# Quantum Computing Quantum Monte Carlo

Yukun Zhang,[1,2,*] Yifei Huang,[3,*] Jinzhao Sun,[4,5] Dingshun Lv,[3] and Xiao Yuan[1,2,†]

[1] *Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China*
[2] *School of Computer Science, Peking University, Beijing 100871, China*
[3] *ByteDance Ltd., Zhonghang Plaza, No. 43, North 3rd Ring West Road, Haidian District, Beijing, China*
[4] *Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom*
[5] *Quantum Advantage Research, Beijing 100080, China*

## I. BACKGROUND

In this section, we give a more detailed background introduction to variational quantum algorithms and quantum Monte Carlo.

### A. Variational Quantum Eigensolver

We focus on variational quantum eigensolver (VQE) as a typical example of general variational quantum algorithms [1–5]. VQE is a hybrid quantum-classical method that utilizes quantum devices to prepare parameterized quantum circuits and optimize the parameters classically [6]. Through encoding the cost function, usually the corresponding Hamiltonian for the system of interest, one is able to estimate its expectation value by measuring the quantum state. In the following discussion, we focus on fermionic Hamiltonians. Yet, our results apply for general Hamiltonians. For a fermionic Hamiltonian, we use the Jordan-Wigner encoding which encodes the occupation numbers of the spin orbitals to qubits. The alternatives to JW encoding include the parity transformation and Bravyi-Kitaev transformation [7, 8].

Several carefully designed ansatz is proposed for the specific quantum systems concerning their unique properties. In this work, we focus on two types of algorithms: the ADAPT-VQE [9], and the Hamiltonian Variational Ansatz (HVA) [10–12].

ADAPT-VQE algorithms have been put forward as general protocols for adaptively growing the ansatz. The ADAPT-VQE methods build the quantum circuit iteratively by adding the operator from the operator pool with the largest gradient with respect to the loss function of the current circuit at every iteration. After adding a new operator to the circuit, one variationally optimizes the new parameter while also re-optimize all previous parameters. An iteration terminates when the norm of the gradient is smaller than a certain threshold. Several variations for this algorithm have been proposed with distinct features. The most commonly adopted operator pool is constructed according to the unitary coupled cluster ansatz with single and double excitations (UCCSD) [13]. For a molecular non-relativistic Hamiltonian of the form

$$H = H_0 + \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \tag{1}$$

where $h_{pq}$ and $h_{pqrs}$ are coefficients obtained from one- and two-electron integrals, the coupled cluster method with single and double excitation heuristic truncated the higher excitations and gives the following ansatz based upon Hartree-Fock orbitals:

$$|\psi_{CCSD}\rangle = e^{T_1+T_2} |\psi_{HF}\rangle . \tag{2}$$

Here $T_1$ and $T_2$ correspond to single and double excitations:

$$T_1 = \sum_{i\in occ, k\in virt} t_i^k a_i^\dagger a_k \tag{3}$$

---

* These two authors contributed equally to this work.
† xiaoyuan@pku.edu.cn

and

$$T_2 = \sum_{i>j\in occ, k>l\in virt} t_{ij}^{kl} a_i^\dagger a_j^\dagger a_k a_l \tag{4}$$

The operator $e^T$ cannot be implemented on a quantum computer due to its non-unitarity. However, its variant $e^{T-T^\dagger}$ is unitary for a given excitation operator $T$. This is dubbed the unitary coupled cluster ansatz. Therefore, the operator pool for the ADAPT-VQE ansatz is formed by the single and double excitation operators from $T_1$ and $T_2$ but every term has to subtract its conjugate transpose. Later we will use this ansatz to construct the VQE basis for the nitrogen molecule.

The HVA ansatz is particularly suitable for the Hubbard model. The Hubbard models are celebrated for modeling strong-correlated electronic effects in many-body quantum physics. The model captures the competition between electronic hopping and on-site repulsion, which as the form

$$H = -\sum_{i,j} t_{ij} \sum_{\sigma=\uparrow,\downarrow} a_{i\sigma}^\dagger a_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}, \tag{5}$$

where $t$ and $U$ is the hopping and repulsion strength, and $n_i$ denotes the number operator. For a square lattice, the HVA algorithm separates the Hamiltonian into three parts $H = H_v + H_h + H_U$. The former two terms are the hopping terms in Eq. (5) in the vertical and horizontal direction, and the last one is the on-site term. For a reference state $|\psi_{\text{ref}}\rangle$, the ansatz takes the following form

$$\Psi = \prod_{b=1}^{B} \left[ \exp\left(i\frac{\theta_U^b}{2} H_U\right) \exp\left(i\frac{\theta_h^b}{2} H_h\right) \exp\left(i\frac{\theta_v^b}{2} H_v\right) \exp\left(i\frac{\theta_U^b}{2} H_U\right) \right] |\psi_{\text{ref}}\rangle \,, \tag{6}$$

where $\theta$ are parameters for the ansatz, and $B$ is the total interation. With recent progress [14], one notices that the one-body terms in the ansatz can be compiled exactly without trotterization for they have the fermionic Gaussian form.

### B. Quantum Monte Carlo

Quantum Monte Carlo (QMC) is a class of classical computation algorithms that invokes the Monte Carlo method to approximate properties of quantum many-body systems. Projector quantum Monte Carlo is a subclass of QMC attempting to stochastically apply walkers to follow the imaginary-time evolution, which however generally causes the infamous sign problem in fermionic systems. The sign problem can be alleviated through different approximations, such as the fix-node approximation in Diffusion Monte Carlo (DMC) and the phaseless approximation in auxiliary-field quantum Monte Carlo (AFQMC). However, most techniques introduce bias while trying to ease the sign problem. In this work, we seek to alleviate the sign problem without introducing bias to the solution by combining quantum computing with a QMC algorithm.

A hybrid algorithm of AFQMC and quantum computing has been proposed in Ref. [15] where they prepared a non-trivial trial state for AFQMC with the quantum device. Shadow tomography is implemented on the quantum device to calculate the overlaps between the trial state and stochastic walker states, which are the values required for energy estimation. These overlaps are easy to estimate to an additive error and friendly for virtual correlation energy calculation without overhead on quantum devices. However, estimation of ground state energy to an additive error with the scheme in its current form requires overlap estimation to a relative error, which can be hard when some of the overlaps vanish exponentially fast as the size of the system increases. Here in this work, we circumvent the overlap estimation by combining quantum computing with another type of projector QMC algorithm in the space of Slater determinants, called the full-configuration interaction quantum Monte Carlo (FCIQMC) algorithm [16, 17]. The FCIQMC algorithm pursues FCI-level accuracy estimation of ground-state properties with fewer resources. Now we describe its main procedure as follows.

Suppose that the wavefunction is decomposed as follows,

$$|\psi(\tau)\rangle = \sum_i c_i(\tau) |i\rangle \tag{7}$$

where $\tau$ is the imaginary time we are evolving and $|i\rangle$s are the Fock basis states, which corresponds to the single determinant configurations in the fermionic representation. Substituting this into the imaginary time Schrodinger equation

$$|\psi(\tau)\rangle \propto e^{-H\tau} |\psi(0)\rangle \,, \tag{8}$$

taking the first-order approximation, one obtains a set of coupled differential equations for the evolution of coefficients $c_i$s:

$$-\frac{\mathrm{d}c_i(\tau)}{\mathrm{d}\tau} = \sum_j (H_{ij} - S\delta_{ij})c_j(\tau) \tag{9}$$

where $H_{ij}$s are the Hamiltonian matrix elements under a certain basis (here the Fock basis) and $S$ is an additional energy shift on the diagonal terms. One usually starts with the Hartree-Fock energy and adjusts the value along the evolution for the purpose of walker population control. Empirical formulas are proposed in Ref. [16] for the adjustment of $S$ as follows

$$S(\tau) = S(\tau - A\Delta\tau) - \frac{\zeta}{A\Delta\tau}\ln\frac{N(\tau)}{N(\tau - A\delta\tau)}, \tag{10}$$

where $N$ is the total walker number at a specific time, $A$ and $\zeta$ are values one can adjust according to the system of interest.

FCIQMC considers a set of walkers, who live in computational basis vector space with coefficients being $\pm 1$. The population $N_i(\tau)$ at each basis vector $|i\rangle$ is therefore an integer, obtained by the signed sum of all walkers located at this basis vector. One needs to design the QMC algorithm such that $N_i(\tau)$ is proportional to the corresponding amplitude $c_i(\tau)$ following the imaginary time dynamics. Hence, if one trotterize the time evolution into steps of size $\Delta\tau$, the FCIQMC algorithm consists of the following three steps at each time step according to eq. (47):

- Spawning: For each walker living in $|i\rangle$, spawn a child walker $|j\rangle$ $(j \neq i)$ with probability $|H_{ij}|\Delta\tau$ up to a normalization constant. Moreover, if this quantity $H_{ij}$ is positive, then the child has the same sign as the parent, and the opposite sign otherwise.
- Death/clone: If the quantity $H_{ii} - S > 0$, then walker $|\phi_i\rangle$ dies with probability $(H_{ii} - S)\Delta\tau$; and if $H_{ii} - S < 0$, then the walker clones itself with probability $|(H_{ii} - S)|\Delta\tau$.
- Annihilation: As we have alluded to, we want the signed sum of walkers to be proportional to the true amplitudes. Therefore, at the end of each time step, we take the signed sum of walkers at each basis vector and annihilate every pair of walkers with opposite signs.

## II. SIGN PROBLEM IN QUANTUM MONTE CARLO

In this section, we give a more detailed analysis of the sign problem in QMC. We consider the non-stoquasticity indicator and derive its upper bounds for different cases.

### A. Definition and Mitigation of The Sign Problem

The projector QMC methods are plagued by the sign problem. We briefly discuss the implications of the sign problem in the main text, and here we give a self-contained introduction and proofs of the bounds of the sign problem. As we intend to implement the imaginary time evolution (ITE) operator to an initial state, we have to normalize the resultant state due to the non-unitarity of the ITE operator. In cases of the projector QMC methods, we represent the overall wave function as a distribution of walker states over the chosen basis, and the walkers evolve upon implementing the ITE operator. The sign problem manifests when computing the expectation value of observable with the final-round walkers as we normalize the wave function in the chosen basis. To see this, let us compute the partition function of the ITE operator on the chosen basis. Here, we first consider the Fock basis for fermionic systems, and we get

$$\begin{aligned}\mathrm{Tr}(e^{-\beta H}) &= \sum_{k=0}^{\infty}\frac{(-\beta)^k}{k!}\mathrm{Tr}(H^k) \\ &= \sum_{k=0}^{\infty}\frac{(-\beta)^k}{k!}\sum_{\{|\zeta\rangle\}}\langle\zeta_0|\,H\,|\zeta_1\rangle\langle\zeta_1|\,H\,|\zeta_2\rangle\cdots\langle\zeta_{k-1}|\,H\,|\zeta_0\rangle\,.\end{aligned} \tag{11}$$

