

# Universal Hamming Weight Preserving Variational Quantum Ansatz

Ge Yan<sup>1</sup>, Kaisen Pan<sup>4</sup>, Ruocheng Wang<sup>4</sup>, Mengfei Ran<sup>3</sup>,  
Hongxu Chen<sup>1</sup>, Xunuo Wang<sup>4</sup>, Junchi Yan<sup>1,2\*</sup>

<sup>1</sup>Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China.

<sup>2</sup>School of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai, 200240, China.

<sup>2</sup>School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, 200240, China.

<sup>4</sup>Zhiyuan College, Shanghai Jiao Tong University, Shanghai, 200240, China.

\*Corresponding author(s). E-mail(s): [yanjunchi@sjtu.edu.cn](mailto:yanjunchi@sjtu.edu.cn);

## Abstract

Understanding the mathematical properties of variational quantum ansätze is crucial for determining quantum advantage in Variational Quantum Eigensolvers (VQEs). A deeper understanding of ansätze not only enriches theoretical discussions but also facilitates the design of more efficient and robust frameworks for near-term applications. In this work, we address the challenge of balancing expressivity and trainability by utilizing a Hamming Weight Preserving (HWP) ansatz that confines quantum state evolution to a symmetry-preserving subspace. We rigorously establish the necessary and sufficient conditions for subspace universality of HWP ansätze, along with a comprehensive analysis of the trainability. These theoretical advances are validated via the accurate approximation of arbitrary unitary matrices in the HWP subspace. Furthermore, the practical utility of the HWP ansatz is substantiated for solving ground-state properties of Fermionic systems, achieving energy errors below  $1 \times 10^{-10}$  Ha. This work highlights the critical role of symmetry-preserving ansätze in VQE research, offering insights that extend beyond supremacy debates and paving the way for more reliable and efficient quantum algorithms in the near term.

**Keywords:** Variational Quantum Eigensolver, Quantum Chemistry, Dynamical Lie Algebra, Symmetry-Preserving Quantum Ansatz

With the advent of the Noisy Intermediate-Scale Quantum (NISQ) era [1, 2], Variational Quantum Algorithms (VQAs) have emerged as a promising framework for tackling complex optimization problems, offering potential pathways to quantum advantage on near-term devices [3, 4]. Although these algorithms have shown significant promise across diverse applications [5–7], the absence of intrinsic evidence for quantum computational supremacy in VQAs underscores the need for deeper exploration into their capabilities and limitations.

A critical factor in evaluating the potential of VQAs lies in understanding the mathematical properties of the ansätze, including expressivity and trainability [8, 9]. Achieving a balance between these properties is particularly challenging in the exponentially large Hilbert space of quantum systems [10]. Symmetry-preserving ansätze, such as Hamming Weight Preserving (HWP) ansätze [4], offer a promising avenue for mitigating this complexity. HWP ansätze constrain the quantum state evolution to subspaces with fixed numbers of  $|1\rangle$ s, requiring specially designed operators to enforce these constraints. Such constraints can be mapped to a wide variety of fundamental physical symmetries, such as particle number conservation [11, 12] and total spin symmetry [13, 14]. Existing approaches [15–20] have largely adopted a bottom-up strategy, constructing HWP ansätze using operators later identified to possess the HWP property, where it remains a challenging task to determine whether the ansatz’s failure to achieve desired accuracy was due to insufficient expressibility, poor trainability leading to barren plateaus, or simply an inadequate number of trainable parameters.

In this work, we propose a novel top-down framework for constructing mathematically interpretable HWP ansätze that are theoretically rigorous, practically implementable, and capable of approximating arbitrary unitary matrices within the HWP subspace. Our approach addresses foundational challenges in ansatz design through three key contributions. First, we derive the conditions for two-qubit HWP operators to achieve universality in the HWP subspace by analyzing the Dynamical Lie Algebra (DLA) dimension, ensuring maximum expressivity. Second, we establish the trainability of the ansatz by deriving gradient variance expressions, revealing its resilience against barren plateaus [8, 21]. Third, we validate these theoretical insights through successful approximations of arbitrary unitary matrices within the HWP subspace [22–24], demonstrating the ansatz’s capability to efficiently capture target quantum states or transformations. The practical utility of the proposed ansatz is further demonstrated as a variational quantum eigensolver (VQE) for ground-state properties of Fermionic systems [7, 13, 25] and achieved remarkable accuracy across models.

