algorithm 4 Estimation of V_k with a shadow-based strategy.

Input: $\rho_{k,i}^{\phi_{k,i}}$, $\{G_{k,i,q}\}_{q=1}^{N_{\rm BA}}$, $\{P_q\}_{q=1}^{N_{\rm BA}}$, $N_{\rm shot}$, N_g
SET result $= 0$.
for $i = 1...N_g$ **do**
  for $i' = 1...N_g$ **do**
    SET $\nu_{k,i} =$ **estimateNu**$(\rho_{k,i}^{\phi_{k,i}}, \{G_{k,i,q}\}_{q=1}^{N_{\rm BA}}, \{X \otimes P_q\}_{q=1}^{N_{\rm BA}}, N_{\rm shot})$.
    SET result = result + $\nu_{k,i}$.
  end for
end for
return result

However, we estimate M and $\vec{V}$ with a finite number of measurements, and therefore, the estimated time derivative includes the shot noise:

$$\frac{\partial \vec{\theta}(t)}{\partial t} = (M + \delta M)^{-1}(\vec{V} + \delta \vec{V}), \tag{4.24}$$

where δM and $\delta \vec{V}$ are the error to M and $\vec{V}$ due to the shot noise. As in literature [9], we use the following infidelity as a metric for evaluating the effect of error:

$$D_I\left(\left|v(\vec{\theta}_{\rm ideal}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right) = \sqrt{1 - \left|\left\langle v(\vec{\theta}_{\rm ideal}(T))\right| v(\vec{\theta}(T))\right\rangle\right|^2}, \tag{4.25}$$

between $\left|v(\vec{\theta}_{\rm ideal}(T))\right\rangle$ evolved by the ideal time derivative (4.23) and $\left|v(\vec{\theta}(T))\right\rangle$ evolved by the distorted time derivative (4.24) at time T.

Suppose that $D_I^{\rm naive}$ is an upper bound of D_I when using the naive strategy and $D_I^{\rm shadow}$ is the one when using the shadow-based strategy. As we show in (C.19) in Appendix C, $D_I^{\rm naive} = D_I^{\rm shadow}$ holds when

$$\frac{N_{\rm shot}}{N_{\rm naive}} \simeq \left(\frac{\Delta_{\rm shadow}^{\rm max}}{\Delta_{\rm naive}^{\rm max}}\right)^2, \tag{4.26}$$

where

$$\begin{aligned}
\Delta_{\rm naive} &= ||M^{-1}||_F \sqrt{\sum_k (\Delta V_k^{\rm naive})^2} + ||M^{-1}||_F^2 ||\vec{V}||_2 \sqrt{\sum_{k\ell} (\Delta M_{k\ell}^{\rm naive})^2}, \\
\Delta_{\rm shadow} &= ||M^{-1}||_F \sqrt{\sum_k (\Delta V_k)^2} + ||M^{-1}||_F^2 ||\vec{V}||_2 \sqrt{\sum_{k\ell} (\Delta M_{k\ell})^2},
\end{aligned} \tag{4.27}$$

$\Delta_{\rm naive}^{\rm max} \equiv \max_{0\le t\le T}(\Delta_{\rm naive})$, $\Delta_{\rm shadow}^{\rm max} \equiv \max_{0\le t\le T}(\Delta_{\rm shadow})$, $||\cdot||_F$ is the Frobenius norm, and $||\cdot||_2$ is the two-norm.

Note that the total number of measurements for each step in the case of the shadow-based strategy is $N_{\rm shadow}^{\rm total} = N_{\rm shot}(N_P N_g + N_P^2 N_g^2)$ while that in the naive strategy is $N_{\rm naive}^{\rm total} = N_{\rm naive}(N_P N_g N_{\rm BA} + N_P^2 N_g^2 N_{\rm BB})$. Thus, the advantage of using a shadow-based strategy arises if $N_{\rm shadow}^{\rm total}/N_{\rm naive}^{\rm total} < 1$ when (4.26) is satisfied, i.e.,

$$\frac{N_{\rm shadow}^{\rm total}}{N_{\rm naive}^{\rm total}} = \frac{1 + N_P N_g}{N_{\rm BA} + N_P N_g N_{\rm BB}} \left(\frac{\Delta_{\rm shadow}^{\rm max}}{\Delta_{\rm naive}^{\rm max}}\right)^2 < 1. \tag{4.28}$$

In the rest of this section, we give a rough estimate of the leftest-hand side of the inequality.

Estimation of $N_{\rm shadow}^{\rm total}/N_{\rm naive}^{\rm total}$

We assume that the values of $||M^{-1}||_F$ and $||M||_F^2||\vec{V}||_2$ when using the naive strategy is almost the same as those when using the shadow-based strategy. The assumption is valid since the time evolutions with

both strategies are close to each other (recall that we consider the regime where $D_I^{\rm naive} = D_I^{\rm shadow}$). Under the assumption, we obtain the inequality:

$$\frac{\Delta^{\max}_{\rm shadow}}{\Delta^{\max}_{\rm naive}} < \sqrt{\frac{\sum_k (\Delta V_k)^2}{\sum_k (\Delta V_k^{\rm naive})^2}} + \sqrt{\frac{\sum_{k\ell} (\Delta M_{k\ell})^2}{\sum_{k\ell} (\Delta M_{k\ell}^{\rm naive})^2}}. \tag{4.29}$$

We evaluate each term on the right-hand side in the following.

For the numerator in the square root of the first term, it holds

$$(\Delta V_k)^2 \leq \sum_{i=1}^{N_g} \mathcal{V}_k^i(\rho_{k,i}^{\phi_{k,i}}), \tag{4.30}$$

where

$$\mathcal{V}_k^i(\rho) = \sum_{r=1}^{N_{\rm BA}} \frac{(G_{k,i,r})^2}{q(P_r)} + \sum_{r=1}^{N_{\rm BA}} \sum_{r'\neq r} G_{k,i,r} G_{k,i,r'} g(\tilde{P}_r, \tilde{P}_{r'}) {\rm Tr}(\tilde{P}_r \tilde{P}_{r'} \rho). \tag{4.31}$$

Since $\mathcal{V}_k^i(\rho)$ depends on the quantum state ρ, it is difficult to directly estimate the value. Instead of evaluating $\mathcal{V}_k^i(\rho)$, let us estimate the typical size of $\mathcal{V}_k^i(\rho)$ in the overall Hilbert space by computing

$$\langle \mathcal{V}_k^i \rangle_{\rm Haar} = \int_{\rm Haar} dU \mathcal{V}_k^i(U|0\rangle\langle 0|U^\dagger), \text{ and } {\rm Var}(\mathcal{V}_k^i)_{\rm Haar} = \int_{\rm Haar} dU \left(\mathcal{V}_k^i(U|0\rangle\langle 0|U^\dagger)\right)^2 - (\langle \mathcal{V}_k^i \rangle_{\rm Haar})^2, \tag{4.32}$$

where $\int_{\rm Haar}$ is the average over the Haar distribution. By calculating those statistical values, we can see how $\mathcal{V}_k^i(\rho)$ distributes when ρ is uniformly sampled. As we show in Appendix D, we can readily show

$$\begin{aligned}
\langle \mathcal{V}_k^i \rangle_{\rm Haar} &= \sum_{r=1}^{N_{\rm BA}} \frac{(G_{k,i,r})^2}{q(P_r)}, \\
{\rm Var}(\mathcal{V}_k^i)_{\rm Haar} &= \frac{1}{2^{n+1}(2^{n+1}+1)} \sum_{r=1}^{N_{\rm BA}} \sum_{r'\neq r} \sum_{u=1}^{N_{\rm BA}} \sum_{u'\neq u} G_{k,i,r} G_{k,i,r'} G_{k,i,u} G_{k,i,u'} g(\tilde{P}_r, \tilde{P}_{r'}) g(\tilde{P}_u, \tilde{P}_{u'}) {\rm Tr}(\tilde{P}_r \tilde{P}_{r'} \tilde{P}_u \tilde{P}_{u'}),
\end{aligned} \tag{4.33}$$

where n is the number of qubits in PQC. By using ${\rm Tr}(\tilde{P}_r \tilde{P}_{r'} \tilde{P}_u \tilde{P}_{u'}) \leq 2^{n+1}$, we can show ${\rm Var}(\mathcal{V}_k^i) \propto O(1/2^n)$, which becomes asymptotically zero as n becomes large. Since the variance is exponentially suppressed, $\mathcal{V}_k^i(\rho_{k,i}^{\phi_{k,i}})$ takes the value almost equal to the average with high probability when uniformly choosing $\rho_{k,i}^{\phi_{k,i}}$. In the rest of this section, given a function of a density operator as $\xi(\rho)$, if $\langle \xi(\rho) \rangle_{\rm Haar} = \xi_0$ holds with ξ_0 as a constant and ${\rm Var}(\xi)_{\rm Haar}$ is exponentially suppressed, we denote $\xi(\rho) \simeq \xi_0$. Also, for given another function of a density operator as $\xi'(\rho)$, if $\xi'(\rho) \leq \xi(\rho) \simeq \xi_0$, we denote $\xi'(\rho) \lesssim \xi_0$. Thus, with a large n,

$$\mathcal{V}_k^i(\rho_{k,i}^{\phi_{k,i}}) \simeq \sum_{r=1}^{N_{\rm BA}} \frac{(G_{k,i,r})^2}{q(P_r)} \tag{4.34}$$

holds and

$$(\Delta V_k)^2 \lesssim \sum_{i=1}^{N_g} \sum_{r=1}^{N_{\rm BA}} \frac{(G_{k,i,r})^2}{q(P_r)}. \tag{4.35}$$

In the Eq. (4.29), for the denominator of the square root in the first term,

$$(\Delta V_k^{\rm naive})^2 = \sum_{i=1}^{N_g} \sum_{r=1}^{N_{\rm BA}} \mathcal{G}_k^{i,r}(\rho_{k,i}^{\phi_{k,i}}), \tag{4.36}$$

where

$$\mathcal{G}_k^{i,r}(\rho) = (G_{k,i,r})^2 \left[1 - \left({\rm Tr}(\tilde{P}_r \rho)\right)^2\right]. \tag{4.37}$$

As in the case of $\mathcal{V}_k^i(\rho)$, we can compute the average in the overall Hilbert space (See Appendix D) as

$$\begin{aligned}\langle \mathcal{G}_k^{i,r}(\rho)\rangle_{\mathrm{Haar}} &= (G_{k,i,r})^2\left(1-\frac{1}{2^{n+1}+1}\right),\\ \mathrm{Var}\left(\mathcal{G}_k^{i,r}(\rho)\right)_{\mathrm{Haar}} &= \frac{(G_{k,i,r})^4(2\cdot 2^{2(n+1)}+8\cdot 2^{2(n+1)})}{(2^{n+1}+1)^2(2^{n+1}+3)}.\end{aligned} \tag{4.38}$$

Thus, for a large enough n, $\mathcal{G}_k^{i,r}(\rho)\simeq (G_{k,i,r})^2$ and, we obtain

$$\left(\Delta V_k^{\mathrm{naive}}\right)^2 \simeq \sum_{i=1}^{N_g}\sum_{r=1}^{N_{\mathrm{BA}}}(G_{k,i,r})^2. \tag{4.39}$$

By combining (4.34) and (4.39), the first term in the right hand side of (4.29) is bounded as

$$\sqrt{\frac{\sum_k(\Delta V_k)^2}{\sum_k(\Delta V_k^{\mathrm{naive}})^2}} \lesssim \sqrt{\left\langle \frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}}, \tag{4.40}$$

where $\langle\cdot\rangle_{G_{k,i,r}}$ is a weighted average defined by

$$\langle \mathcal{F}_r\rangle_G = \frac{\sum_{k=1}^{N_P}\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\mathrm{BA}}}(G_{k,i,r})^2\mathcal{F}_r}{\sum_{k=1}^{N_P}\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\mathrm{BA}}}(G_{k,i,r})^2}. \tag{4.41}$$