Here, $\{|\zeta\rangle\}$ is a set of complete orthogonal bases, i.e., the Fock basis in our case. Eq. (11) gives us the path integral form of the partition function, which consists of all cyclic paths in the chosen basis. One notices that both positive

and negative paths could be contained in Eq. (11). By definition of the sign problem, the system is sign-problem-free if all the paths are positive in Eq. (11) [18]. In particular, fermionic systems are known to be notoriously affected by the sign problem because the cyclic paths include cases where odd pairs of particles are exchanged, which changes the sign of the paths. The standard way to deal with the negative path is to sample according to its absolute value in Monte Carlo simulation. To quantify the sign problem, we follow the deduction of Troyer and Wiese [18] and consider computing the expectation value of the operator $A$ with respect to the thermal state in a similar form

$$\begin{aligned}\langle A\rangle &= \frac{\mathrm{Tr}[Ae^{-\beta H}]}{\mathrm{Tr}[e^{-\beta H}]} = \frac{\mathrm{Tr}[Ae^{\beta G}]}{\mathrm{Tr}[e^{\beta G}]} = \frac{\sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{i_0,i_1,\ldots i_k} A_{i_0 i_1}G_{i_1 i_2}\cdots G_{i_k i_0}}{\sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{i_1,i_2\ldots i_k} G_{i_1 i_2}G_{i_2 i_3}\cdots G_{i_k i_1}} \\ &= \frac{\sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{i_0,i_1,\ldots i_k} A_{i_0 i_1}|G_{i_1 i_2}\cdots G_{i_k i_0}|s_{i_1\cdots i_k,i_0}/\sum_{i_1,i_2\ldots i_k}|G_{i_1 i_2}\cdots G_{i_k i_0}|}{\sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{i_1,i_2\ldots i_k}|G_{i_1 i_2}G_{i_2 i_3}\cdots G_{i_k i_1}|s_{i_1\cdots i_k,i_1}/\sum_{i_1,i_2\ldots i_k}|G_{i_1 i_2}\cdots G_{i_k i_1}|} = \frac{\langle As\rangle}{\langle s\rangle},\end{aligned} \tag{12}$$

where $G$ is the same as in the main text, which equals to $\alpha I - H$ with $\alpha = \max_i H_{ii}$, and we extract the sign of $G_{i_1 i_2}G_{i_2 i_3}\cdots G_{i_k i_1}$ as $s_{i_1\cdots i_k,i_1}$. In cases where all off-diagonal terms in the Hamiltonian are non-positive, the system is sign-problem-free as can be seen from Eq. (12), and these kinds of Hamiltonians are dubbed as "stoquastic" [19] or "Bosonic". We notice that the denominator is the average sign $\langle s\rangle$ in Eq. (12) is the ratio of the partition function and its bosonic counterpart. By definition of the partition function that is exponential of the free energy, the average sign goes down exponentially as $\exp(-\beta n\Delta f)$, where $\Delta f$ is the free energy difference between the original and its bosonic form system. Upon computing the expectation of an observable, the estimator is enlarged exponentially by the partition function, so that its variance

$$\frac{\Delta s}{\langle s\rangle} = \frac{\sqrt{\left(\langle s^2\rangle - \langle s\rangle^2\right)/M}}{\langle s\rangle} = \frac{\sqrt{1-\langle s\rangle^2}}{\sqrt{M}\langle s\rangle} \propto \frac{e^{\beta n\Delta f}}{\sqrt{M}}. \tag{13}$$

This means that the number $M$ of sampled walkers needs to grow exponentially to counterbalance the variance growth. In practice, because we can only afford a finite size of samples, the sign problem often manifests itself as great statistical fluctuation in the observable estimation step. Note that it is generally hard to resolve the sign problem as Troyer and Wiese [18] proved that the problem is **NP**-complete. In certain cases, the sign problem can be shown to be removed by a local transformation or basis rotation. However, the intrinsic hardness of solving the sign problem is characterized by their topological nature [20], and hence local transformations are not capable of curing the sign problem in the most general cases.

It is natural to think of easier ways to quantify the sign problem other than the path integral form in Eq. (11), and ways to mitigate the sign problem. It is also of independent interest to develop measures for the sign problem as they serve as loss functions for mitigating the sign problem [21, 22], which is of great significance for extending numerical ability in classical QMC simulation of quantum many-body systems and materials. The criteria for the measure of the sign problem include i) faithfully reflect the severity of the sign problem for a Hamiltonian in a given basis, e.g. deviation of the sample complexity for the Hamiltonian from that of the effective bosonic form Hamiltonian; ii) general enough that it is independent of the details of implantation of the QMC algorithms; iii) efficiently computable. The non-stoquasticity, being directly related to the Hamiltonian and a computationally tractable description of the sign problem, is adopted as the measure in Ref. [21]. The authors also provide numerical evidence with randomly generated Hamiltonians to support the statement that the non-stoquasticity measures the severity of the sign problem. However, as pointed out in Ref. [23] the off-diagonal terms do not directly induce the sign problem, but the cumulative phases of the cyclic paths in the path-integral expression do. Unfortunately, it is left unanswered to what extent the non-stoquasticity (quantitatively) contributes to the sign problem. Furthermore, it is desirable to propose a measure other than the cumulative phases as there are exponentially many of them. In the next section, we derive rigorously the measure of the sign problem following the criteria discussed above. It is found that the upper bound for the overall effect of path-integral for the negative cyclic paths (i.e., the direct origin of the sign problem) adds up to be directly related to the negative and positive off-diagonal terms in the Hamiltonian separately. This renders our measure very much like the properties of non-stoquasticity, and we thus name the measure as non-stoquasticity indicator (NSI). Our result elucidates the relationship between non-stoquasticity and the measure of severity of the sign problem. We discuss this in detail in the next section.

To suppress the sign problem without altering the spectrum of the Hamiltonian, the general way is to implement similarity transformation with a unitary operator $U(\vec{\theta})$ [21]

$$H_T(\vec{\theta}) = U(\vec{\theta})^\dagger H U(\vec{\theta}). \tag{14}$$

Generally speaking, an arbitrary unitary will scramble the original Hamiltonian, i.e., the Hamiltonian will become highly non-local. As a consequence, it is not possible to implement the ITE operator both in classical and quantum

cases. To preserve the locality after the transformation, several proposals [21, 22, 24] have been put forward. For spin systems, local Clifford and special orthogonal group [21, 24] transformation have been considered. The former is due to the fact that the Pauli group is the normal subgroup of the Clifford group, and the latter will only scramble the observable locally while maintaining its elements real. For fermionic systems, the basis rotation transformations [22] are ideal choices as they shall not alter the locality. However, both the three types of transformations are limited in their expressive power, and basically, it is **NP**-complete [21, 24] to find the curing transformation in these settings. The above-mentioned three types of transformations are efficient-simulatable classically. To further extend the capacity of mitigating the sign problem, we suggest thinking of similarity transformation realized by quantum computing-related protocols as quantum circuits are natural building blocks to construct arbitrary unitaries. As we will discuss in the next session, we propose to alter the basis from the Fock basis to the VQE-rotated basis, which is $\{|\phi_i\rangle \equiv U(\vec{\theta})\,|\zeta_i\rangle\}$ with $U(\vec{\theta})$ prepared by the VQE algorithm. Let us consider the path integral form of the partition function in the VQE-rotated basis

$$\begin{aligned}&\sum_{k=0}^{\infty}\frac{(-\beta)^k}{k!}\sum_{\{|\zeta\rangle\}}\langle\zeta_0|\,U(\vec{\theta})^\dagger HU(\vec{\theta})\,|\zeta_1\rangle\,\langle\zeta_1|\,U(\vec{\theta})^\dagger HU(\vec{\theta})\,|\zeta_2\rangle\cdots\langle\zeta_{k-1}|\,U(\vec{\theta})^\dagger HU(\vec{\theta})\,|\zeta_0\rangle\\&=\sum_{k=0}^{\infty}\frac{(-\beta)^k}{k!}\mathrm{Tr}(H_T(\vec{\theta})^k)=\mathrm{Tr}(e^{-\beta H_T(\vec{\theta})}),\end{aligned}\tag{15}$$

which matches the general way to suppress the sign problem as Eq. (14). From this perspective, our algorithms allow an effective implementation of a rather complicated unitary to realize Eq. (14) without actually scrambling the Hamiltonian. Therefore, with the similarity transformation effectively implemented by a quantum circuit, we expect our approach to achieving better performance in easing the sign problem, and we provide quantitative measures for gauging the sign problem in the following.

### B. Non-Stoquasticity Indicator For Real Hamiltonians

For any given Hamiltonian $H$, we refer to $\tilde{H}$ Bosonic form of the original Hamiltonian, which has the following form

$$\begin{aligned}\tilde{H}_{ij}&=-\,|H_{ij}|\,,\qquad \text{if } i\neq j;\\ \tilde{H}_{ij}&=H_{ij},\qquad \text{otherwise.}\end{aligned}\tag{16}$$

To see the relationship between the severity of the sign problem and the non-stoquasticity of the Hamiltonian after the similarity transformation, we derive an easily computing formula by relating it to the partition function difference between $H_T$ and its Bosonic form $\tilde{H}_T$. We refer this quantity to the Non-Stoquasticity Indicator (NSI), $S(H)$ for any given Hamiltonian $H$ spanned by the basis $\{|\phi\rangle\}$, which has the form

$$\begin{aligned}S(H)&=\mathrm{Tr}(e^{-\beta\tilde{H}})-\mathrm{Tr}(e^{-\beta H})\\&=\mathrm{Tr}(e^{-\beta(\alpha I-\tilde{G})})-\mathrm{Tr}(e^{-\beta(\alpha I-G)})\\&=e^{-\beta\alpha}\left(\sum_{k=0}^{\infty}\frac{\beta^k}{k!}\sum_{\{|\phi\rangle\}}\left(|\langle\phi_0|\,G\,|\phi_1\rangle\,\langle\phi_1|\,G\,|\phi_2\rangle\cdots\langle\phi_{k-1}|\,G\,|\phi_0\rangle|-\langle\phi_0|\,G\,|\phi_1\rangle\,\langle\phi_1|\,G\,|\phi_2\rangle\cdots\langle\phi_{k-1}|\,G\,|\phi_0\rangle\right)\right)\end{aligned}\tag{17}$$

Note that the diagonal terms in $G$ are all non-negative. We claim that the NSI is essentially the most general form that accounts for all possible negative paths in the path-integral formula for measuring the severity of the sign problem.

To arrive at the final expression for the NSI, we need to sort out all the negative path integrals in Eq. (17). We define $H_+$ and $H_-$ as

$$\begin{aligned}(H_+)_{ij}&=H_{ij},\text{if}\,H_{ij}>0\ \text{ and }\quad i\neq j;(H_+)_{ij}=0,\text{ otherwise;}\\(H_-)_{ij}&=H_{ij},\text{if}\,H_{ij}<0\ \text{ or }\quad i=j;(H_-)_{ij}=0,\text{ otherwise.}\end{aligned}\tag{18}$$

According to the above definition, the corresponding $G$ will be

$$\begin{aligned}G_+&=\alpha-H_-;\\G_-&=\alpha-H_+.\end{aligned}\tag{19}$$

We first focus on cases where all terms in the given Hamiltonian $H$ are real. The NSI is supposed to be $-2$-fold of sum of all negative path integrals

$$\begin{aligned}
&S(H)/e^{-\beta\alpha}\\
=&-2\sum_{k=0}^{\infty}\frac{(\beta)^{2k+1}}{(2k+1)!}\sum_{\{\eta_1,\eta_2\}}\left(G_-^{(1)}(\eta_1|\eta_2)G_+^{(2k)}(\eta_2|\eta_1)+G_-^{(3)}(\eta_1|\eta_2)G_+^{(2k-2)}(\eta_2|\eta_1)+\cdots+G_-^{(2k+1)}(\eta_1|\eta_1)\right)\\
&-2\sum_{k=0}^{\infty}\frac{(\beta)^{2k}}{(2k)!}\sum_{\{\eta_1,\eta_2\}}\left(G_-^{(1)}(\eta_1|\eta_2)G_+^{(2k-1)}(\eta_2|\eta_1)+G_-^{(3)}(\eta_1|\eta_2)G_+^{(2k-3)}(\eta_2|\eta_1)+\cdots+G_-^{(2k-1)}(\eta_1|\eta_2)G_+^{(1)}(\eta_2|\eta_1)\right),
\end{aligned}\tag{20}$$

where $G_{-/+}^{(i)}$ stands for the path integral that contains $i$ positive/negative elements from matrix $G$, and we sum over all the possible set of paths denote by $\eta_1,\eta_2$ that contains the right amount of positive and negative components. The path integral form is equivalent to choosing elements from $G_-$ and $G_+$, and we have

$$\begin{aligned}
=&2\sum_{k=0}^{\infty}\frac{(\beta)^{2k+1}}{(2k+1)!}\left(\binom{2k+1}{1}\|G_-\|_{L_1}\|G_+^{2k}\|_{L_1}+\binom{2k+1}{3}\|G_-^3\|_{L_1}\|G_+^{2k-2}\|_{L_1}+\cdots+\binom{2k+1}{2k+1}\|G_-^{2k+1}\|_{L_1}\right)\\
&+2\sum_{k=0}^{\infty}\frac{(\beta)^{2k}}{(2k)!}\left(\binom{2k}{1}\|G_-\|_{L_1}\|G_+^{2k-1}\|_{L_1}+\binom{2k}{3}\|G_-^3\|_{L_1}\|G_+^{2k-3}\|_{L_1}+\cdots+\binom{2k}{2k-1}\|G_-^{2k-1}\|_{L_1}\|G_+\|_{L_1}\right).
\end{aligned}\tag{21}$$