### Theoretical Results of Universality

We begin by establishing the necessary and sufficient conditions for an HWP ansatz to be universal within the HWP subspace. By leveraging quantum optimal control theory [26], we identify these conditions through the analysis of the Dynamical Lie Algebra (DLA) generated by the HWP ansatz (refer to the Methods section for more details). In order to preserve the number of  $|1\rangle$ s in the quantum state, a two-qubit HWP gate acting on qubits  $i$  and  $j$  should have the following Hamiltonian, which only

operates on basis states  $|01\rangle$  and  $|10\rangle$ :

$$\mathbf{H}_{HWPij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a & b & 0 \\ 0 & \bar{b} & c & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{a+c}{2}\mathbf{E}_{ij} + \frac{a-c}{2}\mathbf{S}_{ij} + \frac{b+\bar{b}}{2}\mathbf{R}_{ij} + \frac{b-\bar{b}}{2i}\mathbf{J}_{ij}, \quad (1)$$

where  $a, c \in \mathbb{R}$ ,  $b \in \mathbb{C}$ , and  $\bar{b}$  denotes the complex conjugate of  $b$  since  $\mathbf{H} = \mathbf{H}^\dagger$ . We further decompose the Hamiltonian with four basis matrices:

$$\mathbf{R}_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{J}_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{E}_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{S}_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2)$$

To simplify the decomposition in Eq. 1, we denote  $r = \text{Re } b$ ,  $j = \text{Im } b$ ,  $e = \frac{a+c}{2}$ , and  $s = \frac{a-c}{2}$ , where  $r$  and  $j$  are the real and imaginary parts of  $b$ , respectively. For an HWP subspace of  $n$  qubits with an Hamming weight of  $k$ , the dimension is given by  $d_k = \binom{n}{k}$ . Since the number of initial generators for the DLA depends on the physical qubit connectivity, we begin by analyzing an ideal case where two-qubit HWP gates are applied between any pair of qubits. Theorem 1 outlines the conditions under which a two-qubit HWP gate is universal when full connectivity (FC) is available. These conditions provide a rigorous framework for understanding the expressiveness of the HWP ansatz.

**Theorem 1.** *For any  $n$  and  $k$ , a two-qubit HWP gate is universal with FC if and only if the coefficients satisfy one of the following two conditions:*

$$(1) \ e \neq 0, j \neq 0; \quad (2) \ e \neq 0, r \neq 0, s \neq 0. \quad (3)$$

By calculating the DLA dimension, we explore the commutation relations between the four basis matrices, where there are 16 possible combinations (whether  $e, r, j, s \neq 0$ ). These combinations are then classified into eight distinct DLA types, each with a unique dimensionality. Of the 16 combinations, only five yield full dimensionality in the FC configuration, meeting the conditions set forth in Theorem 1.

Given that achieving FC on most quantum processors remains a significant challenge, especially in the NISQ era, we also extend our analysis to a more realistic scenario involving nearest-neighbor (NN) connectivity. In this configuration, each physical qubit is only connected to its two adjacent qubits, forming a circular topology. The impact of this limited connectivity on the universality of the HWP ansatz is explored in Lemma 2.

**Lemma 2.** *If the set of generators contains all four basis matrices on NN qubits, which are  $\mathbf{R}_{i,i+1}$ ,  $\mathbf{J}_{i,i+1}$ ,  $\mathbf{E}_{i,i+1}$ ,  $\mathbf{S}_{i,i+1}$ , the dimension of DLA is  $d_k^2$ .*

The lemma reveals that if the generator set contains all four basis matrices acting on NN qubits, the DLA dimension remains  $d_k^2$ . This can be done simply by deriving the basis matrices on the nearest connected qubits  $\mathbf{R}_{i,i+1}$ ,  $\mathbf{J}_{i,i+1}$ ,  $\mathbf{E}_{i,i+1}$ ,  $\mathbf{S}_{i,i+1}$  to any two qubits  $\mathbf{R}_{i,j}$ ,  $\mathbf{J}_{i,j}$ ,  $\mathbf{E}_{i,j}$ ,  $\mathbf{S}_{i,j}$ . This result establishes a strong condition for

NN connectivity, and we proceed by reducing the number of basis matrices in the generator set. Given that NN connectivity allows for fewer generators than FC, the DLA dimension for NN is either smaller than or equal to that for FC. Thus, we focus on verifying the five combinations (including the one in Lemma 2) from the two primary conditions in Theorem 1.