Similarly,

$$\sqrt{\frac{\sum_{k\ell}(\Delta M_{k\ell})^2}{\sum_{k\ell}(\Delta M_{k\ell}^{\mathrm{naive}})^2}} \lesssim \sqrt{\left\langle \frac{1}{q(\tilde{P}_q)}\right\rangle_{G_{k,i,\ell,i',q}}}, \tag{4.42}$$

where $\langle\cdot\rangle_{G_{k,i,r}}$ is another weighted average defined by

$$\langle \mathcal{F}_q\rangle_{G_{k,i,\ell,i',q}} = \frac{\sum_{k,\ell}^{N_P}\sum_{i,i'}^{N_g}\sum_{q=1}^{N_{\mathrm{BB}}}(G_{k,i,\ell,i',q})^2\mathcal{F}_q}{\sum_{k,\ell}^{N_P}\sum_{i,i'}^{N_g}\sum_{q=1}^{N_{\mathrm{BB}}}(G_{k,i,\ell,i',q})^2}. \tag{4.43}$$

By substituting (4.40) and (4.42) to (4.29), we obtain

$$\frac{\Delta_{\mathrm{shadow}}^{\max}}{\Delta_{\mathrm{naive}}^{\max}} \lesssim \sqrt{\left\langle \frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}} + \sqrt{\left\langle \frac{1}{q(\tilde{P}_q)}\right\rangle_{G_{k,i,\ell,i',q}}}. \tag{4.44}$$

Combining with (4.28), we obtain

$$\frac{N_{\mathrm{shadow}}^{\mathrm{total}}}{N_{\mathrm{naive}}^{\mathrm{total}}} \lesssim \frac{1+N_PN_g}{N_{\mathrm{BA}}+N_PN_gN_{\mathrm{BB}}}\left(\sqrt{\left\langle \frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}} + \sqrt{\left\langle \frac{1}{q(\tilde{P}_q)}\right\rangle_{G_{k,i,\ell,i',q}}}\right)^2. \tag{4.45}$$

The factor $(1+N_PN_g)/(N_{BA}+N_PN_gN_{BB})<1$ corresponds to the benefit of simaltaneously evaluating the N_{BA} terms in the calcualtion of V_k and the N_{BB} terms in the calculation of $M_{k\ell}$ in the shadow-based strategy. The other factor $\sqrt{\langle 1/q(\tilde{P}_r)\rangle_{G_{k,i,r}}}+\sqrt{\langle 1/q(\tilde{P}_q)\rangle_{G_{k,i,\ell,i',q}}}$ corresponds to how well measurements in the shadow-based strategy cover the observables; if observables are not covered efficiently, the benefit of the simaltaneous evaluation may disappear due to this factor. Particularly, when we use the classical shadow as the shadow-based strategy, we can explicitly write the latter factor as

$$\left(\sqrt{\left\langle \frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}} + \sqrt{\left\langle \frac{1}{q(\tilde{P}_q)}\right\rangle_{G_{k,i,\ell,i',q}}}\right)^2 = \left(\sqrt{\langle 3^{\mathrm{locality}}(\tilde{P}_r)\rangle_{G_{k,i,r}}} + \sqrt{\langle 3^{\mathrm{locality}}(\tilde{P}_q)\rangle_{G_{k,i,\ell,i',q}}}\right)^2, \tag{4.46}$$

where we use (3.13). Thus, the locality of the observables should be small in order for enjoying the benefit of the classical shadow in the VQS. It should be noted that

$$\mathrm{locality}(\tilde{P}_r) = \mathrm{locality}(P_r)+1, \tag{4.47}$$

because $\tilde{P}_r = X\otimes P_r$.

Advantages of using the shadow-based strategy in RTE and ITE

Finally, we discuss the advantage of using the shadow-based strategy in RTE and ITE. As we noted at the end of Section 4.2, we take the hybrid strategy that utilizes the shadow-based strategy for computing V_k and the naive strategy for computing $M_{k\ell}$. Unlike the general case (where we measure both V_k and $M_{k\ell}$ by using the shadow-based strategy), we can reduce the number of measurements for estimating each $M_{k\ell}$ to smaller than that for estimating V_k ($=N_{\rm shot}$); we set the number of measurements as $N_{\rm shot}\alpha$ with $0<\alpha<1$.

Let

$$\Delta_{\rm hybrid} = ||M^{-1}||_F\sqrt{\sum_k (\Delta V_k)^2} + ||M^{-1}||_F^2||\vec{V}||_2\sqrt{\frac{\sum_{k\ell}(\Delta M_{k\ell}^{\rm naive})^2}{\alpha}}. \tag{4.48}$$

Then, as shown in (C.24) in Appendix C, we can show that

$$\frac{N_{\rm shot}}{N_{\rm naive}} \simeq \left(\frac{\Delta_{\rm hybrid}^{\rm max}}{\Delta_{\rm naive}^{\rm max}}\right)^2, \tag{4.49}$$

if the error due to the shot noise in the naive strategy is almost the same as that in the hybrid strategy, where we define $\Delta_{\rm hybrid}^{\rm max} \equiv \max_{0\le t\le T}(\Delta_{\rm hybrid})$. Given $N_{\rm hybrid}^{\rm total}$ as the total number of measurements for each step in the hybrid strategy, it holds $N_{\rm hybrid}^{\rm total} = N_{\rm shot}(N_P N_g + \alpha(N_P N_g)^2)$. Under the same assumption to obtain (4.29), it holds

$$\frac{N_{\rm hybrid}^{\rm total}}{N_{\rm naive}^{\rm total}} \lesssim \frac{1+N_P N_g\alpha}{N_{\rm BA}+N_P N_g}\left(\sqrt{\left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}}+\sqrt{\frac{1}{\alpha}}\right)^2 < \frac{1+N_P N_g\alpha}{N_{\rm BA}}\left(\sqrt{\left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}}+\sqrt{\frac{1}{\alpha}}\right)^2. \tag{4.50}$$

By setting $\alpha = 4/\left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}$,

$$\frac{N_{\rm hybrid}^{\rm total}}{N_{\rm naive}^{\rm total}} \lesssim \frac{1+N_\alpha}{N_{\rm BA}} \times \left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}} \times \frac{9}{4}, \tag{4.51}$$

with

$$N_\alpha = N_p N_g \alpha = \frac{4N_P N_g}{\left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}}}. \tag{4.52}$$

Note that the literature [49] states that given a circuit structure, the number of measurements for estimating $\{M_{k\ell}\}_{k,\ell}^{N_P}$ asymptotically becomes negligible compared to that for estimating $\{V_k\}_{k=1}^{N_{\rm BA}}$ as the number of qubits n increases and $N_{\rm BA}$ increases faster than quadratically with n. Though the literature assumes the naive strategy to estimate $\{V_k\}_{k=1}^{N_{\rm BA}}$, the asymptotic behavior would be the same in our hybrid strategy. The condition that the number of measurements for $\{M_{k\ell}\}_{k,\ell}^{N_P}$ is negligible, corresponds to the case when $N_\alpha \ll 1$. In that case, it holds

$$\frac{N_{\rm hybrid}^{\rm total}}{N_{\rm naive}^{\rm total}} \lesssim \frac{1}{N_{\rm BA}} \times \left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}} \times \frac{9}{4}. \tag{4.53}$$

For example, let us calculate the right-hand side value of the right-hand side of (4.53) when the time evolution Hamiltonian is given by (3.22). We can write

$$X\otimes H = XXXXZ + XXXII + XIIXZ + XYYZX + XYYII + XIIZX, \tag{4.54}$$

and therefore, the number of terms is given by $N_{\rm BA}=6$. Recall that each $\tilde{P}_r$ in (4.53) corresponds to each term in (4.54). Let us first calculate when we use the classical shadow. Since the locality of each term in (4.54) is equal to or larger than three,

$$\left\langle\frac{1}{q(\tilde{P}_r)}\right\rangle_{G_{k,i,r}} \ge 3^3, \tag{4.55}$$

where we use (3.13). Thus, in the case of the classical shadow, the right-hand side of (4.53) is

$$\frac{1}{N_{\mathrm{BA}}} \times \left\langle \frac{1}{q(\tilde{P}_r)} \right\rangle_{G_{k,i,r}} \times \frac{9}{4} \geq 10, \tag{4.56}$$

and there is no advantage in using the classical shadow with the Hamiltonian (3.22). On the contrary, when using derandomization, half of the chosen measurements are $XXXXZ$, and the other half is $XYYZX$ (see the discussion around (3.22)). Therefore,

$$\left\langle \frac{1}{q(\tilde{P}_r)} \right\rangle_{G_{k,i,r}} = \frac{1}{2}, \tag{4.57}$$

and the right-hand side of (4.53) is

$$\frac{1}{N_{\mathrm{BA}}} \times \left\langle \frac{1}{q(\tilde{P}_r)} \right\rangle_{G_{k,i,r}} \times \frac{9}{4} = \frac{3}{4}. \tag{4.58}$$

Thus, there is merit in using derandomization with the Hamiltonian (3.22). We note that we utilize the Hamiltonian with just six terms in this example, and the advantage of using the shadow-based methods is limited. In Section 5, we will demonstrate the result of VQS when using the molecular Hamiltonian and see the significant advantages of using the shadow-based methods.

4.4 Remark on the general shadow-based strategies

In Section 4.3, we showed that

$$(\Delta V_k)^2 \lesssim \sum_{i=1}^{N_g} \sum_{r=1}^{N_{\mathrm{BA}}} \frac{G_{k,i,r}}{q(P_r)}, \text{ and } (\Delta V_k^{\mathrm{naive}})^2 \simeq \sum_{i=1}^{N_g} \sum_{r=1}^{N_{\mathrm{BA}}} (G_{k,i,r}^2)$$

in the context of VQS. The same argument holds even in the general estimation problem of $\langle \mathcal{H} \rangle = \sum_{j=1}^{K} a_j \mathrm{Tr}(P_j \rho)$ discussed in Section 3. Let ν_{naive} be the estimator of $\langle \mathcal{H} \rangle$ when estimating each term of $\langle \mathcal{H} \rangle$ one by one with N_{naive} measurements. Also, recall that ν is the estimator of $\langle \mathcal{H} \rangle$ when using a shadow-based strategy with N_{shot} measurements. Then, with the same arguments in Section 4.3, it can be shown that

$$\mathrm{Var}(\nu) \lesssim \frac{1}{N_{\mathrm{shot}}} \sum_{j=1}^{K} \frac{a_j^2}{q(P_j)}, \tag{4.59}$$

$$\mathrm{Var}(\nu_{\mathrm{naive}}) \simeq \frac{1}{N_{\mathrm{naive}}} \sum_{j=1}^{K} a_j^2. \tag{4.60}$$

The total number of measurements is $N_{\mathrm{shadow}}^{\mathrm{total}} = N_{\mathrm{shot}}$ in the case of using the shadow-based strategy and $N_{\mathrm{naive}}^{\mathrm{total}} = K N_{\mathrm{naive}}$ in the case of naively measuring each term one by one. For achieving $\mathrm{Var}(\nu) = \mathrm{Var}(\nu_{\mathrm{naive}})$, the following relation is necessary

$$\frac{N_{\mathrm{shadow}}^{\mathrm{total}}}{N_{\mathrm{naive}}^{\mathrm{total}}} \lesssim \frac{1}{K} \frac{\sum_{j=1}^{K} \frac{a_j^2}{q(P_j)}}{\sum_{j=1}^{K} a_j^2}. \tag{4.61}$$

Notably, as long as the covering probability $q(P_j)$ is $O(1)$, we can reduce the number of measurements by the factor $O(1/K)$ in the shadow-based strategy.