For $L_1$ norm of some matrix $x$, its $L_1$ norm of $m$th power is no larger than its $L_1$ norm's $m$th power for any positive integer $m$, which is $\|x^m\|_{L_1}\leq\|x\|_{L_1}^m$. We get the upper bound for the NSI,

$$\begin{aligned}
\leq&2\sum_{k=0}^{\infty}\frac{(\beta)^{2k+1}}{(2k+1)!}\left(\binom{2k+1}{1}\|G_-\|_{L_1}\|G_+\|_{L_1}^{2k}+\binom{2k+1}{3}\|G_-\|_{L_1}^3\|G_+\|_{L_1}^{2k-2}+\cdots+\binom{2k+1}{2k+1}\|G_+\|_{L_1}^{2k+1}\right)\\
&+2\sum_{k=0}^{\infty}\frac{(-\beta)^{2k}}{(2k)!}\left(\binom{2k}{1}\|G_-\|_{L_1}\|G_+\|_{L_1}^{2k-1}+\binom{2k}{3}\|G_-\|_{L_1}^3\|G_+\|_{L_1}^{2k-3}+\cdots+\binom{2k}{2k-1}\|G_-\|_{L_1}^{2k-1}\|G_+\|_{L_1}\right).
\end{aligned}\tag{22}$$

The above equation is just the binomial theorem with only the odd terms. By adjusting the binomial theorem to $\frac{1}{2}((b+a)^n-(b-a)^n)=\sum_{i=1,\text{odd}}^{n}\binom{n}{i}a^xb^{n-x}$, we get

$$\begin{aligned}
=&\sum_{k=0}^{\infty}\frac{(\beta)^{2k+1}}{(2k+1)!}\left((\|G_+\|_{L_1}+\|G_-\|_{L_1})^{2k+1}-(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{2k+1}\right)\\
&+\sum_{k=0}^{\infty}\frac{(\beta)^{2k}}{(2k)!}\left((\|G_+\|_{L_1}+\|G_-\|_{L_1})^{2k}-(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{2k}\right).
\end{aligned}\tag{23}$$

At this point, one finds that the power series in Eq. (23) contains either odd or even terms, which fit the definitions for the hyperbolic sine and cosine functions,

$$\begin{aligned}
=&\sinh\left(\beta(\|G_+\|_{L_1}+\|G_-\|_{L_1})\right)-\sinh(\beta(\|G_+\|_{L_1}-\|G_-\|_{L_1}))+\cosh(\beta(\|G_+\|_{L_1}+\|G_-\|_{L_1}))-\cosh(\beta(\|G_+\|_{L_1}-\|G_-\|_{L_1}))\\
=&2\exp(\beta\|G_+\|_{L_1})\sinh(\beta\|G_-\|_{L_1})\\
=&2\exp(\beta\|\alpha I-H_-\|_{L_1})\sinh(\beta\|H_+\|_{L_1}).
\end{aligned}\tag{24}$$

The bound for the NSI formula in Eq. (24) indicates that the sign problem will become exponentially severe as the off-diagonal terms in the Hamiltonian increase. However, when $\|H_+\|_{L_1}$ reduces to precisely zero, the sign problem disappears, which meets what one would expect for stoquastic Hamiltonian. Our bound for the NSI directly links the sign problem's severity with the off-diagonal terms in the Hamiltonian.

The message we learn from the NSI formula is that a similar transformation should decrease the absolute value for both positive and negative off-diagonal elements simultaneously to mitigate the sign problem.

### C. Non-Stoquasticity Indicator For Complex Hamiltonians

More generally, we consider the situation where the Hamiltonian $H$ is spanned by a complex space, the analysis becomes more complicated. Note that this is the case in the following discussion about the Hubbard model as the

HVA algorithms contain complex amplitudes. The NSI by definition is the difference between the partition function of the Hamiltonian $H$ and its Bosonic form Hamiltonian $\tilde{H}$. Importantly, the diagonal terms of the Hamiltonian should be real in any basis, because they correspond to the energy $E_i = \langle\phi_i| H |\phi_i\rangle$ so that $\alpha = \max_i H_{ii}$ is also a real number. To keep the NSI a real measure, we apply modulus to each trace for expansion of the exponential function, which gives us

$$S(H)/e^{-\beta\alpha} = \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \mathrm{Tr}(\tilde{G}^k) \; - \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \left|\mathrm{Tr}(G^k)\right|. \tag{25}$$

We again expand the trace in the path integral form

$$= \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \sum_{\{|\phi_i\rangle,\forall i\}} \langle\phi_0| \tilde{G} |\phi_1\rangle \langle\phi_1| \tilde{G} |\phi_2\rangle \cdots \langle\phi_{k-1}| \tilde{G} |\phi_0\rangle - \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \left| \sum_{\{|\phi_i\rangle,\forall i\}} \langle\phi_0| G |\phi_1\rangle \langle\phi_1| G |\phi_2\rangle \cdots \langle\phi_{k-1}| G |\phi_0\rangle \right|. \tag{26}$$

We separate the phase factor from the modulus of $H$ for each element in the matrix, $G = |G| \odot P(G)$, where the $\odot$ and $P(G)$ denote element-wise multiplication and the phase factor matrix for $H$, respectively. We have

$$\begin{aligned} =& \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \sum_{\{|\phi_i\rangle,\forall i\}} \tilde{G}_{0,1}\tilde{G}_{1,2}\cdots\tilde{G}_{k-1,0} \cdot P_{0,1}(\tilde{G})P_{1,2}(\tilde{G})\cdots P_{k-1,0}(\tilde{G}) \\ & - \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \left| \sum_{\{|\phi_i\rangle,\forall i\}} |G_{0,1}||G_{1,2}|\cdots|G_{k-1,0}| \cdot P_{0,1}(G)P_{1,2}(G)\cdots P_{k-1,0}(G) \right|, \end{aligned} \tag{27}$$

where we denote the $i$-th row, $j$-th colomn element in matrix $X$ as $X_{i,j}$. Note that every nonzero element in $P(\tilde{G})$ equals 1 in Eq. (27), so we can omit this term. Note also that $\tilde{G}$ equals to $|G|$ so that

$$\begin{aligned} =& \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \sum_{\{|\phi_i\rangle,\forall i\}} \tilde{G}_{0,1}\tilde{G}_{1,2}\cdots\tilde{G}_{k-1,0} - \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \left| \sum_{\{|\phi_i\rangle,\forall i\}} \tilde{G}_{0,1}\tilde{G}_{1,2}\cdots\tilde{G}_{k-1,0} \cdot P_{0,1}(G)P_{1,2}(G)\cdots P_{k-1,0}(G) \right| \\ =& \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \left( \sum_{\{|\phi_i\rangle,\forall i\}} \tilde{G}_{0,1}\tilde{G}_{1,2}\cdots\tilde{G}_{k-1,0} - \left| \sum_{\{|\phi_i\rangle,\forall i\}} \tilde{G}_{0,1}\tilde{G}_{1,2}\cdots\tilde{G}_{k-1,0} \cdot P_{0,1}(G)P_{1,2}(G)\cdots P_{k-1,0}(G) \right| \right). \end{aligned} \tag{28}$$

Notably, the last summation in Eq. (28) involves complex terms that would cancel each out. We loose this condition and attain the upper bound

$$\leq \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \sum_{\{|\phi_i\rangle,\forall i\}} \tilde{G}_{0,1}\tilde{G}_{1,2}\cdots\tilde{G}_{k-1,0} \left|1 - P_{0,1}(G)P_{1,2}(G)\cdots P_{k-1,0}(G)\right|. \tag{29}$$

In the next step, we relax the bound by decoupling the product of $\tilde{G}_{i,j}$ from the modulus term. Also, because all non-zero terms in $\tilde{G}$ are positive, we also apply bounds for its products of elements following the same deduction as Eq. (21). We have

$$\leq \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \|G\|_{L_1}^k \sum_{\{|\phi_i\rangle,\forall i\}} \left|1 - P_{0,1}(G)P_{1,2}(G)\cdots P_{k-1,0}(G)\right|. \tag{30}$$

Now, let us consider two arbitrary phase factors $p_i = e^{i\theta_i}$ and $p_j = e^{i\theta_j}$. We can establish the triangle inequality $|1 - p_i p_j| \leq |1 - p_i| + |1 - p_j|$. By applying the inequality iteratively to the last modulus part of Eq. (30), we extract one phase factor out each time from the product. The resultant formula is

$$\begin{aligned} \leq& \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \|G\|_{L_1}^k \sum_{\{|\phi_i\rangle,\forall i\}} \left( \sum_{r=0}^{k-1} |1 - P_{r,r+1}(G)| \right) \\ =& \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \|G\|_{L_1}^k \sum_{r=0}^{k-1} \left( \sum_{\{|\phi_i\rangle,\forall i\}} |1 - P_{r,r+1}(G)| \right), \end{aligned} \tag{31}$$

where we have implicitly prescribed that the next term for $k-1$ is 0. In the second line of Eq (31), we also exchange the last two summations. By doing so, one finds that the summation runs over all possible elements for two independent basis sets $\{|\phi_r\rangle\}$ and $\{|\phi_{r+1}\rangle\}$. Then, it is obvious that the outside summation will get $k$ equal results. We get

$$= \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \|G\|_{L_1}^k \cdot k \left( \sum_{r,l} |1 - P_{r,l}(G)| \right). \tag{32}$$

Because the two basis sets index $\{|\phi_r\rangle\}$ and $\{|\phi_{r+1}\rangle\}$ are independent, we change the indices to $r$ and $l$. We note that all non-zero elements in the $\tilde{G}$ matrix have phase factor 1. Thus, we change the element-wise subtraction in Eq. (32) to matrix-wise subtraction in the form of the phase matrix $P$ and obtain

$$\begin{aligned} &= \beta \|G\|_{L_1} \|P(\tilde{G}) - P(G)\|_{L_1} \sum_{k=0}^{\infty} \frac{(\beta)^k}{k!} \|G\|_{L_1}^k \\ &= \beta \|G\|_{L_1} \|P(\tilde{G}) - P(G)\|_{L_1} \exp(\beta \|G\|_{L_1}). \end{aligned} \tag{33}$$

It is worth noting that Eq. (33) states that the sign problem relates to the norm of the Hamiltonian and also its phases of off-diagonal terms. As all phase factors of elements in $G$ approximate 1, the sign problem alleviates. Equivalently, the more the Hamiltonian $H$ is closer to its stoquastic counterpart $\tilde{H}$, the more the sign problem is relieved.

### D. Initial-State-Related Non-Stoquasticity Indicator

For our setting, it is particularly intriguing to investigate to what extent the quality of the initial state influences the sign problem. By quality, we mean the fidelity of the initial state compared to the ground state. For projector QMC algorithms, as we sum up the walkers at the final step, the overall stochastic implementation of the ITE process is equivalent to the following formula,

$$\frac{e^{-\beta H} |\zeta_0\rangle}{\langle \zeta_0 | e^{-\beta H} |\zeta_0\rangle}. \tag{34}$$

Here, $|\zeta_i\rangle$ is the $i$-th Fock state, and the $|\zeta_0\rangle$ is the Hartree Fock state, which is usually chosen to be the initial state for various kinds of projector QMC algorithms [16, 25].