**Corollary 3.** *For any  $n$  and  $k$ , a two-qubit HWP gate is universal with NN connectivity if and only if the coefficients satisfy the following conditions:*

$$(1) e \neq 0, j \neq 0, r \neq 0; \quad (2) e \neq 0, j \neq 0, s \neq 0; \quad (3) e \neq 0, r \neq 0, s \neq 0. \quad (4)$$

Corollary 3 indeed only rules out the combination  $j, e \neq 0, r = s = 0$ , the DLA dimension of which does not reach  $d_k^2$  when  $n = 2k$ . The above theorems offer top-down practical guidance for designing HWP ansätze and a robust framework for evaluating their effectiveness. While FC ensures maximum dimensionality, our findings show that even with the constraints of NN connectivity, universality can still be achieved under specific conditions.

### Theoretical Results of Trainability

We now turn our attention to the trainability of the proposed HWP ansatz, which is critical due to the known barren plateaus problem in VQAs [8, 21]. Typically, the gradients of the cost function decay exponentially with the number of qubits, scaling as  $\mathcal{O}(1/2^n)$ , since VQAs operate in the  $2^n$ -dimensional Hilbert space. However, as the HWP ansatz is confined to a  $d_k$ -dimensional subspace, we conjecture that the gradient decay should correspond to the subspace dimension, scaling as  $\mathcal{O}(1/d_k)$ . The following theorem provides the variance of the cost function gradients, using the framework established in [8]:

**Theorem 4.** *Consider an  $n$ -qubit quantum ansatz operating in the subspace with Hamming weight  $k$ . The variance of the cost function partial derivative is  $\text{Var}_\theta[\partial_l C] \approx \frac{16k^2(n-k)^2}{n^4 d_k}$ .*

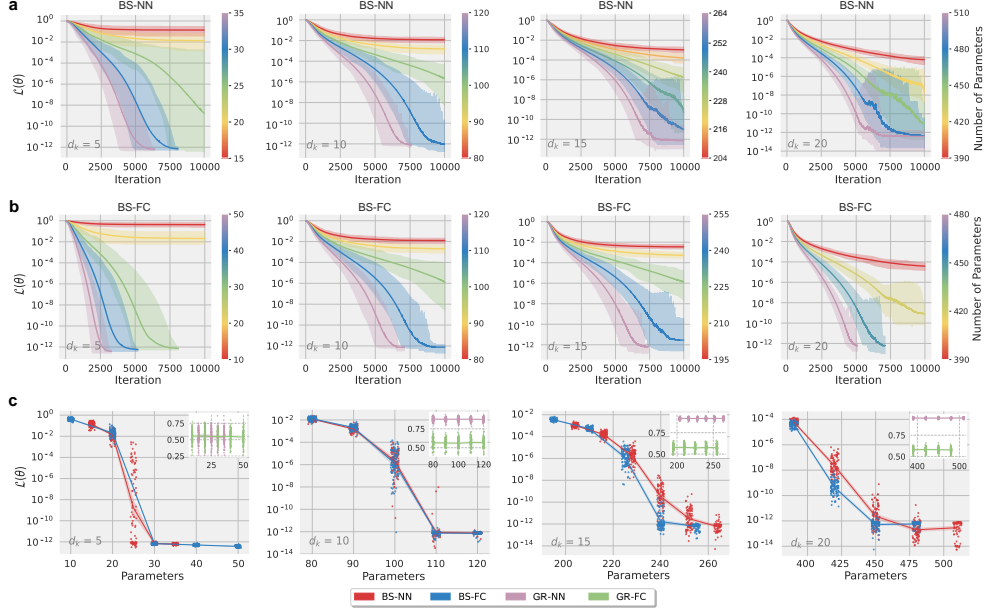
Theorem 4 supports our conjecture that the trainability of the circuit is governed by the subspace dimensionality  $d_k$  rather than the exponentially large qubit number  $2^n$ . As highlighted in [10], there exists a tradeoff between expressivity and trainability: more expressive ansätze might offer greater representational power, but they also suffer from diminished trainability, as optimization becomes increasingly difficult due to flat cost landscapes. Therefore, designing ansätze that operate within a restricted subspace, such as the HWP ansatz, may offer a potential solution to this tradeoff.