The above observations consolidate the validity of using the shadow-based strategies [17–21] for estimating the sum of Pauli observables. In Eq. (18) of [18], they also give the upper bound for the variance of ν. However, it is as loose as

$$\mathrm{Var}(\nu) \leq \frac{K}{N_{\mathrm{shot}}} \sum_{j=1}^{K} \frac{a_j^2}{q(P_j)}, \tag{4.62}$$

in our notation (in [18], they write the number of terms as $|H_0|$ instead of K). We can easily check that if $\mathrm{Var}(\nu)$ takes the value of the upper-bound (4.62), $N_{\mathrm{shadow}}^{\mathrm{total}} > N_{\mathrm{naive}}^{\mathrm{total}}$, and there are no advantages of using shadow-based strategies. Conversely, previous research does not theoretically show the advantages of using a shadow-based strategy to estimate the sum of Pauli observables. Our upper bound, however, successfully removes the effect from the second term in (3.10) by using the average over the whole quantum state and shows that in most of the quantum states ρ, there is an advantage to using the shadow-based strategy.

5 Numerical Demonstration

Here we demonstrate our shadow-based measurement optimization strategy for the VQS. In Section 4, we generally discussed our strategy, but in this demonstration, we focus on RTE and ITE, which is the algorithm most actively studied in the VQS algorithms. As we noted at the end of Section 4.2, we take the hybrid strategy, meaning that we obtain $M_{k\ell}$ by the naive strategy and V_k by the shadow-based strategy.

We perform two types of demonstrations. The first one is the evaluation of the infidelity due to the shot noise,

$$D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right) = \sqrt{1 - \left|\left\langle v(\vec{\theta}_{\mathrm{ideal}}(T))\right| v(\vec{\theta}(T))\right\rangle\right|^2},$$

which is defined in (4.25), in various Hamiltonians, and in various measurement strategies. The second one is the evaluation of the variances in each strategy. Particularly, as a result of the Haar average in the Hilbert space, we obtained

$$(\Delta V_k)^2 \lesssim \sum_{i=1}^{N_g}\sum_{r=1}^{N_{\mathrm{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)}, \quad \left(\Delta V_k^{\mathrm{naive}}\right)^2 \simeq \sum_{i=1}^{N_g}\sum_{r=1}^{N_{\mathrm{BA}}} (G_{k,i,r})^2,$$

in (4.34) and (4.39) in Section 4.3. We check if it holds in our numerical experiments.

Before going into each experiment, let us give the setting of numerical experiments in Section 5.1. Then, in the subsequent sections, we demonstrate the results of the numerical experiments.

5.1 Setting of numerical experiments

Here we summarize Hamiltonians, the reference state, and the ansatz used in the experiments. Also, we describe strategies to estimate $M_{k\ell}$ and V_k.

Hamiltonians

We use three types of problem Hamiltonians H:

- H_2 molecule in the 6-31g basis with the Bravyi-Kitaev transformation (8-qubit, $N_{\mathrm{BA}} = 184$);
- LiH molecule in the STO-3g basis with the Bravyi-Kitaev transformation (12-qubit, $N_{\mathrm{BA}} = 630$);
- Heisenberg model Hamiltonian (6-qubit, $N_{\mathrm{BA}} = 24$).

We write the number of qubits in the Hamiltonian as n. Here we define our Heisenberg model Hamiltonian by

$$H = 0.1\sum_{j=1}^{6}\sum_{a=x,y,z} \sigma_a^j \sigma_a^{j+1} + 0.1\sum_{j=1}^{6}\sigma_z^j, \tag{5.1}$$

with σ_a^j as a Pauli operator operating on the j-th qubit, and $\sigma_a^7 = \sigma_a^1$.

Reference state and ansatz

As the reference state $|v_{\mathrm{ref}}\rangle$, we use the $|v_{\mathrm{ref}}\rangle = \mathrm{H}^{\otimes n}|0\rangle^{\otimes n}$. As the choice of the ansatz R (defined in (2.10)), we use a hardware efficient ansatz, which contains multiple layers of single-qubit rotation gates and entanglers. Each parameter is embedded in each single-qubit rotation gate as $\exp(-i\sigma_a\theta_j/2)$, where σ_a is a Pauli matrix with $a \in \{x, y, z\}$ (therefore, $N_g = 1$ and $g_{k,1} = -i/2$). The type of each single-qubit rotation gate 'a' is randomly chosen at the initialization. As the two-qubit gates in entanglers, we use the controlled-Z gates. At each layer, we first perform the single qubit gates and, secondly, perform the controlled-Z gates. The way of operating the controlled-Z gates is different in the odd layers and the even layers. Let us define a set of integers $\mathcal{N}_1$ by $\mathcal{N}_1 = \{m_1 | m_1 \in \mathbf{N} \textit{ and } 1 < 2m_1 \leq n\}$ and another set of integers $\mathcal{N}_2$ by $\mathcal{N}_2 = \{m_2 | m_2 \in \mathbf{N} \textit{ and } 1 < 2m_2 + 1 \leq n\}$. For example, when $n = 8$, $\mathcal{N}_1 = \{1, 2, 3, 4\}$ and $\mathcal{N}_2 = \{1, 2, 3\}$. Also, we denote by (j, k) the pair of control and target qubits on which we perform a two-qubit operation. Then, in each odd layer, pairs of $(2m_1 - 1, 2m_1)$ for all $m_1 \in \mathcal{N}_1$ are firstly entangled by the controlled-Z gates, and next, pairs of $(2m_2, 2m_2+1)$ for all $m_2 \in \mathcal{N}_2$ are entangled. On the contrary, in each even layer, pairs $(2m_2, 2m_2 + 1)$ for all $m_2 \in \mathcal{N}_2$ are firstly entangled by the controlled-Z gates, and next, pairs $(2m_1 - 1, 2m_1)$ for all $m_1 \in \mathcal{N}_1$ are entangled. In Fig. 4, we show our ansatz with the reference state $|v_{\mathrm{ref}}\rangle$ when $n = 8$ as an example. For all experiments, we fix the number of layers as four, and therefore, $N_P = 4n$. All parameters are initialized as zero, and therefore, the initial state generated by the ansatz is $|v_{\mathrm{ref}}\rangle$.

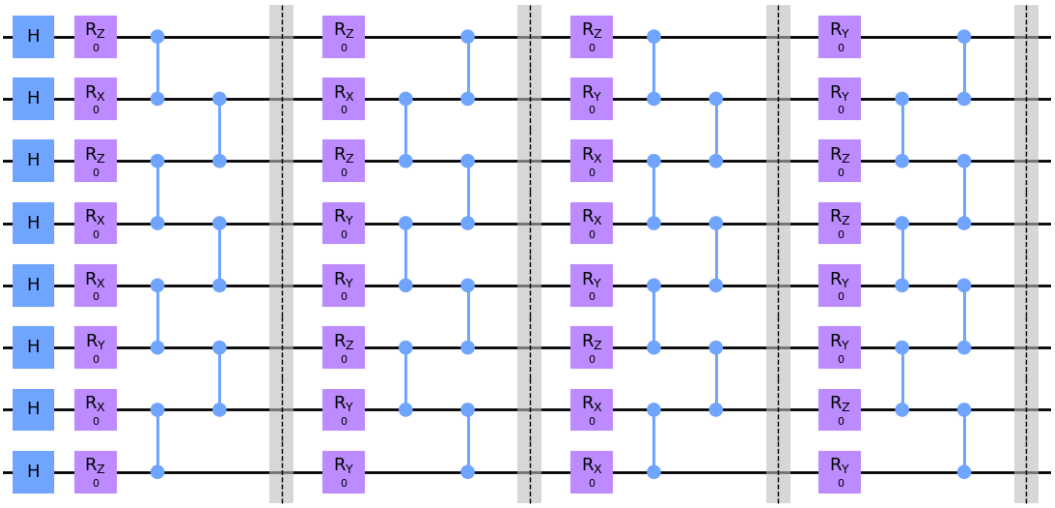


Figure 4: The reference state and the ansatz used in the numerical experiment. As an example, we shot the case when $n = 8$.

Strategies to estimate $M_{k\ell}$ and V_k

For the estimation of V_k, we compare following four strategies including two shadow-based strategies:

- Classical shadow (=a shadow-based strategy);
- Derandomization (=another shadow-based strategy);
- Naively estimate each term of V_k one by one;
- Largest degree first (LDF) grouping [16, 18].

The LDF grouping is a method to reduce the number of measurements by grouping the simultaneously measurable Pauli observables. In the LDF grouping, the measurement basis is chosen so that all of the observables in one of the groups should be covered by each measurement. The group to be covered by each measurement is chosen according to the probability distribution weighted by the importance of each group,

i.e., the larger the norm of the coefficients of the observables in the group is, the larger the probability becomes. As shown in previous literature [18–20], the LDF grouping successfully decreases the number of measurements in the VQE problems without increasing the depth of the circuit. Due to the circuit synthesis in Section 4.1, $M_{k\ell}$ and V_k are written as the sum of Pauli observables, and the LDF grouping is also usable for the evaluation of V_k and $M_{k\ell}$. Thus, it can be a good benchmark for evaluating the shadow-based strategy. For a more detailed explanation of the LDF grouping, please see [18].

For the estimation of $M_{k\ell}$, there is no room for measurement optimization since $N_{\mathrm{BB}} = 1$. To emphasize the difference in the performances of the above four estimation strategies of V_k, we assume that $M_{k\ell}$ can be exactly computed (which corresponds to the case when the number of measurements for estimating $\{M_{k\ell}\}$ is negligible). We can obtain the exact value of $M_{k\ell}$ in Eq. (2.16) by directly calculating $\langle v_{\mathrm{ref}}| R^{\dagger}_{k,i} P_q R_{\ell,i'} |v_{\mathrm{ref}}\rangle$ with the state vector representation. Finally, throughout our numerical experiment, for simulating the quantum process, we use Qulacs [50].

5.2 Evaluation of the infidelity

Under the setting we described in Section 5.1, we execute the RTE and ITE for the initial five steps such as $T = x\delta t$ ($x \in \{0, 1, 2, 3, 4, 5\}$) where we set $\delta t = 0.01$, and see how $D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right)$ evolves depending on the strategies to estimate V_k. For $T = 0$, we have $D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right) = 0$. For computing $\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle$, we exactly compute V by using the state vector representation, while for estimating $\left|v(\vec{\theta}(T))\right\rangle$, we evaluate V by using each measurement strategy. Given the total number of measurements for estimating each V_k as N_{total}, it is set to be $N_{\mathrm{total}} = 5 \times N_{\mathrm{BA}}$ for each strategy (thus, $N_{\mathrm{shot}} = 5 \times N_{\mathrm{BA}}$ for the shadow-based strategy and $N_{\mathrm{naive}} = 5$ for the naive strategy). For example, in the case of the H_2 molecule Hamiltonian, $N_{\mathrm{BA}} = 184$, and the total number of measurements is 920.

The resulting evolutions of $D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right)$ for each T are shown in Fig. 5. There are four subfigures in it, which correspond to the followings: (a) ITE with H_2 molecule Hamiltonian, (b) RTE with H_2 molecule Hamiltonian, (c) ITE with Heisenberg model Hamiltonian and (d) ITE with LiH molecule Hamiltonian. We perform five trials and write the mean and std error of the mean for each data point. For each trial, different gate types are randomly chosen in PQC at initialization, but the chosen gate types are the same in every strategy as far as they are in the same trial. For example, the gate types chosen in the first trial of the naive strategy are the same as those in the first trial of the classical shadow strategy.