Evidently, the partition function in Eq. (34) is related to the initial state and the basis we choose. Following the above initial-state-agnostic analysis for the partition function, here we discuss bounds for NSI equation related to the initial state. For our cases, we consider the setting that the initial state to be the VQE-optimized state $|\psi_0\rangle = U(\vec{\theta}) |\zeta_0\rangle$. That is we are interested in the phenomenon that when the VQE-prepared state approaches the ground state, how does the NSI behave? The main difference of the partition function from the above analysis is that the begins and ends of path integrals are restricted to the initial state $|\psi_0\rangle$. When the expression of Hamiltonian $H$ is real in the basis $\{\phi_i\}$, the deduction of the NSI is similar to Eq. (20). We have

$$\begin{aligned} S(H, \psi_0)/e^{-\beta\alpha} &= \frac{\langle \psi_0 | e^{-\beta \tilde{H}} |\psi_0\rangle - \langle \psi_0 | e^{-\beta H} |\psi_0\rangle}{e^{-\beta\alpha}} \\ = \sum_{k=0}^{\infty} \frac{\beta^k}{k!} \sum_{\{|\phi\rangle\}} &\left( |\langle \psi_0 | G |\phi_1\rangle \langle \phi_1 | G |\phi_2\rangle \cdots \langle \phi_{k-1} | G |\psi_0\rangle| - \langle \psi_0 | G |\phi_1\rangle \langle \phi_1 | G |\phi_2\rangle \cdots \langle \phi_{k-1} | G |\psi_0\rangle \right) \end{aligned} \tag{35}$$

It is not too hard to conceive that as the fidelity of the VQE-prepared state goes up, the off-diagonal terms in the 0th row of matrix $G$, $G(0)$, will be suppressed, and become zeros when it reaches the ground state. To study the relationship between the fidelity $|\psi_0\rangle$ and bound for the NSI, we consider the case that the second line in Eq. (35) contains $(l-1)$-degree of $G(0,0)$ from the beginning, and $(r-1)$-degree of $G(0,0)$ from the end of the path integral. Here, we denote $G(0,0)$ as the 0-th row and column term in $G$. We sum over all possible $l$ and $r$. There are two cases where negative paths involve, which are: i) a negative intermediate path both ends either both positive or negative;

ii) a positive intermediate path with one end positive and the other negative. This gives us

$$\begin{aligned}\leq & 2\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\Bigg(\sum_{0\leq l<r\leq k-1}G(0,0)^{l-1+k-r}\Bigg(\sum_{w,x,y,z=1}^{k-1}(G(0,w)_+\cdot G(0,x)_+ + G(0,y)_-\cdot G(0,z)_-)\\ & \cdot\sum_{\{\eta_1,\eta_2\}}\left|G_-^{(1)}(\eta_1|\eta_2)G_+^{(r-l-2)}(\eta_2|\eta_1)+G_-^{(3)}(\eta_1|\eta_2)G_+^{(r-l-4)}(\eta_2|\eta_1)+\cdots\right|\\ & +\binom{2}{1}\sum_{x,y=1}^{n-1}G(0,x)_+\cdot|G(0,y)_-|\cdot\sum_{\{\eta_1,\eta_2\}}(G_+^{(r-l-1)}(\eta_1|\eta_1)+G_+^{(r-l-3)}(\eta_1|\eta_2)G_-^{(2)}(\eta_2|\eta_1)+\cdots)\Bigg)+G(0,0)^{k-1}\sum_{x=1}^{n-1}|G(0,x)_-|\Bigg),\end{aligned} \tag{36}$$

where we denote the $i$-th row, $j$-th column element of $G$ as $G(i,j)$. With the same claim as Eq. (21), we have

$$\begin{aligned}\leq & \sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\Bigg(\sum_{0\leq l<r\leq k-1}G(0,0)^{l-1+k-r}\Big((\|G(0,\setminus 0)_+\|_{L_1}\cdot\|G(0,0)_+\|_{L_1}+\|G(0,\setminus 0)_-\|_{L_1}\cdot\|G(0,\setminus 0)_-\|_{L_1})\\ & \cdot\big((\|G_+\|_{L_1}+\|G_-\|_{L_1})^{r-l-1}-(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{r-l-1}\big)\\ & +2\|G(0,\setminus 0)_+\|_{L_1}\cdot\|G(0,\setminus 0)_-\|_{L_1}\cdot\big((\|G_+\|_{L_1}+\|G_-\|_{L_1})^{r-l-1}+(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{r-l-1}\big)\Big)\\ & +2\binom{k-1}{1}G(0,0)^{k-1}\|G(0,\setminus 0)_-\|_{L_1}\Bigg).\end{aligned} \tag{37}$$

Here, $G(i,\setminus j)$ denotes the set of elements of $G$ in row $i$ and columns other than $j$. To proceed, we rearrange the terms in Eq. (37), and complete two squares

$$\begin{aligned}= & \sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\Bigg(\sum_{0\leq l<r\leq k-1}G(0,0)^{l-1+k-r}\Big((\|G(0,\setminus 0)_+\|_{L_1}+\|G(0,\setminus 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}+\|G_-\|_{L_1})^{r-l-1}\\ & -(\|G(0,\setminus 0)_+\|_{L_1}-\|G(0,\setminus 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{r-l-1}\Big)+2(k-1)G(0,0)^{k-1}\|G(0,\setminus 0)_-\|_{L_1}\Bigg).\end{aligned} \tag{38}$$

Next, we substitute $r-l$ to $\Delta$, and Eq. (38) becomes

$$\begin{aligned}= & \sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\Bigg(\sum_{1\leq\Delta\leq k-1}G(0,0)^{k-1-\Delta}\Bigg(\binom{k-\Delta}{1}\cdot(\|G(0,\setminus 0)_+\|_{L_1}+\|G(0,\setminus 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}+\|G_-\|_{L_1})^{\Delta-1}\\ & -\binom{k-\Delta}{1}\cdot(\|G(0,\setminus 0)_+\|_{L_1}-\|G(0,\setminus 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{\Delta-1}\Bigg)+2(k-1)G(0,0)^{k-1}\|G(0,\setminus 0)_-\|_{L_1}\Bigg).\end{aligned} \tag{39}$$

We use the gradient trick to the $G(0,0)$ terms in Eq. (39) to get rid of the binomial coefficient $\binom{k-\Delta}{1}$, and this gives us

$$\begin{aligned}= & \sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\Bigg(\sum_{1\leq\Delta\leq k-1}\frac{\mathrm{d}G(0,0)^{k-\Delta}}{\mathrm{d}G(0,0)}\Big((\|G(0,\setminus 0)_+\|_{L_1}+\|G(0,\setminus 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}+\|G_-\|_{L_1})^{\Delta-1}\\ & -(\|G(0,\setminus 0)_+\|_{L_1}-\|G(0,\setminus 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{\Delta-1}\Big)\Bigg)\\ & +2\beta\|G(0,\setminus 0)_-\|_{L_1}\sum_{k=1}^{\infty}\frac{(\beta)^{k-1}}{(k-1)!}G(0,0)^{k-1}-\frac{2\|G(0,\setminus 0)_-\|_{L_1}}{G(0,0)}\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}G(0,0)^k.\end{aligned} \tag{40}$$

We first compute the last term in Eq. (40)

$$=\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\left(\frac{\mathrm{d}}{\mathrm{d}G(0,0)}\sum_{1\leq\Delta\leq k-1}G(0,0)^{k-\Delta}\Big((\|G(0,\backslash 0)_+\|_{L_1}+\|G(0,\backslash 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}+\|G_-\|_{L_1})^{\Delta-1}\right.$$
$$\left.-(\|G(0,\backslash 0)_+\|_{L_1}-\|G(0,\backslash 0)_-\|_{L_1})^2\cdot(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{\Delta-1}\Big)\right)+2\|G(0,\backslash 0)_-\|_{L_1}\cdot\frac{2\beta G(0,0)-1}{G(0,0)}\exp(\beta G(0,0)). \tag{41}$$

Note that the summation for $\Delta$ in Eq. (41) is nothing but two geometric progressions. Applying the summation formula of geometric progression, we acquire

$$\begin{aligned}=&\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\frac{\mathrm{d}}{\mathrm{d}G(0,0)}\left((\|G(0,\backslash 0)_+\|_{L_1}+\|G(0,\backslash 0)_-\|_{L_1})^2\cdot\frac{G(0,0)^k-G(0,0)(\|G_+\|_{L_1}+\|G_-\|_{L_1})^{k-1}}{G(0,0)-(\|G_+\|_{L_1}+\|G_-\|_{L_1})}\right.\\&\left.+(\|G(0,\backslash 0)_+\|_{L_1}-\|G(0,\backslash 0)_-\|_{L_1})^2\cdot\frac{G(0,0)^k-G(0,0)(\|G_+\|_{L_1}-\|G_-\|_{L_1})^{k-1}}{G(0,0)-(\|G_+\|_{L_1}-\|G_-\|_{L_1})}\right)\\&+2\|G(0,\backslash 0)_-\|_{L_1}\cdot\frac{2\beta G(0,0)-1}{G(0,0)}\exp(\beta G(0,0))\end{aligned} \tag{42}$$

The next step is to compute the gradient, and we have

$$\begin{aligned}=&\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\frac{(\|G(0,\backslash 0)_+\|_{L_1}+\|G(0,\backslash 0)_-\|_{L_1})^2}{G(0,0)-(\|G_+\|_{L_1}+\|G_-\|_{L_1})}\cdot\left((\|G_+\|_{L_1}+\|G_-\|_{L_1})^k+\left(k-1-\frac{k(\|G_+\|_{L_1}+\|G_-\|_{L_1})}{G(0,0)}\right)G(0,0)^k\right)\\&-\frac{(\|G(0,\backslash 0)_+\|_{L_1}-\|G(0,\backslash 0)_-\|_{L_1})^2}{G(0,0)-(\|G_+\|_{L_1}-\|G_-\|_{L_1})}\cdot\left((\|G_+\|_{L_1}-\|G_-\|_{L_1})^k+\left(k-1-\frac{k(\|G_+\|_{L_1}-\|G_-\|_{L_1})}{G(0,0)}\right)G(0,0)^k\right)\\&+2\|G(0,/0)_-\|_{L_1}\cdot\frac{2\beta G(0,0)-1}{G(0,0)}\exp(\beta G(0,0)).\end{aligned} \tag{43}$$

By definition of the series expansion, we finally get

$$\begin{aligned}=&\frac{(\|G(0,\backslash 0)_+\|_{L_1}+\|G(0,\backslash 0)_-\|_{L_1})^2}{(G(0,0)-(\|G_+\|_{L_1}+\|G_-\|_{L_1}))^2}\exp(\beta(\|G_+\|_{L_1}+\|G_-\|_{L_1}))\\&-\frac{(\|G(0,\backslash 0)_+\|_{L_1}-\|G(0,\backslash 0)_-\|_{L_1})^2}{(G(0,0)-(\|G_+\|_{L_1}-\|G_-\|_{L_1}))^2}\exp(\beta(\|G_+\|_{L_1}-\|G_-\|_{L_1}))\\&+\frac{(\|G(0,\backslash 0)_+\|_{L_1}+\|G(0,\backslash 0)_-\|_{L_1})^2(\beta(G(0,0)+\|G_+\|_{L_1}+\|G_-\|_{L_1})-2)}{(G(0,0)-(\|G_+\|_{L_1}+\|G_-\|_{L_1}))^2}\exp(\beta G(0,0))\\&-\frac{(\|G(0,\backslash 0)_+\|_{L_1}+\|G(0,\backslash 0)_-\|_{L_1})^2(\beta(G(0,0)+\|G_+\|_{L_1}-\|G_-\|_{L_1})-2)}{(G(0,0)-(\|G_+\|_{L_1}-\|G_-\|_{L_1}))^2}\exp(\beta G(0,0))\\&+2\|G(0,\backslash 0)_-\|_{L_1}\cdot\frac{2\beta G(0,0)-1}{G(0,0)}\exp(\beta G(0,0)).\end{aligned} \tag{44}$$

Now, it is evident from Eq. (44) that every term in the NSI is suppressed at least linearly by the off-diagonal terms in $G(0)$. The implication is that even if we only optimize the energy as for the VQE algorithms, we nonetheless mitigate the sign problem as off-diagonal terms vanish more and more in $G(0)$ as the prepared state gets closer to the ground state.