### Numerical Results for Unitary Approximation

To verify the theoretical results on the universality of different gates and connectivity, we provide numerical results on the unitary approximation problem [22–24]. The unitary approximation aims to solve the problem of whether  $\{\mathbf{U}(\boldsymbol{\theta})\}_\theta$  is equal to  $SU(N)$  (see lemma. 6). For a target unitary matrix  $\hat{\mathbf{U}}$  in  $d_k$ -dimensional HWP subspace, the loss function for unitary approximation is

$$\mathcal{L}_{UA}(\boldsymbol{\theta}) = 1 - |\text{Tr}(\hat{\mathbf{U}}^\dagger \mathbf{U}(\boldsymbol{\theta}))|^2 / d_k^2. \quad (5)$$

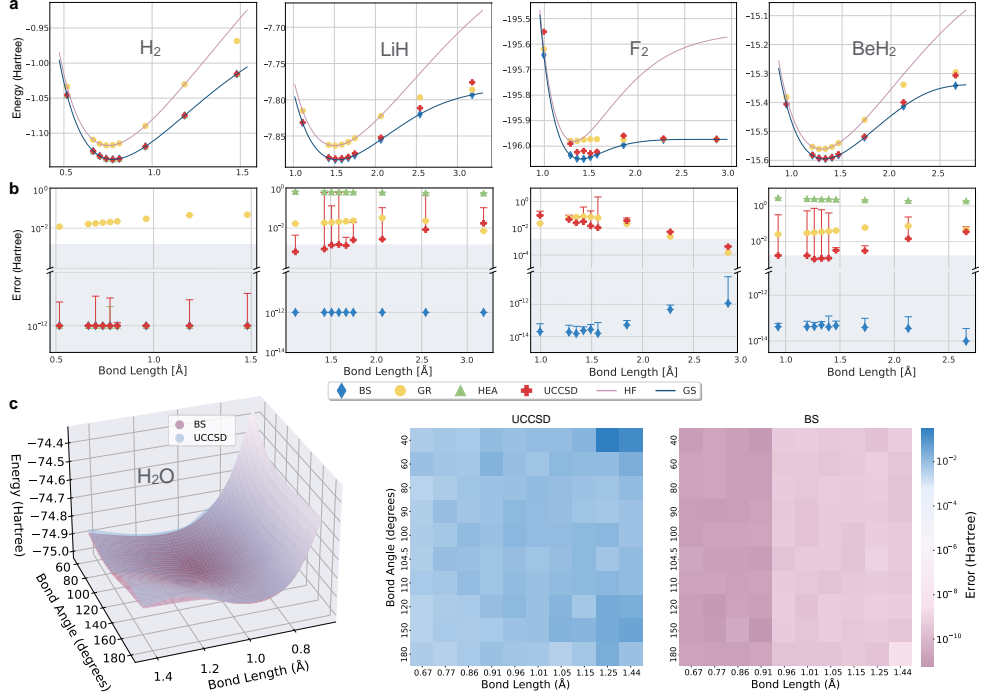




**Fig. 1 Results for unitary approximation.** We iterate through all the cases for  $d_k = \{\binom{5}{1}, \binom{5}{2}, \binom{6}{2}, \binom{6}{3}\} = \{5, 10, 15, 20\}$ , with both GR and BS gates for NN and FC connectivity. For each  $d_k$ , 100 unitary matrices are randomly sampled based on Haar measure [27, 28]. **a** The training curves for BS with NN connectivity. **b** The training curves for BS with FC connectivity. **c** The loss function is plotted versus the number of parameters. Both BS-NN and BS-FC show a similar decreasing pattern with the number of parameters needed for exact approximation at around  $d_k^2$ . Inset shows the results for GR-NN and GR-FC, with neither method getting close to  $\mathcal{L}(\theta) = 0$ .