In Fig. 5, we see that the classical shadow outperforms the naive method other than the problems with the Heisenberg model Hamiltonian. The reason why the classical shadow does not work well in the case of the Heisenberg model Hamiltonian is as follows. The Heisenberg model Hamiltonian (5.1) has $N_{\mathrm{BA}} = 24$ terms, and in the classical shadow, it holds

$$\left\langle \frac{1}{q(\tilde{P}_r)} \right\rangle_{G_{k,i,r}} \simeq 3^3, \tag{5.2}$$

where we use that the locality of most of the terms in (5.1) is two and in the observable to evaluate V_k, the locality increases by one (see (4.47)). By substituting these into (4.53),

$$\frac{N^{\mathrm{total}}_{\mathrm{hybrid}}}{N^{\mathrm{total}}_{\mathrm{naive}}} \lesssim \frac{1}{N_{\mathrm{BA}}} \times \left\langle \frac{1}{q(\tilde{P}_r)} \right\rangle_{G_{k,i,r}} \times \frac{9}{4} \simeq 2.5, \tag{5.3}$$

meaning that there is not an advantage of using the classical shadow algorithm in the Heisenberg model Hamiltonian as far as we see the upper bound.

Even though the classical shadow algorithm successfully works in the other problems compared to the naive strategy, the LDF grouping outperforms the classical shadow in every problem. That is because the classical shadow contains inefficiency regarding the choice of the measurement basis. The most significant inefficiency might exist in the choice of the measurement basis for the ancilla qubit. For evaluating V_k, we

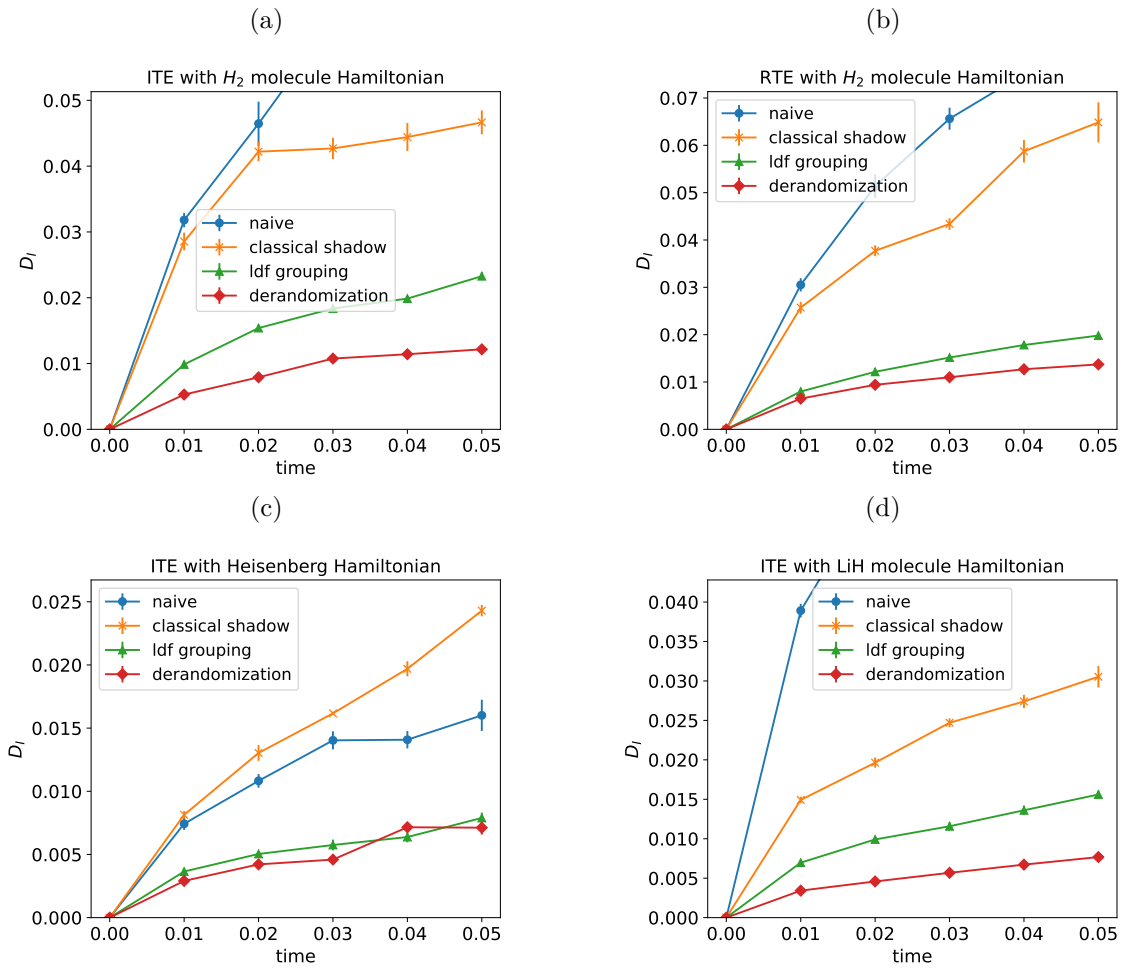


Figure 5: The evolutions of $D_I\left(\left|v(\vec{\theta}_{\text{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right)$ for each T. Four subfigures correspond to the followings: (a) ITE with H_2 molecule Hamiltonian, (b) RTE with H_2 molecule Hamiltonian, (c) ITE with Heisenberg model Hamiltonian, and (d) ITE with LiH molecule Hamiltonian.

need to evaluate the following summation of the observables:

$$\sum_{r}^{N_{\text{BA}}} \alpha_r X \otimes P_r. \tag{5.4}$$

Since all the observables have X in the first qubit, the measurement basis that does not have X in the first qubit does not give any information about V_k. Conversely, only the measurement basis that has X in the first qubit covers the terms in (5.4). However, the measurement basis is chosen with the probability of $1/3$, which is obviously inefficient in the computation of V_k. The inefficiency is the reason why the LDF grouping outperforms the classical shadow.

We enjoy the benefit of the shadow-based strategy with a more intelligent selection of the measurement basis. The derandomization, which optimizes the measurement basis so that $1/q(P_r)$ becomes smaller, outperforms the LDF grouping other than the problem with the Heisenberg model Hamiltonian. In the Heisenberg model Hamiltonian, all observables are covered by the only three measurement basis:

$$XXXXXXX, XYYYYYY, XZZZZZZ. \tag{5.5}$$

Finding such a measurement basis is easy both for the LDF grouping and the derandomization. In our numerical experiment, the LDF grouping successfully groups all the observables into three groups, where

each group is covered by one of the above measurement basis. Thus, the measurements in the LDF grouping are nearly optimal and equal to the derandomization in the Heisenberg model Hamiltonian. However, in the other problems, the observables are not covered by a few measurement basis. Thus the derandomization, which is a relatively intelligent way of finding the measurement basis, outperforms the LDF grouping.

We also compute $D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right)$ at $T=1$ with various total numbers of measurements N_{total}. Similarly to Fig 5, each data point is the mean and the standard error of the mean in five trials. The results are shown in Fig. 6 (the labels of the subfigures are the same as those of Fig. 5). We note that both the horizontal and vertical axes are in the logarithmic scale in the figure. We see that the same conclusion holds for the performance of each measurement strategy. Additionally, we see that $D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right)$ almost decreases with the square root of N_{total}, which is consistent with (C.23) in Appendix C.

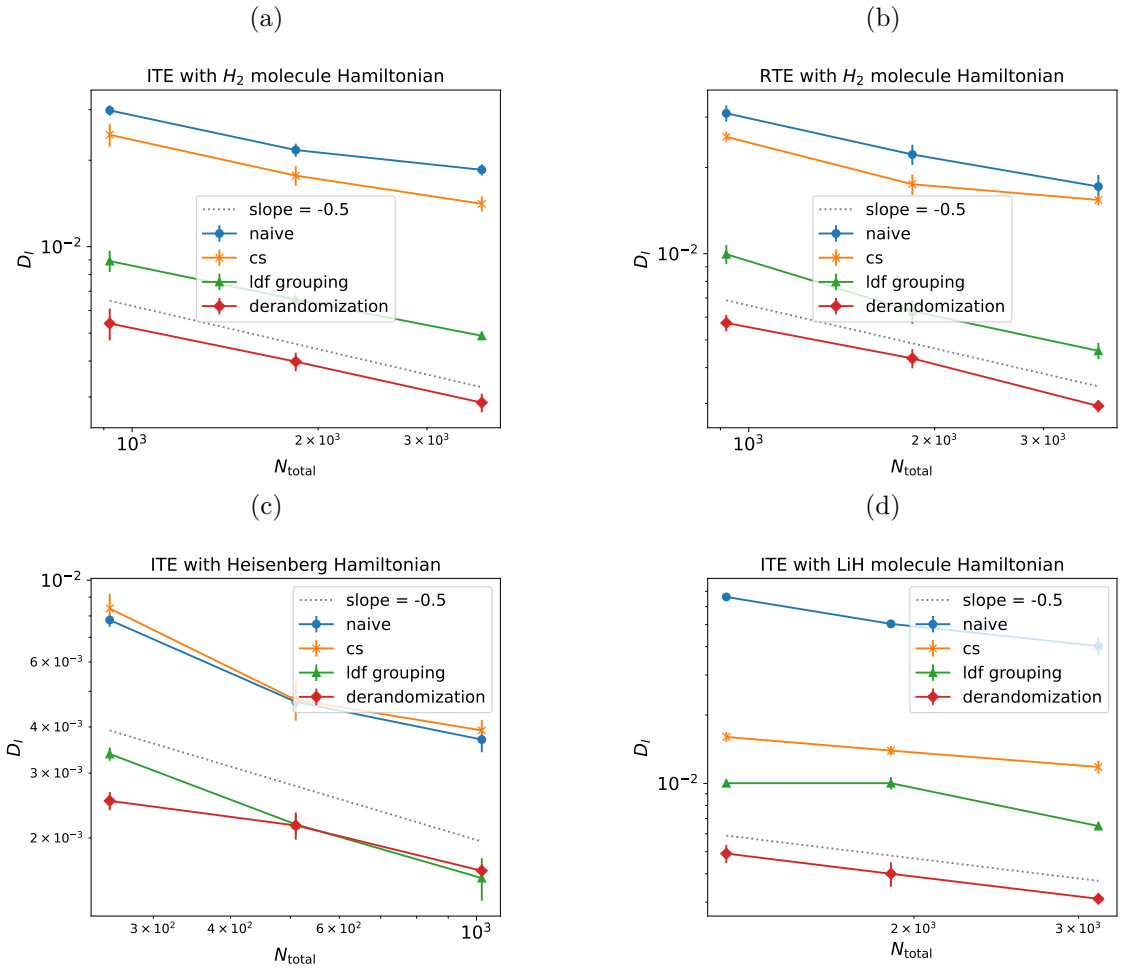


Figure 6: The value of $D_I\left(\left|v(\vec{\theta}_{\mathrm{ideal}}(T))\right\rangle, \left|v(\vec{\theta}(T))\right\rangle\right)$ at $T=1$ for each N_{total}. As in Fig. 5, four subfigures correspond to the followings: (a) ITE with H_2 molecule Hamiltonian, (b) RTE with H_2 molecule Hamiltonian, (c) ITE with Heisenberg model Hamiltonian, and (d) ITE with LiH molecule Hamiltonian. Both the horizontal and the vertical axes are in the logarithmic scale. As a reference, we plot a slope proportional to $N_{\mathrm{total}}^{-0.5}$ by a gray dotted line.