Whereas the Hamiltonian is complex, we follow a similar definition of NSI as Eq. (45) for the initial state-related case

$$S(H,\psi_0)/e^{-\beta\alpha}=\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\left\langle\psi_0\right|e^{-\beta\tilde{G}}\left|\psi_0\right\rangle\ -\sum_{k=0}^{\infty}\frac{(\beta)^k}{k!}\left|\left\langle\psi_0\right|e^{-\beta\tilde{H}}\left|\psi_0\right\rangle\right|. \tag{45}$$

Compared to cases where the Hamiltonians are real, the formula is different from Eq. (36) in that the paths other than $G(0,0)$ consist of complex elements. Thus, we can simply follow the deduction in the initial-state-agnostic situation. Following the same idea as in the real case, it is natural to expect that all terms in the bound contain terms quadratic or linear to $\|G(0,\backslash 0)\|_{L_1}$.

To summarize, as we discussed in subsection II A, during alternating the walker basis, we effectively rotate the Hamiltonian with $U(\vec{\theta})$. This has two-fold implications for mitigating the sign problem. First, as one suppresses the off-diagonal terms in the effective Hamiltonian $H_T(\vec{\theta})$ the sign problem is exponentially mitigated. Second, from the initial state-related analysis, we know that the sign problem is alleviated at least linearly to $\|\Pi_\perp H |\psi_0\rangle\|$ with $\Pi_\perp = I - |\psi_0\rangle\langle\psi_0|$. This instructs us that rather than approximately diagonalizing the Hamiltonian, we can seek an easier way that we only optimize our ansatz state $|\psi_0\rangle$ towards the ground state. Compared to the diagonalization of the Hamiltonian, the latter is more near-term amenable.

## III. QUANTUM CIRCUIT-BASED QMC

### A. Motivation

Although VQE algorithms have become standard prototypes for solving ground state properties regarding quantum many-body and molecular systems in the NISQ era. Several drawbacks hinder their performance in both practical and theoretical ways:

- Due to the presence of all sorts of noises, the VQE algorithms are limited to relatively shallow quantum circuits for successful implementation even aided with quantum error mitigation approaches. Thus, the expressiveness of VQE is restricted, so the performance could be compromised especially when dealing with strong-correlated systems.
- For their variational nature, the VQE approaches are not guaranteed to converge to the optimal solution because they can get trapped in a local minimum easily as [26] proved that the loss landscape is packed out with local minimums.
- Intrinsically, the gradients of the VQE algorithms vanish exponentially as one extends the circuit depth. These types of phenomena known as "barren plateau" [27] are found ubiquitous for any highly expressive circuits that reach 2-design over the Haar measure.

From the discussion above, we know that insights are required for one to carefully choose a suitable ansatz that is both expressive enough to include the optimal solution and not too strong to bear barren plateaus. Unfortunately, no understanding has been found for systematically building reasonable VQE circuits in generic cases. Besides, for practical concerns, most error mitigation strategies are not scalable and it is hardly possible for implementing VQE experiments to go beyond the reach of classical simulation.

To facilitate research for applying quantum algorithms to study quantum many-body systems in the NISQ era, herein we suggest combining the VQE with ITE algorithms, which are well-known approaches that guarantee to produce the ground state as we go to the long-time limit. Although the operations in the ITE approaches are not unitary, classical projector QMC methods have provided the solution to implement each evolution step in a statistical way. Therefore, each step in the whole process can be made unitary. Besides, the repeated-state-preparation feature of QMC allows shallow quantum circuits to be used in the experiment. Hence, the combination of the two cutting-edge methods enables NISQ-device-friendly implementations of our Quantum Circuit-based (QC)-QMC schemes. From another perspective, we prove and numerically demonstrate in the following session that the VQE methods are found to be helpful for easing sign problem for the QC-QMC algorithms. Thus, we expect our approaches can push the limit for studying quantum many-body and chemical problems.

### B. Algorithm

In this section, we describe our hybrid algorithm. Following our previous discussion on sign problem, choosing an appropriate basis helps to reduce the sign problem and single Slater determinants are rather restricted in this sense. Though being an NP-hard problem, the sign problem may be alleviated using more complicated basis sets provided by NISQ devices. In the following, we will introduce our algorithm using shallow VQE circuits to transform computational basis states (single determinant state after Jordan-Wigner encoding) into entangled walkers that are closer to the diagonal basis of the Hamiltonian, which will be adopted into the FCIQMC framework to reduce the sign problem.

Now our wavefunction is expanded in a new orthonormal basis $\{|\phi_i\rangle\}_i$:

$$|\psi(\tau)\rangle = \sum_i \widetilde{c}_i(\tau)\,|\phi_i\rangle \tag{46}$$

Substituting this into Eq. (8) again and taking the first-order approximation, one obtains a set of coupled differential equations for the evolution of coefficients $\widetilde{c_i}$s:

$$-\frac{\mathrm{d}\widetilde{c_i}(\tau)}{\mathrm{d}\tau} = \sum_j (\widetilde{H_{ij}} - S\delta_{ij})\widetilde{c_j}(\tau) \tag{47}$$

where $\widetilde{H_{ij}}$s are the matrix elements of $H$ under the new basis. Here we can start the energy shift $S$ with the VQE energy and adjusts the value in a similar way as in the classical FCIQMC setting. We generate the new basis by bringing the ground state of the mean-field Hamiltonian, i.e., the Hartree-Fock state, closer to the true ground state of the Hamiltonian. We achieve this by running a variational algorithm on a quantum device starting with the Hartree-Fock state and then obtaining an optimized circuit denoted as $U(\vec{\theta})$. With this circuit, we now act it on the Fock basis to form a set of new basis, $\{|\phi_i\rangle := U(\vec{\theta})\,|i\rangle\}$, whose orthogonality can be easily checked as follows:

$$\langle j|\, U(\vec{\theta})^\dagger U(\vec{\theta})\,|i\rangle = \langle j|i\rangle = \delta_{ij}, \tag{48}$$

where $\delta$ is the Dirac delta function. These new basis states are multi-configurational, and the optimized ground state $\{U(\vec{\theta})\,|HF\rangle\}$ has a bigger overlap with the exact ground state than the initial Hartree-Fock state. Given Pauli decomposition of the Hamiltonian $H = \sum_k h_k P_k$, the matrix elements of the Hamiltonian $\widetilde{H_{ij}}$ under basis $\{|\phi_i\rangle\}$ can be estimated on a quantum processor with the circuit in Fig. 1 (a) for each $P_k$. With the VQE energy and the new $\widetilde{H_{ij}}$s under the new basis, we reformulate the FCIQMC procedure as follows for the sake of completeness and we summarize the whole workflow in Algorithm 1.

- Spawning: For each walker living in $|\phi_i\rangle$, spawn a child walker $|\phi_j\rangle$ ($j \neq i$) with probability $\left|\widetilde{H_{ij}}\right|\Delta\tau$ up to a normalization constant. Moreover, if this quantity $\widetilde{H_{ij}}$ is positive, then the child has the same sign as the parent, and the opposite sign otherwise.
- Death/cloning: If the quantity $\widetilde{H_{ii}} - S > 0$, then walker $|\phi_i\rangle$ dies with probability $(\widetilde{H_{ii}} - S)\Delta\tau$; and if $\widetilde{H_{ii}} - S < 0$, then the walker clones itself with probability $\left|(\widetilde{H_{ii}} - S)\right|\Delta\tau$.
- Annihilation: At the end of each time step, we take the signed sum of walkers at each basis vector and annihilate every pair of walkers with opposite signs.

**Algorithm 1** QC-FCIQMC

**Input:** Hamiltonian $H$, total evolution time $T$, time step $\Delta\tau$.
**Output:** Ground state energy estimation.

```
Run VQE to determine U and hence {|φ_i⟩ = U|i⟩}.
Generate N_0 walkers |φ_0⟩ and let walker set D = {0}.
for n in range(T/Δτ) do
    for i in D do
        Estimate |H̃_ij| using the circuits in Fig. 1(b).
        for j with nonzero |H̃_ij| do                          ▷ Spawning step
            For each walker |φ_i⟩, spawn a new walker |φ_j⟩ with probability Δτ|H̃_ij|.
            if New walker |φ_j⟩ spawned then
                Estimate H̃_ij/|H̃_ij| using the circuit in Fig. 1(a)
                Label the new walker |φ_j⟩ with the sign of |φ_i⟩ multiplied by −H̃_ij/|H̃_ij|.
                Add j to D.
        Estimate p_i = Δτ(H̃_ii − S)
        if p_i < 0 then                                       ▷ Death/cloning step
            Clone each walker |φ_i⟩ with probability |p_i|.
        else
            Kill each walker |φ_i⟩ with probability p_i.
        for i in D do                                         ▷ Annihilation step
            Annihilate the walkers |φ_i⟩ with opposite signs.
Output the mixed energy.
```

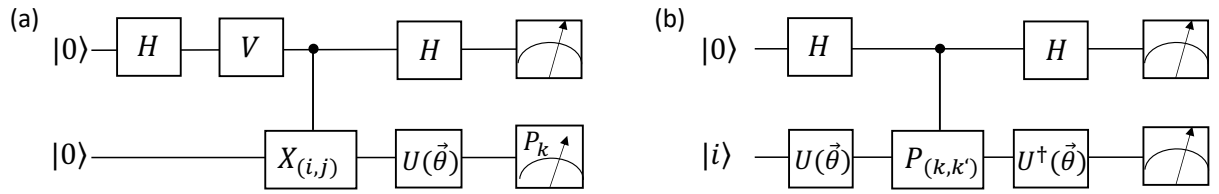


Figure 1. (a) Circuit for estimating amplitudes $\mathrm{Re}(\langle\phi_j|\,H\,|\phi_i\rangle)$. One let $V = I$ when estimating the real parts and $V = S^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$ for the imaginary parts. The gate $X_{i,j}$ implements $X_i$ if the control register is $|0\rangle$ and $X_j$ if the control register is $|1\rangle$. $X_i$ implements $X$ gates on those qubits where the corresponding digits in $i$ are 1. (b) Circuit for sampling the significant amplitudes in the distribution $|\widetilde{H_{ji}}|^2$ for a fixed $i$. The gate $P_{k,k'}$ implements $P_k$ if the control register is $|0\rangle$ and $P_{k'}$ if the control register is $|1\rangle$.

### C. Complexity Analysis

In this section, we estimate the time complexity of our algorithm. This can be divided into two parts, which are the complexity using Fig. 1 (a) to sample the entries in the Hamiltonian matrix for estimation; and the estimation of these entries using Fig. 1 (b). The rest is classical post-processing just like classical FCIQMC.

Assuming for a given walker $i$, we have garnered enough sample, and the walker set is denoted as $\mathcal{D}^i = \{|\phi_j\rangle\}$ using the sampling method specified in the last section. Now, let us discuss the complexity for estimation of each $\widetilde{H_{ji}} := \langle\phi_j|H|\phi_i\rangle$ within given accuracy $\epsilon$ for all $|\phi_j\rangle$ in the walker set $\mathcal{D}^i$ using quantum circuit depicted in Figure 1 (a). The sample complexity for achieving this goal is analyzed as follows.