We take condition (1) in Corollary 3 with  $e = r = j = 1$  and normalization coefficients to construct the ansatz, denoted as BS ansatz, with implementation details in the Method section. To illustrate ansatz with DLA dimension less than  $d_k^2$  can not approximate an arbitrary unitary matrix to any desired accuracy, we also provide results for the Givens Rotations (GR) [15, 16], which is widely used in Fermionic system simulation. Fig. 1 shows the results of minimizing the loss in Eq. 5 for four cases with HWP subspace dimension  $d_k = \{\binom{5}{1}, \binom{5}{2}, \binom{6}{2}, \binom{6}{3}\} = \{5, 10, 15, 20\}$ .

Our numerical results demonstrate that the theoretically derived BS gate can approximate a target unitary matrix with arbitrary precision, achieving a loss function value as low as  $1 \times 10^{-12}$ . In contrast, other commonly used HWP gates in quantum chemistry, such as the GR gate, fail to approximate arbitrary unitary matrices with similar accuracy. These numerical findings corroborate our theoretical predictions, confirming that the ansatz constructed from BS gates is universal within the HWP subspace, regardless of whether NN or FC connectivity is employed. Additionally, Fig. 1c illustrates the number of parameters (or layers) required by the ansatz to achieve this approximation. Our results support the conclusion of [29] in subspace that the bound on overparameterization is proportional to the DLA dimension of the ansatz, specifically  $d_k^2$  in our case. This further validates the applicability of overparameterization theory within the HWP subspace. The insights gained here are crucial

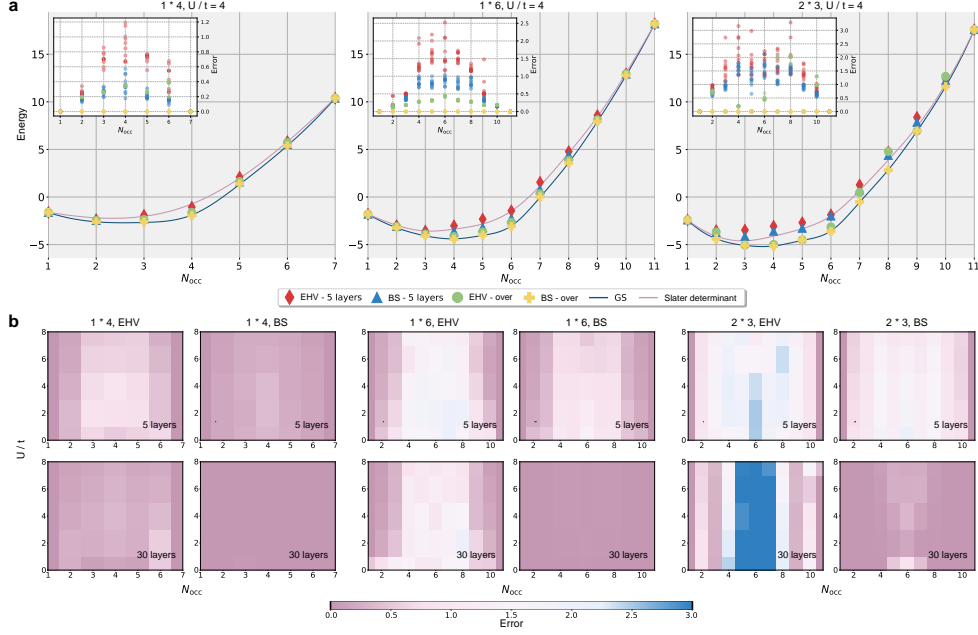


**Fig. 2 Results for simulating molecular electronic structures.** **a** Potential energy curves of four molecules w.r.t. the bond length, with the corresponding absolute errors compared to the exact energy shown in **b**. HF stands for the energy with Hartree-Fock state, and GS stands for the exact energy. The grey region shows results within chemical accuracy (less than 1.6 milli-hartrees). **c** Potential energy surface of  $H_2O$ , and the energy error of 100 sampled molecule structures for both UCCSD and BS. All the data points are a minimum of 10 random seeds, with the error bars indicating the range from minimum to maximum.

for guiding the design of HWP ansätze and their practical application in ground-state energy estimation challenges.