5.3 Evaluation of the variance

Finally, we numerically compare the variance $\mathrm{Var}(\tilde{V}_k)$ and $\mathrm{Var}(\tilde{V}_k^{\mathrm{naive}})$ for a given total number of measurements. We set the total number of measurements as $5 \times N_{\mathrm{BA}}$. We compute the variance in the two

problems: ITE with H_2 molecule Hamiltonian, which is shown in TABLE 1, and ITE with Heisenberg model Hamiltonian, which is shown in TABLE 2.

The detail of the contents of the tables is as follows. In the 'Variance' column, we show the average of the variance; namely,

$$\text{Variance} = \frac{1}{N_P}\sum_{k=1}^{N_P}\text{Var}(\tilde{V}_k), \tag{5.6}$$

for the shadow-based strategy and

$$\text{Variance} = \frac{1}{N_P}\sum_{k=1}^{N_P}\text{Var}(\tilde{V}_k^{\text{naive}}), \tag{5.7}$$

for the naive strategy. Additionally, checking

$$(\Delta V_k)^2 \lesssim \sum_{i=1}^{N_g}\sum_{r=1}^{N_{\text{BA}}}\frac{(G_{k,i,r})^2}{q(P_r)}, \quad \left(\Delta V_k^{\text{naive}}\right)^2 \simeq \sum_{i=1}^{N_g}\sum_{r=1}^{N_{\text{BA}}}(G_{k,i,r})^2,$$

in the 'Approximation' column, we show the approximate value of 'Variance', which is denoted by

$$\text{Approximation} = \frac{1}{N_P}\sum_{k=1}^{N_P}\left(\frac{1}{N_{\text{shot}}}\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\text{BA}}}\frac{(G_{k,i,r})^2}{q(P_r)}\right), \tag{5.8}$$

for the shadow-based strategy, and

$$\text{Approximation} = \frac{1}{N_P}\sum_{k=1}^{N_P}\left(\frac{1}{N_{\text{naive}}}\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\text{BA}}}(G_{k,i,r})^2\right), \tag{5.9}$$

for the naive strategy. Finally, in the 'Diff' row, we show the averege difference between the true value and the approximate value of the variance, namely,

$$\text{Diff} = \frac{1}{N_P}\sum_{k=1}^{N_P}\left|\text{Var}(\tilde{V}_k) - \frac{1}{N_{\text{shot}}}\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\text{BA}}}\frac{(G_{k,i,r})^2}{q(P_r)}\right|, \tag{5.10}$$

for the shadow-based strategy, and

$$\text{Diff} = \frac{1}{N_P}\sum_{k=1}^{N_P}\left|\text{Var}(\tilde{V}_k^{\text{naive}}) - \frac{1}{N_{\text{naive}}}\sum_{i=1}^{N_g}\sum_{r=1}^{N_{\text{BA}}}(G_{k,i,r})^2\right|, \tag{5.11}$$

for the naive strategy. The value of 'Variance' depends on the parameters of the ansatz. We choose parameters in two different ways when computing 'Variance': (i) all zero, where all parameters are set to be zero, and (ii) random, where all parameters are randomly selected. For the random parameters choice case, we generate five patterns of parameters.

There are several implications from the results in the tables. First, both in TABLE 1 and TABLE 2, we see that all 'Variance' are almost equal to the corresponding 'Approximation', which shows the validity of the approximation in (4.34) and (4.39) in Section 4.3. Second, the variances are consistent with the results in Section 5.2. Namely, the smaller variance results in the smaller infidelity in Fig. 5 and Fig. 6. Finally, variances in the derandomization are much smaller than those in the other measurement methods. Particularly, the large difference in the variances between the classical shadow and the derandomization shows the importance of optimizing the covering probability $q(P_r)$ for decreasing the variance (hence the infidelity) and implies that other shadow-based strategies [18, 20], which also optimizes the covering probability, also achieve good performance. Studying the effect of using [18, 20] in VQS is left for future work.

Table 1: Variance, Approximation, and Diff in the problem of ITE with H_2 molecule Hamiltonian.

Parameters	Measurement	Variance	Approximation	Diff
all zero	**naive**	0.731	0.743	0.012
	classical shadow	0.524	0.525	0.001
	derandomization	0.020	0.020	0.000
random (pattern1)	**naive**	0.741	0.743	0.001
	classical shadow	0.525	0.525	0.000
	derandomization	0.020	0.020	0.000
random (pattern2)	**naive**	0.731	0.743	0.011
	classical shadow	0.528	0.525	0.002
	derandomization	0.019	0.020	0.000
random (pattern3)	**naive**	0.733	0.743	0.010
	classical shadow	0.526	0.525	0.000
	derandomization	0.020	0.020	0.000
random (pattern4)	**naive**	0.736	0.743	0.006
	classical shadow	0.528	0.525	0.002
	derandomization	0.020	0.020	0.000
random (pattern5)	**naive**	0.732	0.743	0.010
	classical shadow	0.510	0.525	0.014
	derandomization	0.019	0.020	0.000

Table 2: Variance, Approximation, and Diff in the problem of ITE with Heisenberg model Hamiltonian.

Parameters	Measurement	Variance	Approximation	Diff
all zero	**naive**	0.0475	0.0480	0.0004
	classical shadow	0.0457	0.0450	0.0007
	derandomization	0.0063	0.0059	0.0004
random (pattern1)	**naive**	0.0469	0.0480	0.0010
	classical shadow	0.0464	0.0450	0.0014
	derandomization	0.0055	0.0059	0.0003
random (pattern2)	**naive**	0.0472	0.0480	0.0007
	classical shadow	0.0436	0.0450	0.0013
	derandomization	0.0054	0.0059	0.0004
random (pattern3)	**naive**	0.0471	0.0480	0.0008
	classical shadow	0.0443	0.0450	0.0006
	derandomization	0.0054	0.0059	0.0005
random (pattern4)	**naive**	0.0467	0.0480	0.0012
	classical shadow	0.0449	0.0450	0.0000
	derandomization	0.0065	0.0059	0.0006
random (pattern5)	**naive**	0.0469	0.0480	0.0010
	classical shadow	0.0456	0.0450	0.0006
	derandomization	0.0058	0.0059	0.0001

6 Discussion

The variational quantum simulation (VQS) successfully gives a way of simulating quantum systems in those devices, which is considered as one of the practical applications the near-term quantum devices. However, when applying VQS to a large quantum system where we have practical interests, the number of running a quantum device, i.e., the number of measurements, dramatically increases to reduce the effect of shot noise. Even though the grouping methods and the shadow-based strategies have recently been proposed to decrease the number of measurements in the variational quantum optimization (VQO), they are not utilized in VQS because the way of evaluating the observables in VQS is different from the one in VQO and we cannot straightforwardly apply the shadow-based strategies to VQS. In this paper, by changing the circuit and observables in VQS, we make it possible to apply shadow-based strategies and grouping methods to VQS. We particularly focus on the shadow-based strategies; we theoretically show that with the shadow-based strategy, we can greatly decrease the number of measurements in VQS compared to the naive strategy, which has been utilized since the original proposal of VQS [9,13,14]. Our numerical demonstration of ITE and RTE using various Hamiltonians, including chemical Hamiltonians, also implies the validity of using the shadow-based strategy. Particularly, the derandomization [19], one of the shadow-based strategies, achieves the best performance in most of the Hamiltonians.

Though we examine only the classical shadow and the derandomization in our numerical experiments as the shadow-based strategies, other shadow-based strategies [20, 21] can also be utilized to reduce the number of measurements in VQS. For example, the literature [20] shows that their proposed shadow-based strategy using decision diagrams effectively reduces the estimator's variance to smaller than that in the LDF grouping, which achieves the second-best performance in our numerical demonstration. Additionally, as we note in Section 1, recent literature [31] reports that their grouping strategies outperform shadow-based strategies, such as derandomization, by optimizing parameters in the grouping so that the variance of the observable is minimized. Studying the effect of using those shadow-based strategies and the grouping strategies in VQS is a good direction for future work. We believe that our work takes the first step toward measurement optimization in VQS and paves the way for building practical applications for simulating quantum systems.

Acknowledgement

This work was supported by Grant-in-Aid for JSPS Research Fellow 22J01501 and by Leading Initiative for Excellent Young Researchers MEXT Japan and JST presto (Grant No. JPMJPR1919) Japan. This work was supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Number JPMXS0120319794. This paper is partly based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

A Statistical properties of ν

Proof that ν is an unbiased estimator of $\mathrm{Tr}(H\rho)$

In this Section, we show that ν is the unbiased estimator of $\mathrm{Tr}(H\rho)$. When measurements are chosen probabilistically,

$$\begin{aligned}
\mathbf{E}[\nu_r] &= \sum_{j=1}^{K} a_j \frac{1}{q(P_j)} \sum_{M_r \in \{X,Y,Z\}^{\otimes n}} f(P_j, M_r) q_{\mathrm{m}}(M_r) \sum_b p(\rho, M_r; b) \mu(P_j, b) \\
&= \sum_{j=1}^{K} a_j \frac{1}{q(P_j)} \sum_{M_r \in \{X,Y,Z\}^{\otimes n}} f(P_j, M_r) q_{\mathrm{m}}(M_r) \mathrm{Tr}(P_j \rho) \\
&= \sum_{j=1}^{K} a_j \mathrm{Tr}(P_j \rho) \\
&= \mathrm{Tr}(\mathcal{H}\rho),
\end{aligned} \tag{A.1}$$

where we use (3.9) in the second equality, and we use (3.5) in the third equality. Thus, $\mathbf{E}[\nu] = \sum_{r=1}^{N_{\rm shot}} \mathbf{E}[\nu_r]/N_{\rm shot} = {\rm Tr}(\mathcal{H}\rho)$.