**Lemma 1** (Sample complexity for estimating $\widetilde{H_{ji}}$)**.** *Given Hamiltonian $H = \sum_{k=1}^K h_k P_k$ and target $i, j$ for $\widetilde{H_{ij}}$, let $\{\boldsymbol{A}_{ijk}^{(l)}, \boldsymbol{B}_{ijk}^{(l)}\}_{k,l=1}^{K,L_k}$ be $M = \sum_{k=1}^K L_k$ samples that are measurement outcomes from quantum circuits given in Figure 1 (a) with $V = I$ and $S^\dagger$, respectively, and $i, j, k$ are set to given indices. Define $\bar{\boldsymbol{C}}_{ijk}$ as*

$$\bar{\boldsymbol{C}}_{ijk} := \frac{1}{L_k}\sum_{l=1}^{L_k} h_k(\boldsymbol{A}_{ijk}^{(l)} + i\boldsymbol{B}_{ijk}^{(l)}). \tag{49}$$

*Then, the estimator for $\widetilde{H_{ji}}$ defines as*

$$\bar{\boldsymbol{C}}_{ij} := \sum_{k=1}^K \bar{\boldsymbol{C}}_{ijk}. \tag{50}$$

*Also, define $h_s := \sum_{k=1}^K |h_k|$. Then for any $\epsilon > 0$ and $\mu \in (0, 1)$, let each $L_k$ takes the following value*

$$L_k := \left\lceil \frac{|h_k|^2 \ln(4K/\mu)}{\epsilon_k^2} \right\rceil, \tag{51}$$

*where $\epsilon_k$ is given by*

$$\epsilon_k^2 := \frac{|h_k|}{h_s}\epsilon^2. \tag{52}$$

*Then, when the total number of samples $M$ is given by*

$$M := \left\lceil \frac{h_s^2 \ln(4K/\mu)}{\epsilon^2} \right\rceil, \tag{53}$$

*we have*

$$\mathbb{P}\left[|\bar{\boldsymbol{C}}_{ij} - \widetilde{H_{ji}}| < \epsilon\right] \geq 1 - \mu. \tag{54}$$

*Proof.* Note that Eq. (50) serves as an unbiased estimator for evaluating the $h_k\langle\phi_j|P_k|\phi_i\rangle$, such that $\mathbb{E}[\bar{\mathbf{C}}_{ijk}] \equiv h_k\langle\phi_j|P_k|\phi_i\rangle$. Therefore, Eq. (50) serves as an unbiased estimator for estimating the $\widetilde{H_{ji}}$, and the empirical mean value approaches the expectation value when more samples are collected. For each $k \in [1, K]$, the value of $\mathrm{Re}(\bar{\mathbf{C}}_{ijk})$ and $\mathrm{Im}(\bar{\mathbf{C}}_{ijk})$ are in $[-|h_k|, |h_k|]$ as $\mathbf{A}^{(l)}_{ijk}$ and $\mathbf{B}^{(l)}_{ijk}$ take value in $[-1, 1]$. When taking the number of samples as Eq. (51) for all $k \in [1, K]$, the following formulas hold by applying Hoeffding's inequality [28],

$$\begin{cases} \mathbb{P}\left[\left|\mathrm{Re}\left(\bar{\mathbf{C}}_{ijk}\right) - \mathrm{Re}\left(h_k\langle\phi_j|P_k|\phi_i\rangle\right)\right| > \dfrac{\epsilon_k}{\sqrt{2}}\right] < \dfrac{\mu}{2K}, \\ \mathbb{P}\left[\left|\mathrm{Im}\left(\bar{\mathbf{C}}_{ijk}\right) - \mathrm{Im}\left(h_k\langle\phi_j|P_k|\phi_i\rangle\right)\right| > \dfrac{\epsilon_k}{\sqrt{2}}\right] < \dfrac{\mu}{2K}. \end{cases} \tag{55}$$

By applying the triangle inequality and the union bound of the above two formulas, we have

$$\mathbb{P}\left[\left|\bar{\mathbf{C}}_{ijk} - h_k\langle\phi_j|P_k|\phi_i\rangle\right| > \epsilon_k\right] < \frac{\mu}{K}. \tag{56}$$

Finally, when choosing each $\epsilon_k$ as Eq. (52), the overall $\epsilon$ accuracy is achieved for evaluating $\widetilde{H_{ji}}$ by applying the union bound to all $k \in [1, K]$:

$$\mathbb{P}\left[|\bar{\mathbf{C}}_{ij} - \widetilde{H_{ji}}| > \epsilon\right] < \mu. \tag{57}$$

And the total number of samples $M$ is obtained by substituting Eq. (52) into Eq. (51)

$$\begin{aligned} M &= \sum_{k=1}^{K} L_k \\ &= \sum_{k=1}^{K} \left\lceil \frac{h_s|h_k|\ln(4K/\mu)}{\epsilon^2} \right\rceil \\ &= \left\lceil \frac{h_s^2 \ln(4K/\mu)}{\epsilon^2} \right\rceil. \end{aligned} \tag{58}$$

Note that the sample complexity is similar to that of the energy measurement process for near-term quantum devices using the frequentist approach [10, 13] as the quantity that we estimate here has a similar form to the energy.

### D. Comparison with Other Methods

In this section, we compare our method with recent states-of-the-art ground state energy estimation quantum algorithms [29–33] in a computational complexity and resource-consumption sense in the following.

Recently, quantum algorithms [29–33] are designed regarding the ground state problems that allow a shorter circuit-depth comparing to quantum phase estimation methods and also achieve (near)-optimal scaling [29] in query complexity at the same time. Those quantum algorithms are designed for early-fault-tolerant quantum devices, where the decoherence time is supposed to be longer than near-term devices but is still limited. Besides, the number of logical qubits is limited; as such, it is favorable to employ as less as possible ancillary qubits. We provide the query complexity and maximal circuit-depth comparison between our method and state-of-the-art quantum algorithms. Particularly, we focus on three specific properties of an algorithm: i) Maximal circuit depth that needs to be implemented in one round; ii) Number of ancillary qubits needed; iii) The main subroutine that determines whether or not the method needs multi-qubit-control operations. These three properties determine whether we should implement the quantum algorithm on near-term, early-fault-tolerant, or full-fledged quantum computers. In addition to the advanced algorithms, we also consider two types of textbook algorithms, i.e quantum phase estimation (QPE) method [34, 35] and iterative quantum phase estimation (IQPE) [36, 37]. We denote the accuracy of estimating the ground state energy as $\varepsilon$, and for quantum molecular systems, a standard choice is the chemical accuracy, which is $\varepsilon_{\text{chemical}} = 1.6 * 10^{-3}$ Hartree. Also, it is typically assumed that a lower-bounded estimation $\Delta$ to the energy gap is given such that $E_1 - E_0 \geq \Delta$, where $E_0$ and $E_1$ is the ground- and first-excited state energy.

Two assumptions are made for discussion of the query complexity: i) an initial state $\phi_0$ is known and its low bound $\gamma$ of overlap between the initial state and the ground state is provided, such that $|\langle\phi_0|\Psi_0\rangle| \geq \gamma$, where $|\Psi_0\rangle$ is the ground state of $H$; ii) a lower bound $\Delta$ of the energy gap between the ground and first-excited state of the Hamiltonian

is given, such that $E_1 - E_0 \geq \Delta$, where $E_0$ and $E_1$ are ground and first-excited state energy of $H$, respectively. We summarize the main properties of each quantum algorithm in the following table. We summarize the main properties of each quantum algorithm in the following table. compared to quantum phase estimation [34, 35, 38] and other recently proposed ground state projection/filtering methods [29–33]. Briefly speaking, we would like to point out that our method requires much shallower quantum circuits and hence is much more resource-friendly than existing methods [29–35, 38]. Our method is particularly tailored for NISQ hardware, whereas these existing methods [29–35, 38] generally require deep quantum circuits, which generally require fault-tolerant quantum computing. Here, we give a detailed comparison from the perspectives of theoretical complexity and resource analysis.

1. *Complexity analysis.*— Here, we compare the complexity (query complexity/circuit depth) of the methods.

   - We first analyze the circuit depth complexity of our QC-FCIQMC algorithm, which effectively realizes the imaginary time evolution (ITE) process to the initial walker. Since our sampling circuit (Fig. (f,g) in the main text) only requires one ancillary qubit and a circuit depth of $2D(U) + \mathcal{O}(n)$, where $D(U)$ represents the circuit depth of $U$ and $\mathcal{O}(n)$ corresponds to the controlled Pauli operation. Since $U$ is obtained from variational quantum eigensolver, we generally have $D(U) \sim \mathcal{O}(n)$. Therefore, the circuit depth complexity of our method scales as $\mathcal{O}(n)$. We note that the circuit depth does not explicitly depend on the accuracy $\varepsilon$ or the energy gap $\Delta$, because the quantum circuit is only used as the walker basis and the sampling of the walkers. Instead, the imaginary evolution time might depend on $\varepsilon$ and $\Delta$, which might thus introduce more repetitions of the quantum circuits (could be sped up with multiple quantum hardware). The feature of our method (shallower circuit depth at a cost of more circuit repetitions) exactly reflects the power of NISQ hardware, where the circuit depth is generally limited owing to non-negligible gate errors yet we could repeatedly run the circuits. In particular, we note that our circuit-depth requirement is suitable for near-term quantum devices according to a recent experimental demonstration of VQE methods [39–42].

Table I. Quantum algorithms performance for the ground state energy estimation problem

| | maximal query/circuit depth | number of ancillary qubits | main subroutine |
|---|---|---|---|
| QPE [34, 35] | $\widetilde{\mathcal{O}}(\varepsilon^{-1})$ | $\mathcal{O}(\text{polylog}(\gamma^{-1}\varepsilon^{-1}))$ | controlled Hamiltonian evolution/qubitization + Quantum Fourier transform |
| Iterative QPE [36, 37] | $\widetilde{\mathcal{O}}(\varepsilon^{-1}\gamma^{-2})$ | $\mathcal{O}(1)$ | controlled Hamiltonian evolution + Hadamard test |
| LT [30] | $\widetilde{\mathcal{O}}(\varepsilon^{-1})$ | $\mathcal{O}(1)$ | controlled Hamiltonian evolution + Hadamard test |
| Wan *et al.* [31] | $\widetilde{\mathcal{O}}(\varepsilon^{-2}\lambda^{2})$ | $\mathcal{O}(1)$ | controlled Pauli rotation + Hadamard test |
| Wang *et al.* [32] | $\widetilde{\mathcal{O}}(\Delta^{-1})$ | $\mathcal{O}(1)$ | controlled Hamiltonian evolution + Hadamard test |
| DLT [33] | $\widetilde{\mathcal{O}}(\varepsilon^{-1}\gamma^{-1})$ | $\mathcal{O}(1)$ | controlled Hamiltonian evolution + quantum signal processing |
| This work | $\mathcal{O}(n)$ | $\mathcal{O}(1)$ | controlled Pauli rotation + Hadamard test |

   - As for the other existing methods [30–32, 37], they generally need to implement the Hadamard-test quantum circuit with the controlled unitary being the Hamiltonian evolution operator with evolution time $t$. Further operations such as quantum Fourier transform might be applied, which would cause more gate depth. Here, we first focus on the maximal evolution time of these methods regarding $\varepsilon$ and $\Delta$. The complexity of the circuit depth also needs to take into account the circuits for realizing the time evolution. We summarize the maximal query of the evolution time $t$ in Table I. For example, the work of Lin and Tong [30] applies the maximum evolution time for the Hamiltonian evolution at a single iteration is $t = \widetilde{\mathcal{O}}(\varepsilon^{-1})$. Ref. [32] managed to decrease the maximal evolution time to $\widetilde{\mathcal{O}}(\Delta^{-1})$. Since $\varepsilon^{-1}$ or $\Delta^{-1}$ are also polynomial functions of $n$, the complexity of the evolution time is already generally larger than $\Omega(n)$. Furthermore, there also exists a cost for implementing $e^{-iHt}$. For example, even just considering a simple 1D Ising Hamiltonian with nearest-neighbor interaction, the gate complexity for implementing $e^{-iHt}$ is $\mathcal{O}(n^3)$ even for the best existing Hamiltonian method [43]. The cost would even be larger for more complex chemistry Hamiltonians. Therefore, the overall circuit complexity would be much larger than linear (our method). This explains why these existing methods [30–32, 34–37] are generally considered (early) fault-tolerant algorithms [44], which are beyond the capability of near-term quantum computing.

| Gate | Gate Count | Depth |
|---|---|---|
| Single-qubit gates | $43m - 40\sqrt{m}$ | 16 |
| CNOT | $2(15m - 20\sqrt{m} + 6)$ | $6(2\sqrt{m} + 1)$ |

Table II. Adapted from Ref. [47]. Total resources (single-qubit gates including Clifford, non-Clifford $R_z$) for implementing single time step sweep for a-$m$ modes (sites) Hubbard model.

2. *Resource analysis.*— Here, we also present the resource (circuit depth/gate complexity) analyses of our and other methods [30–32, 37] for some example problems to explicitly show the advantage of our method. This comparison will not include the quantum circuits for the initial state preparation for the early fault-tolerant algorithms [30–32, 37] and the VQE-prepared state in our method.