### Numerical Results for Solving Electronic Structures

We then utilize the proposed BS ansatz to address the challenges of quantum chemistry simulation [7, 25, 30]. The Hamiltonians of the molecules are under the Born-Oppenheimer approximation in the second quantization and mapped to qubits with the Jordan-Wigner transformation. The ground state of molecules can be prepared on a quantum computer by designing a parameterized quantum circuit  $U(\theta)$  and minimizing the loss function  $\mathcal{L}(\theta) = \langle \psi_0 | U(\theta)^\dagger \mathcal{H} U(\theta) | \psi_0 \rangle$ , where  $\mathcal{H}$  is the Hamiltonian of the molecule. In Fig. 2, we show the calculation of the potential energy curves w.r.t. bound length for  $H_2$  (4 qubits),  $LiH$  (12 qubits),  $BeH_2$  (14 qubits) and  $F_2$  (20 qubits), and the potential energy surface for  $H_2O$  (14 qubits). Detailed statistics of the molecules and the hyperparameter analysis can be found in Supplementary Note D.2. To better illustrate the efficiency of the proposed BS ansatz, we select HEA [7] and UCCSD [6]



**Fig. 3 Results for the Fermi-Hubbard model.** We conduct experiments for Fermi-Hubbard model instances on lattices of shapes  $1 \times 4$ ,  $1 \times 6$ , and  $2 \times 3$ . **a** Energy of the Fermi-Hubbard model (at  $U/t = 4$ ) w.r.t. integer number of occupations. The inset shows the energy error with five random seeds for each method. The "over" in EHV-over and BS-over represents the number of layers beyond the overparameterization bound. **b** The energy error of BS and EHV at low circuit depth. In all panels, the X-axis gives the occupation number, and the Y-axis gives the  $U/t$ . All the results are an average of 5 random seeds.

ansatz as the baselines. We also involve GR, representing the single excitation term of the UCC model [6]. Both BS and GR are under NN connectivity.

Our experiments demonstrate that the ansatz constructed from BS gates achieves an error level of  $1 \times 10^{-10}$  Ha for molecular ground state energies, consistent across all tested bond lengths and bond angles. This precision significantly exceeds that of existing VQE methodologies. Furthermore, in our unitary approximation task, we established that any unitary matrix within the HWP subspace can be accurately approximated by our proposed BS-gate ansatz. Consequently, this implies that any state within the HWP subspace is accessible through our ansatz, effectively rendering it a quantum circuit-based implementation of the FCI [31] method.

### Numerical Results for Fermi-Hubbard Model

To further illustrate the ansatz-free HWP circuit is capable of solving various Fermionic systems, we conduct experiments on one more iconic model in condensed-matter physics [32], the Fermi-Hubbard model [13]. It is the simplest system that includes non-trivial correlations and has been widely proposed as an early target for variational quantum algorithms [15, 16, 33–36]. The Hamiltonian of the

Fermi-Hubbard model can be written in the following form:

$$H_{HF} = -t \sum_{i,j} \sum_{\sigma} t_{ij} (a_{i\sigma}^{\dagger} a_{j\sigma} + a_{j\sigma}^{\dagger} a_{i\sigma}) + U \sum_i a_{i\uparrow}^{\dagger} a_{i\uparrow} a_{i\downarrow}^{\dagger} a_{i\downarrow}, \quad (6)$$

where  $i, j$  denotes adjacent sites, and  $\sigma \in \{\uparrow, \downarrow\}$  denotes the spin. Each site of the Fermi-Hubbard model contains two fermionic modes, and each mode is mapped to one qubit under the Jordan-Wigner transformation.

Numerical results for Fermi-Hubbard model instances on lattices of shape  $1 \times 4$ ,  $1 \times 6$ , and  $2 \times 3$  are presented in Fig. 3. To assess the efficiency of the proposed HWP ansatz, we compared its performance against the ground state (GS, the exact energy), the Slater determinant state (classical ansatz) [37, 38], and the Efficient Hamiltonian Variational (EHV) ansatz[36], a VQE approach composed of GR, XY, and onsite gates (details on these methods are provided in Supplementary Note E).