When measurements are chosen deterministically,

$$\begin{aligned}\mathbf{E}[\nu_r] &= \sum_{j=1}^{K} a_j \frac{1}{q(P_j)} f(P_j, M_r) \sum_b p(\rho, M_r; b)\mu(P_j, b) \\ &= \sum_{j=1}^{K} a_j \frac{1}{q(P_j)} f(P_j, M_r){\rm Tr}(P_j\rho) \\ &= \sum_{j=1}^{K} a_j \frac{N_{\rm shot} f(P_j, M_r){\rm Tr}(P_j\rho)}{\sum_{r'=1}^{N_{\rm shot}} f(P_j, M_{r'})},\end{aligned} \tag{A.2}$$

where we use (3.6) in the last equality. Therefore,

$$\begin{aligned}\mathbf{E}[\nu] &= \frac{1}{N_{\rm shot}} \sum_{r=1}^{N_{\rm shot}} \mathbf{E}[\nu_r] \\ &= \sum_{j=1}^{K} a_j \frac{\sum_{r=1}^{N_{\rm shot}} f(P_j, M_r){\rm Tr}(P_j\rho)}{\sum_{r'=1}^{N_{\rm shot}} f(P_j, M_{r'})} \\ &= \sum_{j=1} a_j {\rm Tr}(P_j\rho) \\ &= {\rm Tr}(\mathcal{H}\rho).\end{aligned} \tag{A.3}$$

Variance of ν

The variance of ν is defined by

$${\rm Var}(\nu) = \mathbf{E}[\nu^2] - (\mathbf{E}[\nu])^2 = \frac{1}{N_{\rm shot}} \left(\frac{1}{N_{\rm shot}} \sum_{r=1}^{N_{\rm shot}} \mathbf{E}[\nu_r^2] - [{\rm Tr}(\mathcal{H}\rho)]^2 \right). \tag{A.4}$$

When measurements are probabilistically chosen,

$$\mathbf{E}[\nu_r^2] = \sum_{j=1}^{K}\sum_{\ell=1}^{K} a_j a_\ell \sum_{M_r \in \{X,Y,Z\}^{\otimes n}} \frac{f(P_j, M_r)}{q(P_j)} \frac{f(P_\ell, M_r)}{q(P_\ell)} q_m(M_r) \sum_{b_r \in \{-1,1\}^{\otimes n}} p(\rho, M_r; b_r)\mu(P_j, b_r)\mu(P_\ell, b_r). \tag{A.5}$$

As long as M_r covers both P_j and P_ℓ

$$\sum_{b_r \in \{-1,1\}^{\otimes n}} p(\rho, M_r; b_r)\mu(P_j, b_r)\mu(P_\ell, b_r) = {\rm Tr}(P_j P_\ell \rho). \tag{A.6}$$

Substituting (A.6) to (A.5), we obtain

$$\mathbf{E}[\nu_r^2] = \sum_{j=1}^{K}\sum_{\ell=1}^{K} a_j a_\ell \sum_{M_r \in \{X,Y,Z\}^{\otimes n}} \frac{f(P_j, M_r)}{q(P_j)} \frac{f(P_\ell, M_r)}{q(P_\ell)} q_m(M_r) {\rm Tr}(P_j P_\ell \rho). \tag{A.7}$$

By using $f(P_j, M_r)^2 = f(P_j, M_r)$,

$$\sum_{M_r \in \{X,Y,Z\}^{\otimes n}} \left(\frac{f(P_j, M_r)}{q(P_j)} \right)^2 q_m(M_r) = \frac{1}{q(P_j)}, \tag{A.8}$$

and therefore,

$$\mathbf{E}[\nu_r^2] = \sum_{j=1}^{K} \frac{a_j^2}{q(P_j)} + \sum_{j\neq\ell} a_j a_\ell \sum_{M_r\in\{X,Y,Z\}^{\otimes n}} \frac{f(P_j,M_r)}{q(P_j)} \frac{f(P_\ell,M_r)}{q(P_\ell)} q_m(M_r)\mathrm{Tr}(P_j P_\ell \rho). \tag{A.9}$$

As a result,

$$\mathrm{Var}(\nu) = \frac{1}{N_{\mathrm{shot}}} \left[\sum_{j=1}^{K} \frac{a_j^2}{q(P_j)} + \sum_{j\neq k} a_j a_\ell \sum_{M_r\in\{X,Y,Z\}^{\otimes n}} \frac{f(P_j,M_r)}{q(P_j)} \frac{f(P_\ell,M_r)}{q(P_\ell)} q_m(M_r)\mathrm{Tr}(P_j P_\ell \rho) - [\mathrm{Tr}(\mathcal{H}\rho)]^2 \right]. \tag{A.10}$$

In the deterministic case, by using (A.6),

$$\frac{1}{N_{\mathrm{shot}}} \sum_{r=1}^{N_{\mathrm{shot}}} \mathbf{E}[\nu_r^2] = \frac{1}{N_{\mathrm{shot}}} \sum_{j=1}^{K}\sum_{\ell=1}^{K} a_j a_\ell \sum_{r=1}^{N_{\mathrm{shot}}} \frac{f(P_j,M_r)}{q(P_j)} \frac{f(P_\ell,M_r)}{q(P_\ell)} \mathrm{Tr}(P_j P_\ell \rho). \tag{A.11}$$

By using a calculation similar to (A.8), we can separate the diagonal term and obtain

$$\mathrm{Var}(\nu) = \frac{1}{N_{\mathrm{shot}}} \left[\sum_{j=1}^{K} \frac{a_j^2}{q(P_j)} + \frac{1}{N_{\mathrm{shot}}} \sum_{j\neq\ell} a_j a_\ell \sum_{r=1}^{N_{\mathrm{shot}}} \frac{f(P_j,M_r)}{q(P_j)} \frac{f(P_\ell,M_r)}{q(P_\ell)} \mathrm{Tr}(P_j P_\ell \rho) - [\mathrm{Tr}(\mathcal{H}\rho)]^2 \right]. \tag{A.12}$$

B Variance in classical shadow

Here we derive the variance of ν in the classical shadow [17]. From (3.11), we obtain

$$g(P_j,P_\ell) = 3^{\mathrm{locality}(P_j)} 3^{\mathrm{locality}(P_\ell)} \left(\frac{1}{3^n} \sum_{M_r\in\{X,Y,Z\}^{\otimes n}} f(P_j,M_r) f(P_\ell,M_r) \right) \tag{B.1}$$

$$= 3^{\mathrm{locality}(P_j)} 3^{\mathrm{locality}(P_\ell)} \prod_{i=1}^{n} \frac{1}{3} \sum_{M_r[i]\in\{X,Y,Z\}} f_{\mathrm{local}}(P_j[i],M_r[i]) f_{\mathrm{local}}(P_\ell[i],M_r[i]). \tag{B.2}$$

We can compute the value inside the product as

$$\frac{1}{3} \sum_{M_r[i]\in\{X,Y,Z\}} f_{\mathrm{local}}(P_j[i],M_r[i]) f_{\mathrm{local}}(P_\ell[i],M_r[i]) = \begin{cases} 1 & P_j[i] = P_\ell[i] = I \\ 1/3 & (P_j[i] = I \text{ or } P_\ell[i] = I) \text{ and } P_j[i] \neq P_\ell[i] \\ 1/3 & P_j[i] = P_\ell[i] \neq I \\ 0 & \textit{otherwise} \end{cases}. \tag{B.3}$$

Also, we can expand

$$3^{\mathrm{locality}(P_j)} 3^{\mathrm{locality}(P_\ell)} = \prod_{i=1}^{n} c_{\mathrm{local}}(P_j[i],P_\ell[i]), \tag{B.4}$$

where

$$c_{\mathrm{local}}(P_j[i],P_\ell[i]) = \begin{cases} 1 & P_j[i] = P_\ell[i] = I \\ 3 & (P_j[i] = I \text{ or } P_\ell[i] = I) \text{ and } P_j[i] \neq P_\ell[i] \\ 9 & P_j[i] = P_\ell[i] \neq I \end{cases}. \tag{B.5}$$

By substituting (B.3) and (B.4) into (B.2), we obtain

$$g(P_j,P_\ell) = g_{\mathrm{cs}}(P_j,P_\ell) \equiv \prod_{i=1}^{n} g_{\mathrm{local}}(P_j[i],P_\ell[i]), \tag{B.6}$$

where

$$g_{\rm local}(P_j[i],P_\ell[i])=\begin{cases}1 & P_j[i]=I \text{ or } P_\ell[i]=I\\ 3 & P_j[i]=P_\ell[i]\neq I\\ 0 & otherwise\end{cases}. \tag{B.7}$$

Consequently,

$$\mathrm{Var}(\nu)=\frac{1}{N_{\rm shot}}\left[\sum_{j=1}^{K}a_j^2 3^{\mathrm{locality}(P_j)}+\sum_{j\neq\ell}a_j a_\ell g_{\rm cs}(P_j,P_\ell)\mathrm{Tr}(P_jP_\ell\rho)-[\mathrm{Tr}(\mathcal{H}\rho)]^2\right]. \tag{B.8}$$

C Derivation of the bound for shot noise

In [9], the following infidelity

$$D_I\left(\left|v(\vec{\theta}_{\rm ideal}(T))\right\rangle,\left|v(\vec{\theta}(T))\right\rangle\right)=\sqrt{1-\left|\left\langle v(\vec{\theta}_{\rm ideal}(T))\right|v(\vec{\theta}(T))\right\rangle\right|^2}$$

is evaluated when not using the shadow-based strategy. In this section, we firstly show the value of the previously computed D_I and next, we evaluate D_I when using the shadow-based strategy.

Suppose that there are N-steps for the update of the parameters, i.e., $N\delta t=T$, and let us write the state evolved by (4.24) and the ideal state evolved by (4.23) at r-th step as $|v^r\rangle$ and $|v^r_{\rm ideal}\rangle$. Then, $\left|v^N\right\rangle=\left|v(\vec{\theta}(T))\right\rangle$ and $\left|v^N_{\rm ideal}\right\rangle=\left|v(\vec{\theta}_{\rm ideal}(T))\right\rangle$. We also assume that $|v^0\rangle=|v^0_{\rm ideal}\rangle$. By using the triangle inequality, we obtain

$$\begin{aligned}D_I\left(|v^r_{\rm ideal}\rangle,|v^r\rangle\right)&=D_I\left(U_r\left|v^{r-1}_{\rm ideal}\right\rangle,|v^r\rangle\right)\\&\leq D_I\left(U_r\left|v^{r-1}_{\rm ideal}\right\rangle,U_r\left|v^{r-1}\right\rangle\right)+D_I\left(U_r\left|v^{r-1}\right\rangle,|v^r\rangle\right)\\&=D_I\left(\left|v^{r-1}_{\rm ideal}\right\rangle,\left|v^{r-1}\right\rangle\right)+D_I\left(|v^r_0\rangle,|v^r\rangle\right),\end{aligned} \tag{C.1}$$

for $r\geq 1$, where U_r is the unitary transformation correspoinding to the r-th ideal update of the parameters by (4.23), and $|v^r_0\rangle\equiv U_r\left|v^{r-1}\right\rangle$. By applying (C.1) repeatedly, we obtain

$$D_I\left(\left|v^N_{\rm ideal}\right\rangle,\left|v^N\right\rangle\right)\leq\sum_{r=1}^{N}D_I\left(|v^r_0\rangle,|v^r\rangle\right). \tag{C.2}$$

Let

$$\delta\vec{\theta}^r\equiv\left.\frac{\partial\vec{\theta}(t)}{\partial t}\right|_{t=r\delta t}-\left.\frac{\partial\vec{\theta}_{\rm ideal}(t)}{\partial t}\right|_{t=r\delta t}. \tag{C.3}$$

Then, the direct calculation shows [9]

$$D_I(|v^r_0\rangle,|v^r\rangle)=\sqrt{\delta(\vec{\theta^r})^TW^r\delta\vec{\theta^r}(\delta t)^2+O((\delta t)^3)}, \tag{C.4}$$

where each element of the matrix W^r is given by

$$W^r_{jk}=\frac{\partial\left\langle v^{r-1}\right|}{\partial\theta_j}\frac{\partial\left|v^{r-1}\right\rangle}{\partial\theta_k}-\frac{\partial\left\langle v^{r-1}\right|}{\partial\theta_j}\left|v^{r-1}\right\rangle\left\langle v^{r-1}\right|\frac{\partial\left|v^{r-1}\right\rangle}{\partial\theta_k}. \tag{C.5}$$

Here, we use $\frac{\partial\langle v^{r-1}|}{\partial t}|v^{r-1}\rangle=-\langle v^{r-1}|\frac{\partial|v^{r-1}\rangle}{\partial t}$, which is derived from $\frac{\partial}{\partial t}\langle v^{r-1}|v^{r-1}\rangle=0$. By ignoring the $O((\delta t)^3)$ term inside the square root function, we obtain

$$\begin{aligned}D_I\left(\left|v^N_{\rm ideal}\right\rangle,\left|v^N\right\rangle\right)&\lesssim\sum_{r=1}^{N}\delta t\sqrt{\delta(\vec{\theta^r})^TW^r\delta\vec{\theta^r}}\\&\leq T\sqrt{||W||_{\max}}||\delta\dot{\vec{\theta}}||_{\max},\end{aligned} \tag{C.6}$$

where $||W||_{\max} = \max_r ||W^r||_F$ and $||\delta\dot{\vec{\theta}}||_{\max} = \max_{0\leq t\leq T} ||\delta\dot{\vec{\theta}}(t)||_2$.