From the above discussion, we know that for early fault-tolerant algorithms [30–32, 37], the most circuit-depth consuming part is the (controlled) Hamiltonian simulation, which is to approximate $e^{-iHt_{\max}}$ with longest evolution time $t_{\max} = \frac{1}{\varepsilon}$. Here, $\varepsilon$ is the accuracy in estimating the ground state energy. We will first provide analysis results for the Hamiltonian simulation and then lift the results to a controlled version. As such, we will compare the resources of our methods with Hamiltonian simulation methods that are suitable for early fault-tolerant devices. Note that the following resource estimation results do not include the overhead for fault tolerance (which generally requires thousands of physical qubits to encode a single logical qubit and additional operations on magic state distillation.). As suggested in the early fault-tolerant algorithms [30, 33], the Trotter formulae are favored for the realization of the Hamiltonian simulation. This is because the Trotter formulae may outperform other advanced methods in gate complexity [30, 43], despite other methods that can be more favorable for query complexity. Moreover, it is easy for the Trotter formulae to be tailored for specific quantum models, for which the 2D Fermionic Hubbard model is extensively studied [45–47], to further reduce the gate complexity. Therefore, we will consider the Hamiltonian simulation of 2D fermionic Hubbard models of $m$ fermionic modes (sites) with the Trotter formulae. The Hamiltonian is given by

$$H = -t \sum_{\langle i,j\rangle,\sigma} \left( a_{i\sigma}^\dagger a_{j\sigma} + a_{j\sigma}^\dagger a_{i\sigma} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow}, \tag{59}$$

where $\sigma \in \{\uparrow, \downarrow\}$ and $n$ is the particle number operator. We will introduce the zigzag index strategy as shown in [Ref. [11], FIG. 1]. That is given the 2D grid, we index each site from left to right, and then move to the next row with reverse ordering, and repeat accordingly. By Jordan-Wigner transformation (JWT), the $m$-model Hubbard model transforms to a $n = 2m$-qubit system as the fermionic operators are transformed as

$$\hat{a}_j \mapsto \frac{1}{2}\left(X_j + iY_j\right) \bigotimes_{i=1}^{j-1} Z_i, \;\; \hat{a}_j^\dagger \mapsto \frac{1}{2}\left(X_j - iY_j\right) \bigotimes_{i=1}^{j-1} Z_i \tag{60}$$

with Pauli operators $X$, $Y$, $Z$ acting on the $j$th qubit. Besides, we will encode spin-up and down sectors on the same site to adjacent indices.

Next, we introduce the Hamiltonian simulation results that are specially tailored for the Hubbard model in Ref. [47]. It is shown in Table II (Table IV in Ref. [47]) of the three types of different gate counts for a single step of the Trotter step. Here, we consider the 1st order Trotter formula as an instance. When considering controlled evolution, the single-qubit and CNOT gates will be transformed into CNOT plus single-qubit and Toffoli gates, respectively. Apply the well-known results that a Toffoli gate can be compiled into 6 CNOT and 8 single-qubit (both Clifford and non-Clifford) gates. This gives us the final result that to realize a single controlled Trotter step, the CNOT gate count is $223m - 280\sqrt{m} + 72$. Because the control qubit of these CNOT gates is the ancillary qubit introduced in the Hadamard test, the circuit depth is the same as the gate count. The (worst-case) single-qubit gate count is $326m - 400\sqrt{m} + 96$. Next, we are going to analyze the total Trotter step. As analyzed in Ref. [30, 33], to guarantee the desired accuracy of energy estimation, we need $r$-Trotter step of a 1st-order Trotterization for the maximal evolution time $t_{\max} = \frac{1}{\varepsilon}$ such that

$$r = \widetilde{\mathcal{O}}\left(\max\left(t_{\max} C_{\text{Trotter}}\, \varepsilon^{-1} \gamma^{-1}, t_{\max}\right)\right) \tag{61}$$

where $C_{\text{Trotter}} = \mathcal{O}\left(\left(\sum_i \|H\|_i\right)^2\right)$, and $\gamma$ is the modulus square of the overlap between the initial state and the ground state. For a $4 \times 4$ system (i.e. $m = 16$), the single Trotter step will introduce 2520 CNOT gates and 3712 single-qubit gates. Next, we consider a rather optimistic case where $\gamma = 0.25$ corresponds to an initial state of

50% fidelity, $\varepsilon = 10^{-1}$ and taking $C_{\text{Trotter}} = 1$. This gives us the total Trotter step $r = 2.5 * 10^3$. Finally, the CNOT and single-qubit gate counts for this case are $6.3 * 10^6$ and $9.28 * 10^6$, respectively.

In comparison, because our method only needs to implement controlled Pauli operations, which involve (off-diagonal) terms in the Hamiltonian, we only need to implement Pualis in the single-body terms (i.e., hopping terms) in the Hamiltonian after the JWT. As such, in the worst case corresponding to the interaction between $j$ and $j + 2\sqrt{m}$-th qubits, $2\sqrt{m} = 8$ CNOT gates (correspond to 8 depth) will be introduced. By contrast, only 4 Clifford gates (depth 2) will be introduced and no non-Clifford gates are involved. Therefore, the maximal circuit depth for our method is 10.

From the above discussion, we can see that the high-circuit-depth overhead for advanced methods such as the RFE algorithms [30–32, 37] demands fault tolerance naturally. Yet, our methods only require a quantum circuit of sublinear depth to the system size, which is already within reach for near-term quantum devices [39, 42]. Although here we only focus on the fermionic Hubbard model, our analysis can be readily extended to quantum molecular systems, where the results we believe can still be held. This is because the controlled Hamiltonian simulation for quantum molecular systems can be even more challenging. Yet, the most complicated controlled Pauli terms will be introduced by the two-body interaction terms such that a linear number of CNOT gates will be employed and the number of non-Clifford gates will stay constant.

## IV. NUMERICAL RESULTS

In this section, we present numerical results demonstrating the effectiveness of the QC-FCIQMC algorithm at easing the sign problem and reducing the number of walkers as well as the energy variance compared to classical FCIQMC. We pick two strongly correlated systems to ensure the robustness of our scheme, the $N_2$ molecule at the dissociation limit and the $2 \times 4$ Hubbard model with $U/t = 4$.

We plot the potential energy surface of $N_2$ molecule with our QC-FCIQMC algorithm in Fig. 2 (a). The new walker bases for all bond lengths are prepared by ADAPT-VQE circuits that add 12 fermionic operators. In Fig. 2 (b), we plot the standard deviation from energy profile along the QC-FCIQMC evolution. We freeze the 1s and 2s orbitals and run ADAPT-VQE with 12 qubits. We also plot data from a single determinant FCIQMC for comparison. We set 10000 walkers to be the threshold for both algorithms to start the energy shift. On the one hand, we see that QC-FCIQMC is able to push the results of shallow ADAPT-VQE circuits to the level of chemical accuracy across all bond lengths, therefore easing the burden on the quantum devices. Meanwhile, using the same amount of walkers, single determinant FCIQMC fails to reach chemical accuracy at some of the bond lengths towards the dissociation limit. On the other hand, the energy deviation of QC-FCIQMC is smaller than that of single determinant FCIQMC. Therefore, QC-FCIQMC would require much fewer energy evaluations to obtain a precise energy estimation and maybe smaller total evolution time. All the energy and standard deviation are evaluated by taking energies after reaching a certain total evolution time, and usually, after the total number of walkers stabilizes. We adopt this way of evaluation for all of the following numerical experiments.

For the $N_2$ molecule at bond length $4.0\mathring{A}$, we use quantum circuits from ADAPT-VQE that add different numbers of operators to generate the new basis of QC-FCIQMC. We plot the effect of energy variance reduction of QC-FCIQMC given a fixed number of walkers in Fig. 2 (c) (d). The standard deviation is reduced by close to 100 folds going from a single determinant basis to ADAPT-VQE adding 24 operators. We highlight that at this depth, QC-FCIQMC achieves a standard deviation within the range of chemical accuracy with only 10000 walkers.

In Fig. 2 (e) (f), we conversely plot the total walker number required for QC-FCIQMC runs assisted by ADAPT-VQE circuits of different depths to maintain a similar level of precision. The total number of walkers decreases by two orders of magnitude going from the basis generated by ADAPT-VQE adding 4 operators to that generated by ADAPT-VQE adding 24 operators.

As we discussed in the last section, one can sample the configurations $j$s for which one calculates the corresponding matrix elements $H_{ij}$s with respect to a given $i$. One can choose a certain threshold for the number of $j$s given $i$ where one stops sampling more configurations. In Fig. 3, we plot the effects of taking a limited number of samples for the selection of $j$ configurations spawning from $i$ in $N_2$ simulation. We again use the VQE basis generated by 12 layers of ADAPT-VQE. We stop the sampling procedure after reaching a certain number of $j$ configurations for each $i$. We adjust this threshold to obtain the five data points in the figure and one can see the fewer samples we take, the larger the deviation from the exact energy. We can also observe that the variance does not vary much with the number of samples, which instead depends more on the number of layers of the ADAPT-VQE as shown in Fig. 2(c).

We also test our QC-FCIQMC algorithm on the Hubbard model. Here we choose the Hubbard model of size 2, which requires 16 qubits and pick $U/t = 4$ to enable stronger static correlation. We also choose the chemical potential as half-filling. In Fig. 4, we compare the performance of QC-FCIQMC with the basis generated from the Hamiltonian