We first computed the energy for all instances across different occupation numbers at  $U/t = 4$ , focusing on scenarios with both extremely low circuit depths (5 layers) and sufficiently deep circuits to achieve overparameterization (100 layers for 4 sites and 200 layers for 6 sites; see Supplementary Note E). Maintaining a relatively low circuit depth is crucial for exploring the practical utility of NISQ devices. The results reveal that the BS ansatz consistently outperforms other baseline methods across all circuit depths, with errors below  $1 \times 10^{-10}$  in the overparameterized regime. We further examined the robustness of the BS ansatz by varying  $U/t$  and found that increasing the number of layers steadily improves its performance, whereas the EHV ansatz shows limited improvement. Additionally, it appears that the noninteracting ( $U/t = 0$ ) Fermi-Hubbard model poses more challenges for both VQE methods when underparameterized, likely due to the introduction of unnecessary phases. Nonetheless, we demonstrate that the BS ansatz consistently outperforms existing Hamiltonian variational ansätze, delivering superior results even with a minimal number of layers.

This work rigorously establishes the necessary and sufficient conditions for an HWP ansatz to achieve subspace universality, along with a comprehensive analysis of its trainability. These conditions can be further extended to accommodate different physical qubit connectivity configurations and tailored for specific values of  $n$  and  $k$ , enabling the design of HWP operators with simpler decompositions for broader applicability in real-world problems. The proposed ansatz represents a significant step toward addressing the challenges of VQEs, striking a critical balance between expressivity and trainability while offering inherent error detection capabilities. Any deviation in Hamming weight serves as a direct indicator of bit-flip errors, making it highly suitable for noisy environments. By providing a mathematically interpretable, noise-resilient framework, this ansatz presents a compelling pathway for advancing discussions of VQE supremacy in the near term.

## Method

### Dynamical Lie Algebra and Controllability of Quantum Systems

We first introduce a well-defined mathematical tool to derive the theoretical findings, which is Dynamical Lie Algebra (DLA) [39, 40]. Lie algebraic techniques have been widely used in discussing the controllability of quantum systems due to the unitary transformation nature of the quantum circuits. For a more intuitive understanding, the Dynamical Lie Group  $\mathbb{G}$  comprises unitary matrices, and the corresponding Dynamical Lie Algebra  $\mathfrak{g}$  comprises anti-Hermitian matrices. For a  $L$ -gate parameterized quantum circuit, we can denote the ansatz as a unitary transformation as

$$\mathbf{U}(\boldsymbol{\theta}) = \prod_{l=1}^L \mathbf{U}_l(\theta_l) = \prod_{l=1}^L e^{i\theta_l \mathbf{H}_l}, \quad (7)$$

where  $\mathbf{U}_l = e^{i\mathbf{H}_l \theta_l}$ ,  $i\mathbf{H}_l$  are anti-Hermitian matrices, and  $\boldsymbol{\theta} = \{\theta_1, \theta_2 \dots \theta_L\}$  are the parameters. We take these distinct Hamiltonians in the circuit as a set of generators  $\mathcal{G} = \{i\mathbf{H}_p\}_{p=1}^P$ , where  $|\mathcal{G}| = P$ . The DLA can be defined as

**Definition 1. Dynamical Lie Algebra (DLA):** Consider the set of generators  $\mathcal{G} = \{i\mathbf{H}_p\}_{p=1}^P$ , the DLA  $\mathfrak{g}$  is defined as:

$$\mathfrak{g} = \text{span}\langle i\mathbf{H}_1, i\mathbf{H}_2, \dots, i\mathbf{H}_P \rangle_{\text{Lie}}, \quad (8)$$

where  $\langle \cdot \rangle_{\text{Lie}}$  denotes the Lie closure.

The DLA  $\mathfrak{g}$  is calculated by repeatedly taking the commutator of the elements in the set of generators. The commutator of matrices  $\mathbf{A}$  and  $\mathbf{B}$  can be defined as  $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$ . The reachable unitary matrices of the parameterized quantum ansatz with arbitrary parameters  $\boldsymbol{\theta}$  can be denoted as  $\{\mathbf{U}(\boldsymbol{\theta})\}_{\boldsymbol{\theta}}$ . We introduce the following lemma

**Lemma 5.** [26] A quantum system  $\hat{\mathbf{H}}$  is completely controllable if  $\{\mathbf{U}(\boldsymbol{\theta})\}_{\boldsymbol{\theta}} = \mathbb{G} = \text{SU}(N)$ .