The value $\delta\vec{\theta}(t)$ is related with the amount of the shot noise. By substituting $(M+\delta M)^{-1} = M^{-1} - M^{-1}\delta M M^{-1} + O((\delta M)^2)$ in (4.24), we obtain

$$\frac{\partial\vec{\theta}(t)}{\partial t} \simeq (M^{-1}\vec{V} + M^{-1}\delta\vec{V} - M^{-1}\delta M M^{-1}\vec{V}), \tag{C.7}$$

where we ignore the higher order terms of δM and δV with the assumption that δM and δV are sufficiently small. Then,

$$\begin{aligned} |\delta\dot{\vec{\theta}}(t)| &= \left| \frac{\partial\vec{\theta}(t)}{\partial t} - M^{-1}\vec{V} \right| \\ &\simeq \left| M^{-1}\delta\vec{V} - M^{-1}\delta M M^{-1}\vec{V} \right| \\ &\leq ||M^{-1}||_F ||\delta\vec{V}||_2 + ||M^{-1}||_F^2 ||\delta M||_F ||\vec{V}||_2, \end{aligned} \tag{C.8}$$

where $||\cdot||_F$ is the Frobenius norm and $||\cdot||_2$ is the two-norm. Therefore, let

$$\Delta_{MV} \equiv ||M^{-1}||_F ||\delta\vec{V}||_2 + ||M^{-1}||_F^2 ||\delta M||_F ||\vec{V}||_2, \tag{C.9}$$

then

$$D_I\left(\left|v_{\text{ideal}}^N\right\rangle, \left|v^N\right\rangle\right) \lesssim T\sqrt{||W||_{\max}}||\Delta_{MV}||_{\max}. \tag{C.10}$$

The value Δ_{MV} depends on how to estimate M and $\vec{V}$ as we see in the following.

The infidelity in the case of using naive strategy

In case we naively measure each term of M and V one by one, from the central limit theorem,

$$||\delta M||_F \simeq \frac{\sqrt{\sum_{k\ell}(\Delta M_{k\ell}^{\text{naive}})^2}}{\sqrt{N_{\text{naive}}}},\ ||\delta\vec{V}||_2 \simeq \frac{\sqrt{\sum_k(\Delta V_k^{\text{naive}})^2}}{\sqrt{N_{\text{naive}}}}, \tag{C.11}$$

where we use (4.21). Therefore, we obtain

$$\Delta_{MV} \simeq \frac{\Delta_{\text{naive}}}{\sqrt{N_{\text{naive}}}}, \tag{C.12}$$

where

$$\Delta_{\text{naive}} = ||M^{-1}||_F \sqrt{\sum_k (\Delta V_k^{\text{naive}})^2} + ||M^{-1}||_F^2 ||\vec{V}||_2 \sqrt{\sum_{k\ell}(\Delta M_{k\ell}^{\text{naive}})^2}. \tag{C.13}$$

Thus, we obtain

$$D_I\left(\left|v_{\text{ideal}}^N\right\rangle, \left|v^N\right\rangle\right) \lesssim D_I^{\text{naive}},\ \textit{with}\ D_I^{\text{naive}} = T\sqrt{||W||_{\max}}\Delta_{\text{naive}}^{\max}/\sqrt{N_{\text{naive}}}, \tag{C.14}$$

where $\Delta_{\max}^{\text{naive}} \equiv ||\Delta_{\text{naive}}||_{\max}$.

The infidelity in the case of using the shadow-based strategy

Similarly, in case of shadow-based measurements, we obtain

$$||\delta M||_F \simeq \frac{\sqrt{\sum_{k\ell}(\Delta M_{k\ell})^2}}{\sqrt{N_{\text{shot}}}},\ ||\delta\vec{V}||_2 \simeq \frac{\sqrt{\sum_k(\Delta V_k)^2}}{\sqrt{N_{\text{shot}}}}, \tag{C.15}$$

from the central limit theorem, where we use (4.19) and (4.20). Thus,

$$\Delta_{MV} \simeq \frac{\Delta_{\text{shadow}}}{\sqrt{N_{\text{shot}}}}, \tag{C.16}$$

where

$$\Delta_{\rm shadow} = ||M^{-1}||_F \sqrt{\sum_k (\Delta V_k)^2} + ||M^{-1}||_F^2 ||\vec{V}||_2 \sqrt{\sum_{k\ell} (\Delta M_{k\ell})^2}. \tag{C.17}$$

As a result,

$$D_I\left(\left|v_{\rm ideal}^N\right\rangle, \left|v^N\right\rangle\right) \lesssim D_I^{\rm shadow} \ with \ D_I^{\rm shadow} = T\sqrt{||W||_{\rm max}}\Delta_{\rm shadow}^{\rm max}/\sqrt{N_{\rm shot}}, \tag{C.18}$$

where $\Delta_{\rm max}^{\rm shadow} \equiv \max_{0\le t\le T}(\Delta_{\rm shadow})$. From the equations (C.14) and (C.18), we see that $D_I^{\rm naive} = D_I^{\rm shadow}$ holds when

$$\frac{N_{\rm shot}}{N_{\rm naive}} \simeq \left(\frac{\Delta_{\rm shadow}^{\rm max}}{\Delta_{\rm naive}^{\rm max}}\right)^2, \tag{C.19}$$

where we assume that the value of $||W||_{\rm max}$ is almost the same in the naive strategy and the shadow-based strategy.

The infidelity in the case of using the hybrid strategy

In case of the hybrid strategy, where we use in case $N_{\rm BB} = 1$, the number of measurements for estimating each $M_{k\ell}$ can be smaller than that for estimating V_k ($=N_{\rm shot}$); we set the number of measurements as $N_{\rm shot}\alpha$ with $0 < \alpha < 1$. Then, in the hybrid strategy,

$$||\delta M||_F \simeq \frac{\sqrt{\sum_{k\ell}(\Delta M_{k\ell}^{\rm naive})^2}}{\sqrt{N_{\rm shot}\alpha}}, \ ||\delta\vec{V}||_2 \simeq \frac{\sqrt{\sum_k (\Delta V_k)^2}}{\sqrt{N_{\rm shot}}}, \tag{C.20}$$

from the central limit theorem. We obtain

$$\Delta_{MV} \simeq \frac{\Delta_{\rm hybrid}}{\sqrt{N_{\rm shot}}}, \tag{C.21}$$

where

$$\Delta_{\rm hybrid} = ||M^{-1}||_F \sqrt{\sum_k (\Delta V_k)^2} + ||M^{-1}||_F^2 ||\vec{V}||_2 \sqrt{\frac{\sum_{k\ell}(\Delta M_{k\ell}^{\rm naive})^2}{\alpha}}. \tag{C.22}$$

Then,

$$D_I\left(\left|v_{\rm ideal}^N\right\rangle, \left|v^N\right\rangle\right) \lesssim D_I^{\rm hybrid} \ with \ D_I^{\rm hybrid} = T\sqrt{||W||_{\rm max}}\Delta_{\rm hybrid}^{\rm max}/\sqrt{N_{\rm shot}}, \tag{C.23}$$

where $\Delta_{\rm max}^{\rm hybrid} \equiv \max_{0\le t\le T}(\Delta_{\rm hybrid})$. From the equations (C.14) and (C.23), we can show that $D_I^{\rm naive} = D_I^{\rm hybrid}$ holds when

$$\frac{N_{\rm shot}}{N_{\rm naive}} \simeq \left(\frac{\Delta_{\rm hybrid}^{\rm max}}{\Delta_{\rm naive}^{\rm max}}\right)^2, \tag{C.24}$$

where we assume that the value of $||W||_{\rm max}$ is almost the same in the naive strategy and the hybrid strategy.

D Derivation of the Haar averages

In this section, we derive results of the Haar integration. We use the following itegration formulae [51]:

$$\int_{\rm Haar} dU {\rm Tr}(AUBU^\dagger) = \frac{{\rm Tr}(A){\rm Tr}({\rm B})}{N}, \tag{D.1}$$

$$\int_{\rm Haar} dU {\rm Tr}(AUBU^\dagger){\rm Tr}(CUDU^\dagger) = \frac{{\rm Tr}(A){\rm Tr}(B){\rm Tr}(C){\rm Tr}(D) + {\rm Tr}(AC){\rm Tr}(BD)}{N(N+1)}, \tag{D.2}$$

$$\int_{\rm Haar} dU \left({\rm Tr}(AU|0\rangle\langle 0|U^\dagger)\right)^4 = \frac{6{\rm Tr}(A^4) + 3[{\rm Tr}(A^2)]^2}{N^4 + 6N^3 + 11N^2 + 6N}, \tag{D.3}$$

where N is the dimension of the Hilbert space of U.

Derivation of (4.33)

By using (D.1) with $A = \tilde{P}_r\tilde{P}_{r'}$ and $B = |0\rangle\langle 0|$, we obtain

$$\begin{aligned}\langle \mathcal{V}_k^i\rangle_{\text{Haar}} &= \sum_{r=1}^{N_{\text{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)} + \int_{\text{Haar}} dU \sum_{r=1}^{N_{\text{BA}}}\sum_{r'\neq r} G_{k,i,r}G_{k,i,r'}g(\tilde{P}_r,\tilde{P}_{r'})\text{Tr}(\tilde{P}_r\tilde{P}_{r'}U|0\rangle\langle 0|U^\dagger),\\ &= \sum_{r=1}^{N_{\text{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)} + \sum_{r=1}^{N_{\text{BA}}}\sum_{r'\neq r} G_{k,i,r}G_{k,i,r'}g(\tilde{P}_r,\tilde{P}_{r'})\frac{\text{Tr}(\tilde{P}_r\tilde{P}_{r'})}{2^{n+1}}.\end{aligned} \tag{D.4}$$

Since $\tilde{P}_r \neq \tilde{P}_{r'}$, it holds $\tilde{P}_r\tilde{P}_{r'} \neq I$, and therefore, $\text{Tr}(\tilde{P}_r\tilde{P}_{r'}) = 0$. As a result,

$$\langle \mathcal{V}_k^i\rangle_{\text{Haar}} = \sum_{r=1}^{N_{\text{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)}. \tag{D.5}$$

By using (D.2) with $A = \tilde{P}_r\tilde{P}_{r'}$, $B = \tilde{P}_u\tilde{P}_{u'}$, and $C, D = |0\rangle\langle 0|$.

$$\begin{aligned}\text{Var}(\mathcal{V}_k^i)_{\text{Haar}} &= \left(\sum_{r=1}^{N_{\text{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)} + \sum_{r=1}^{N_{\text{BA}}}\sum_{r'\neq r} G_{k,i,r}G_{k,i,r'}g(\tilde{P}_r,\tilde{P}_{r'})\text{Tr}(\tilde{P}_r\tilde{P}_{r'}U|0\rangle\langle 0|U^\dagger)\right)^2 - \left(\sum_{r=1}^{N_{\text{BA}}} \frac{(G_{k,i,r})^2}{q(P_r)}\right)^2\\ &= \sum_{r=1}^{N_{\text{BA}}}\sum_{r'\neq r}\sum_{u=1}^{N_{\text{BA}}}\sum_{u'\neq u} G_{k,i,r}G_{k,i,r'}G_{k,i,u}G_{k,i,u'}g(\tilde{P}_r,\tilde{P}_{r'})g(\tilde{P}_u,\tilde{P}_{u'})\frac{\text{Tr}(\tilde{P}_r\tilde{P}_{r'})\text{Tr}(\tilde{P}_u\tilde{P}_{u'}) + \text{Tr}(\tilde{P}_r\tilde{P}_{r'}\tilde{P}_u\tilde{P}_{u'})}{2^{n+1}(2^{n+1}+1)}\\ &= \sum_{r=1}^{N_{\text{BA}}}\sum_{r'\neq r}\sum_{u=1}^{N_{\text{BA}}}\sum_{u'\neq u} G_{k,i,r}G_{k,i,r'}G_{k,i,u}G_{k,i,u'}g(\tilde{P}_r,\tilde{P}_{r'})g(\tilde{P}_u,\tilde{P}_{u'})\frac{\text{Tr}(\tilde{P}_r\tilde{P}_{r'}\tilde{P}_u\tilde{P}_{u'})}{2^{n+1}(2^{n+1}+1)},\end{aligned} \tag{D.6}$$

where in the last equality, we use $\text{Tr}(\tilde{P}_r\tilde{P}_{r'}) = 0$.