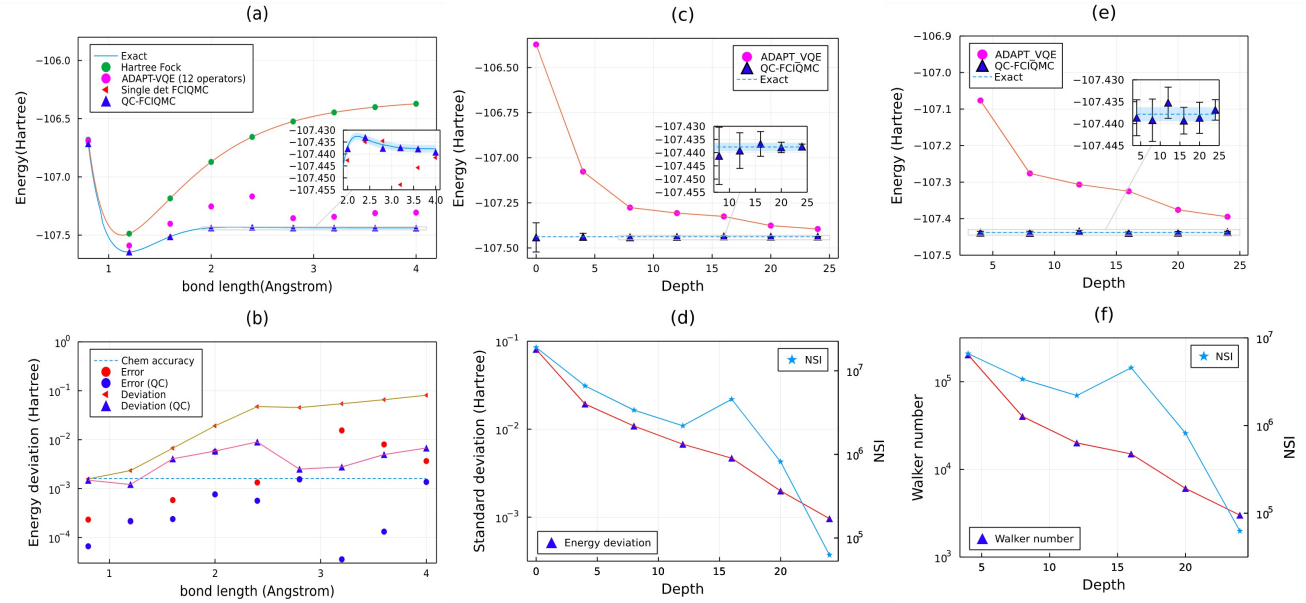


Figure 2. (a) Potential energy surface for the nitrogen molecule with QC-FCIQMC under the STO-3g basis set. (b) The standard deviation of energy evaluations along the QMC evolution. Here the new walker space of QC-FCIQMC is prepared with circuits from ADAPT-VQE (adding 12 fermionic operators for all bond lengths). Even with the assistance of shallow-depth ADAPT-VQE runs, QC-FCIQMC manages to reach chemical accuracy across all bond lengths using only around 10000 walkers. We also include data from FCIQMC with single-determinant walkers for comparison. For both single determinant FCIQMC and QC-FCIQMC, we start the energy shift when the total number of walkers exceeds 10000. Under this setting, single determinant FCIQMC fails to reach chemical accuracy at several bond lengths towards the dissociation limit. Moreover, the standard deviation from the energy profiling of QC-FCIQMC is smaller than the single determinant FCIQMC for all bond lengths. In fact, one observes a bigger variance reduction with a stronger static correlation. (c) FCIQMC-assisted ADAPT-VQE results for nitrogen molecule at bond length 4.0. ADAPT-VQE energies and QC-FCIQMC energies with standard deviations for different depths of ADAPT-VQE. (d) Log plot with standard deviations from (c) as well as the non-stoquastic indicator, where choose $\beta = 10^{-1}$. FCIQMC is able to push ADAPT-VQE results of different depths to chemical accuracy. Here we obtain the expected energy by taking the average of all FCIQMC energies after it converges. We fix the total walker number at the same level by starting to implement an energy shift of $S$ when the total walker number exceeds $10^4$. The deeper the ADAPT-VQE circuit is, the smaller the energy variance FCIQMC would have. One can see with an ADAPT-VQE optimization that picks 20 operators, the energy variance obtained is already at a similar level as chemical accuracy requirement and with 24 operators the variance lies well within chemical accuracy area. (e) ADAPT-VQE energies and QC-FCIQMC energies with variances for different depths of ADAPT-VQE. (f) Log 10 plots of the respective numbers of walkers that give the energies in (e) as well as the natural log of the non-stoquastic indicator, where choose $\beta = 10^{-1}$. To reach a certain level of energy variance, FCIQMC requires fewer walkers assisted with deeper ADAPT-VQE circuits. Here we choose the energy variance upper bound to be 5mHa for the number of walkers to be friendly to the running time of our code. To reach this precision, QC-FCIQMC requires about 100 times fewer walkers going from ADAPT-VQE with 4 operators to that of 24 operators.

variational ansatz (HV ansatz) [10–12] and the classical FCIQMC with single determinant basis. With the basis generated by a 15-layer HV ansatz, we observe the fluctuations of the energy much smaller in (a) and the walker population much more concentrated on a few configurations in (c) than the single-determinant basis in (b). We also plot the suppression of energy variance for HV ansatz with 1, 5, 10, and 15 layers in Fig. 4 (d) (e). We remark that the standard deviation of energy evaluation is reduced by almost one order of magnitude with merely a 1-layer HV ansatz.

[1] J. Preskill, Quantum **2**, 79 (2018).
[2] E. Altman, K. R. Brown, G. Carleo, L. D. Carr, E. Demler, C. Chin, B. DeMarco, S. E. Economou, M. A. Eriksson, K.-M. C. Fu, M. Greiner, K. R. Hazzard, R. G. Hulet, A. J. Kollár, B. L. Lev, M. D. Lukin, R. Ma, X. Mi, S. Misra, C. Monroe, K. Murch, Z. Nazario, K.-K. Ni, A. C. Potter, P. Roushan, M. Saffman, M. Schleier-Smith, I. Siddiqi, R. Simmonds, M. Singh, I. Spielman, K. Temme, D. S. Weiss, J. Vučković, V. Vuletić, J. Ye, and M. Zwierlein, PRX Quantum **2**, 017003 (2021).

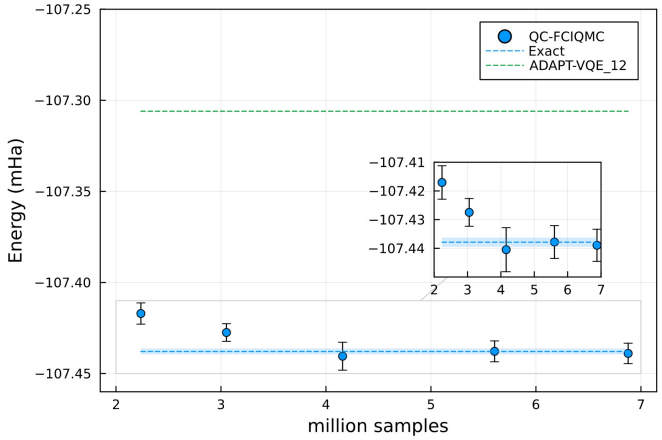


Figure 3. QC-FCIQMC on $N_2$ molecule at bond length 4.0$\mathring{A}$ when taking limited number of samples for the selection of $j$ configurations spawning from $i$. Here we use the VQE basis generated by 12 layers of ADAPT-VQE and the $x$-axis denotes the total number of samples of the whole QMC process. One can see the fewer samples we take, the larger the deviation from the exact energy. We can also observe that the variance does not vary much with the number of samples.

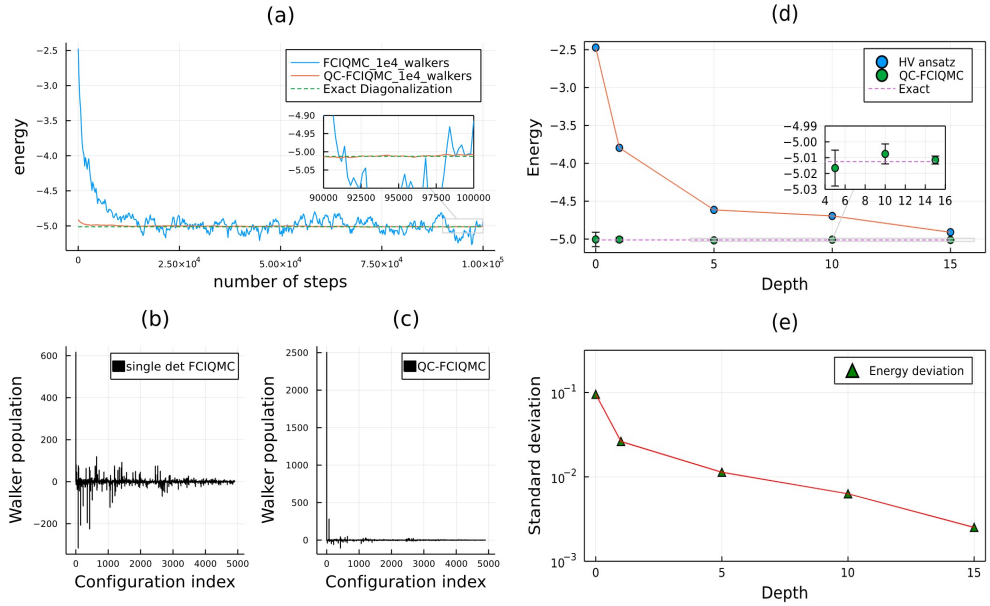


Figure 4. Comparison of the classical FCIQMC and QC-FCIQMC with basis generated by the Hamiltonian variational (HV) ansatz on the $2 \times 4$ Hubbard model ($U/t = 4$). The HV ansatz with $1, 5, 10, 15$ layers is implemented for comparison of variance suppression. The energy shift starts at $10^4$ walkers for all cases. (a) Comparison of energy fluctuation for FCIQMC and QC-FCIQMC. (b) Walker population for single determinant FCIQMC. (c) Walker population for QC-FCIQMC. (d) Energy estimation for different setups of layer number. (e) Effect of variance suppression.


[3] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Journal of the Physical Society of Japan **90**, 032001 (2021), https://doi.org/10.7566/JPSJ.90.032001.
[4] M. Cerezo, A. Poremba, L. Cincio, and P. J. Coles, Quantum **4**, 248 (2020).
[5] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Rev. Mod. Phys. **94**, 015004 (2022).
[6] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, Nature communications **5**, 1 (2014).
[7] S. B. Bravyi and A. Y. Kitaev, Annals of Physics **298**, 210 (2002).
[8] J. T. Seeley, M. J. Richard, and P. J. Love, The Journal of chemical physics **137**, 224109 (2012).
[9] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Nature comm. **10**, 1 (2019).
[10] D. Wecker, M. B. Hastings, and M. Troyer, Physical Review A **92**, 042303 (2015).

[11] C. Cade, L. Mineh, A. Montanaro, and S. Stanisic, Physical Review B **102**, 235122 (2020).
[12] Z. Cai, Physical Review Applied **14**, 014059 (2020).
[13] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Quantum Science and Technology **4**, 014008 (2018).
[14] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Physical review letters **120**, 110501 (2018).
[15] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Nature **603**, 416 (2022).
[16] G. H. Booth, A. J. Thom, and A. Alavi, The Journal of chemical physics **131**, 054106 (2009).
[17] D. Cleland, G. H. Booth, and A. Alavi, The Journal of chemical physics **132**, 041103 (2010).
[18] M. Troyer and U.-J. Wiese, Physical review letters **94**, 170201 (2005).
[19] S. Bravyi, D. P. Divincenzo, R. I. Oliveira, and B. M. Terhal, arXiv preprint quant-ph/0606140 (2006).
[20] M. Iazzi, A. A. Soluyanov, and M. Troyer, Physical Review B **93**, 115102 (2016).
[21] D. Hangleiter, I. Roth, D. Nagaj, and J. Eisert, Science advances **6**, eabb8341 (2020).
[22] R. Levy and B. K. Clark, Physical review letters **126**, 216401 (2021).
[23] L. Gupta and I. Hen, Advanced Quantum Technologies **3**, 1900108 (2020).
[24] M. Marvian, D. A. Lidar, and I. Hen, Nature communications **10**, 1 (2019).
[25] S. Zhang, J. Carlson, and J. E. Gubernatis, Physical Review B **55**, 7464 (1997).
[26] E. R. Anschuetz and B. T. Kiani, arXiv preprint arXiv:2205.05786 (2022).
[27] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature communications **9**, 1 (2018).
[28] W. Hoeffding, Journal of the American statistical association **58**, 13 (1963).
[29] L. Lin and Y. Tong, Quantum **4**, 372 (2020).
[30] L. Lin and Y. Tong, PRX Quantum **3**, 010318 (2022).
[31] K. Wan, M. Berta, and E. T. Campbell, Physical Review Letters **129**, 030503 (2022).
[32] G. Wang, D. Stilck-França, R. Zhang, S. Zhu, and P. D. Johnson, arXiv preprint arXiv:2209.06811 (2022).
[33] Y. Dong, L. Lin, and Y. Tong, PRX Quantum **3**, 040305 (2022).
[34] A. Y. Kitaev, arXiv preprint quant-ph/9511026 (1995).
[35] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (American Association of Physics Teachers, 2002).
[36] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Molecular Physics **109**, 735 (2011).
[37] H. Ni, H. Li, and L. Ying, arXiv preprint arXiv:2302.02454 (2023).
[38] M. Dobšíček, G. Johansson, V. Shumeiko, and G. Wendin, Physical Review A **76**, 030306 (2007).
[39] G. A. Quantum, Collaborators*†, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, *et al.*, Science **369**, 1084 (2020).
[40] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby, and A. Montanaro, Nature Communications **13**, 5743 (2022).
[41] T. E. O'Brien, G. Anselmetti, F. Gkritsis, V. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, *et al.*, arXiv preprint arXiv:2210.10799 (2022).
[42] S. Guo, J. Sun, H. Qian, M. Gong, Y. Zhang, F. Chen, Y. Ye, Y. Wu, S. Cao, K. Liu, *et al.*, arXiv preprint arXiv:2212.08006 (2022).
[43] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Proceedings of the National Academy of Sciences **115**, 9456 (2018).
[44] Q. Liang, Y. Zhou, A. Dalal, and P. D. Johnson, arXiv preprint arXiv:2306.17235 (2023).
[45] I. D. Kivlichan, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, *et al.*, Quantum **4**, 296 (2020).
[46] E. T. Campbell, Quantum Science and Technology **7**, 015007 (2021).
[47] S. Flannigan, N. Pearson, G. H. Low, A. Buyskikh, I. Bloch, P. Zoller, M. Troyer, and A. J. Daley, Quantum Science and Technology **7**, 045025 (2022).