Derivation of (4.38)

By using (D.2), with $A, C = \tilde{P}_r$ and $B, D = |0\rangle\langle 0|$,

$$\begin{aligned}\langle \mathcal{G}_k^{i,r}\rangle_{\text{Haar}} &= (G_{k,i,r})^2 \int_{\text{Haar}} dU \left(1 - \left(\text{Tr}(\tilde{P}_r U|0\rangle\langle 0|U^\dagger)\right)^2\right)\\ &= (G_{k,i,r})^2 \left(1 - \frac{(\text{Tr}(\tilde{P}_r))^2 + (\text{Tr}(\tilde{P}_r{}^2))}{2^{n+1}(2^{n+1}+1)}\right)\\ &= (G_{k,i,r})^2 \left(1 - \frac{1}{2^{n+1}+1}\right).\end{aligned} \tag{D.7}$$

By using (D.3) with $A = \tilde{P}_r$,

$$\begin{aligned}\text{Var}(\mathcal{G}_k^{i,r}) &= (G_{k,i,r})^4 \int_{\text{Haar}} dU \left(1 - \left(\text{Tr}(\tilde{P}_r U|0\rangle\langle 0|U^\dagger)\right)^2\right)^2 - (\langle \mathcal{G}_k^{i,r}\rangle_{\text{Haar}})^2\\ &= (G_{k,i,r})^4 \left[\frac{6\text{Tr}(\tilde{P}_r^4) + 3[\text{Tr}(\tilde{P}_r^2)]^2}{2^{4(n+1)} + 6\cdot 2^{3(n+1)} + 11\cdot 2^{2(n+1)} + 6\cdot 2^{n+1}} + \frac{2\cdot 2^{n+1} - 1}{(2^{2(n+1)}+1)^2}\right]\\ &= \frac{(G_{k,i,r})^4(2\cdot 2^{2(n+1)} + 8\cdot 2^{2(n+1)})}{(2^{n+1}+1)^2(2^{n+1}+3)}.\end{aligned} \tag{D.8}$$

References

[1] John Preskill. "Quantum computing in the NISQ era and beyond". Quantum **2**, 79 (2018).

[2] Suguru Endo, Zhenyu Cai, Simon C Benjamin, and Xiao Yuan. "Hybrid quantum-classical algorithms and quantum error mitigation". Journal of the Physical Society of Japan **90**, 032001 (2021).

[3] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. "Variational quantum algorithms". Nature Reviews Physics **3**, 625–644 (2021).

[4] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm" (2014). arXiv:1411.4028.

[5] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. "A variational eigenvalue solver on a photonic quantum processor". Nature communications **5**, 1–7 (2014).

[6] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry Chow, and Jay Gambetta. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". Nature **549**, 242–246 (2017).

[7] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, et al. "Quantum optimization using variational algorithms on near-term quantum devices". Quantum Science and Technology **3**, 030503 (2018).

[8] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. "The theory of variational hybrid quantum-classical algorithms". New Journal of Physics **18**, 023023 (2016).

[9] Ying Li and Simon C Benjamin. "Efficient variational quantum simulator incorporating active error minimization". Physical Review X **7**, 021050 (2017).

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

[11] Raphael Kaubruegger, Pietro Silvi, Christian Kokail, Rick van Bijnen, Ana Maria Rey, Jun Ye, Adam M Kaufman, and Peter Zoller. "Variational spin-squeezing algorithms on programmable quantum sensors". Physical review letters **123**, 260505 (2019).

[12] Bálint Koczor, Suguru Endo, Tyson Jones, Yuichiro Matsuzaki, and Simon C Benjamin. "Variational-state quantum metrology". New Journal of Physics **22**, 083038 (2020).

[13] Sam McArdle, Tyson Jones, Suguru Endo, Ying Li, Simon C Benjamin, and Xiao Yuan. "Variational ansatz-based quantum simulation of imaginary time evolution". npj Quantum Information **5**, 1–6 (2019).

[14] Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C Benjamin. "Theory of variational quantum simulation". Quantum **3**, 191 (2019).

[15] AD McLachlan. "A variational solution of the time-dependent Schrodinger equation". Molecular Physics **8**, 39–44 (1964).

[16] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F Izmaylov. "Measurement optimization in the variational quantum eigensolver using a minimum clique cover". The Journal of chemical physics **152**, 124114 (2020).

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

[18] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. "Measurements of quantum hamiltonians with locally-biased classical shadows". Communications in Mathematical Physics **391**, 951–967 (2022).

[19] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Efficient estimation of pauli observables by derandomization". Physical review letters **127**, 030503 (2021).

[20] Stefan Hillmich, Charles Hadfield, Rudy Raymond, Antonio Mezzacapo, and Robert Wille. "Decision diagrams for quantum measurements with shallow circuits". In 2021 IEEE International Conference on Quantum Computing and Engineering (QCE). Pages 24–34. IEEE (2021).

[21] Bujiao Wu, Jinzhao Sun, Qi Huang, and Xiao Yuan. "Overlapped grouping measurement: A unified framework for measuring quantum states" (2021).

[22] Pranav Gokhale, Olivia Angiuli, Yongshan Ding, Kaiwen Gui, Teague Tomesh, Martin Suchara, Margaret Martonosi, and Frederic T Chong. "Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families" (2019). arXiv:1907.13623.

[23] Artur F Izmaylov, Tzu-Ching Yen, Robert A Lang, and Vladyslav Verteletskyi. "Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method". Journal of chemical theory and computation **16**, 190–195 (2019).

[24] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley. "Efficient quantum measurement of Pauli operators in the presence of finite sampling error". Quantum **5**, 385 (2021).

[25] William J Huggins, Jarrod R McClean, Nicholas C Rubin, Zhang Jiang, Nathan Wiebe, K Birgitta Whaley, and Ryan Babbush. "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers". npj Quantum Information **7**, 1–9 (2021).

[26] Ikko Hamamura and Takashi Imamichi. "Efficient evaluation of quantum observables using entangled measurements". npj Quantum Information **6**, 1–8 (2020).

[27] Sergey Bravyi, Jay M Gambetta, Antonio Mezzacapo, and Kristan Temme. "Tapering off qubits to simulate fermionic Hamiltonians" (2017). arXiv:1701.08213.

[28] Andrew Zhao, Andrew Tranter, William M Kirby, Shu Fay Ung, Akimasa Miyake, and Peter J Love. "Measurement reduction in variational quantum algorithms". Physical Review A **101**, 062322 (2020).

[29] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F Izmaylov. "Measuring all compatible operators in one series of single-qubit measurements using unitary transformations". Journal of chemical theory and computation **16**, 2400–2409 (2020).

[30] Andrew Jena, Scott Genin, and Michele Mosca. "Pauli partitioning with respect to gate sets" (2019). arXiv:1907.07859.

[31] 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" (2022).

[32] Seonghoon Choi, Tzu-Ching Yen, and Artur F Izmaylov. "Improving Quantum Measurements by Introducing "Ghost" Pauli Products". Journal of Chemical Theory and Computation **18**, 7394–7402 (2022).

[33] 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).

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

[35] Gregory Boyd and Bálint Koczor. "Training variational quantum circuits with CoVaR: covariance root finding with classical shadows" (2022).

[36] Ting Zhang, Jinzhao Sun, Xiao-Xu Fang, Xiao-Ming Zhang, Xiao Yuan, and He Lu. "Experimental quantum state measurement with classical shadows". Physical Review Letters **127**, 200501 (2021).

[37] Suguru Endo, Jinzhao Sun, Ying Li, Simon C Benjamin, and Xiao Yuan. "Variational quantum simulation of general processes". Physical Review Letters **125**, 010501 (2020).

[38] P. A. M. Dirac. "Note on exchange phenomena in the thomas atom". Mathematical Proceedings of the Cambridge Philosophical Society **26**, 376–385 (1930).

[39] Âkov Il'ič Frenkel. "Wave mechanics; advanced general theory". Bull. Amer. Math. Soc **41**, 776 (1935).

[40] Peter Kramer and Marcos Saraceno. "Geometry of the time-dependent variational principle in quantum mechanics". In Group Theoretical Methods in Physics: Proceedings of the IX International Colloquium Held at Cocoyoc, México, June 23–27, 1980. Pages 112–121. Springer (2005).

[41] J Broeckhove, L Lathouwers, E Kesteloot, and P Van Leuven. "On the equivalence of time-dependent variational principles". Chem. Phys. Lett. **149**, 547–550 (1988).

[42] Artur F Izmaylov, Tzu-Ching Yen, Robert A Lang, and Vladyslav Verteletskyi. "Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method". Journal of chemical theory and computation **16**, 190–195 (2019).

[43] Andrew Zhao, Andrew Tranter, William M Kirby, Shu Fay Ung, Akimasa Miyake, and Peter J Love. "Measurement reduction in variational quantum algorithms". Physical Review A **101**, 062322 (2020).

[44] Tzu-Ching Yen, Vladyslav Verteletskyi, and Artur F Izmaylov. "Measuring all compatible operators in one series of single-qubit measurements using unitary transformations". Journal of chemical theory and computation **16**, 2400–2409 (2020).

[45] Giacomo Torlai, Guglielmo Mazzola, Giuseppe Carleo, and Antonio Mezzacapo. "Precise measurement of quantum observables with neural-network estimators". Physical Review Research **2**, 022060 (2020).

[46] Ophelia Crawford, Barnaby van Straaten, Daochen Wang, Thomas Parks, Earl Campbell, and Stephen Brierley. "Efficient quantum measurement of Pauli operators in the presence of finite sampling error". Quantum **5**, 385 (2021).

[47] William J Huggins, Jarrod R McClean, Nicholas C Rubin, Zhang Jiang, Nathan Wiebe, K Birgitta Whaley, and Ryan Babbush. "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers". npj Quantum Information **7**, 1–9 (2021).

[48] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. "Evaluating analytic gradients on quantum hardware". Physical Review A **99**, 032331 (2019).

[49] Barnaby van Straaten and Bálint Koczor. "Measurement cost of metric-aware variational quantum algorithms". PRX Quantum **2**, 030324 (2021).

[50] Yasunari Suzuki, Yoshiaki Kawase, Yuya Masumura, Yuria Hiraga, Masahiro Nakadai, Jiabao Chen, Ken M Nakanishi, Kosuke Mitarai, Ryosuke Imai, Shiro Tamiya, et al. "Qulacs: a fast and versatile quantum circuit simulator for research purpose". Quantum **5**, 559 (2021).

[51] Benoît Collins and Piotr Śniady. "Integration with respect to the Haar measure on unitary, orthogonal and symplectic group". Communications in Mathematical Physics **264**, 773–795 (2006).