Complete controllability indicates that any desired quantum state evolution can be achieved using the ansatz. In other words, the system's dynamics can be fully manipulated to reach any state in the given space from any initial state. We further borrow an important conclusion from [26], which provides a simple way to verify the controllability of a quantum system, namely by computing the dimension of the DLA.

**Lemma 6.** [26] A necessary and sufficient condition for complete controllability of a quantum system  $\hat{\mathbf{H}}$  is that the dimension of the DLA  $\mathfrak{g}$  is  $N^2$ .

A detailed computational method to calculate the dimension of DLA is described in Supplementary Notes Alg. 1 [41]. This lemma is then utilized to derive the conditions for universal HWP operators in the theoretical results.

### Hamming Weight Preserving Operators

Based on the definition of Hamming weight and Hamming distance, we can thus define a type of quantum circuit, namely Hamming Weight Preserving quantum circuit, where the number of non-zero elements in the quantum state remains invariant throughout its evolution. This type of circuit naturally enforces symmetry within the

model by treating it as a hard constraint on the circuit. To ensure that the quantum circuit preserves the Hamming weight of the states, the Hamiltonian of two-qubit HWP gates takes the following form:

$$\mathbf{H}_{HW} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a & b & 0 \\ 0 & \bar{b} & c & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (9)$$

where  $a$  and  $c$  are real values and  $\bar{b}$  denotes the complex conjugate of  $b$ . Two-qubit HWP gates with this structure act exclusively on the basis states  $|01\rangle$  and  $|10\rangle$ , thereby maintaining the Hamming weight of the quantum state throughout computation. This characteristic not only enforces particle number symmetry but also provides resilience against bit-flip errors, as any such error would alter the Hamming weight, making it easily detectable with a simple parity check. Furthermore, incorporating symmetry verification as an additional layer can enhance the robustness and performance of the HWP ansatz on Noisy Intermediate-Scale Quantum (NISQ) devices, making it particularly suited for near-term quantum applications. In Fig. 4, we provide several two-qubit gates that have been identified as HWP gates in previous works [5, 15, 16, 18–20, 36], fitting the form in Equation 9. However, none of these gates have been rigorously analyzed in terms of their expressivity within the HWP subspace, and their performance has been evaluated only on specific tasks with limited ansatz layers, leaving them underparameterized.

#### Sketch of the Proof of Theorem 1

The primary aspect to understand is the commutation relations of the four basis matrices. We begin by analyzing the commutators for two qubits in the same set ( $ij$  and  $ij$ ) and for different qubits ( $ij$  and  $jk$ ). It is important to note that basis matrices acting on entirely distinct qubits, such as  $ij$  and  $kl$ , do not commute, and therefore this case is excluded from our analysis. As shown in Table 4, the commutators between basis matrices on  $ij$  and  $jk$  involve the Pauli-Z operator. Consequently, we extend the analysis to include commutators between basis matrices and those tensored with the Pauli-Z operator. The results in Table 4 demonstrate that these commutators form a closed algebra, generating no new elements.

There are fifteen distinct configurations of the HWP Hamiltonian, obtained by combining the four basis matrices (excluding the trivial all-zero case). Utilizing the commutation relations derived earlier, we compute the final commutation outcomes for these configurations and determine the corresponding DLA dimensions. As shown in Table 4, the analysis reveals eight distinct types of DLA, with only one having a dimension of  $d_k^2$ . The five configurations that result in this specific DLA type can be reduced to the following two conditions:

$$(1) \ e \neq 0, \ j \neq 0; \quad (2) \ e \neq 0, \ r \neq 0, \ s \neq 0. \quad (10)$$

### Proof of Lemma 2

We first recall Lemma 2, which is derived from Theorem 1 and critical for analyzing the DLA dimension with NN connectivity.

**Lemma.** *If the set of generators contains  $\mathbf{R}_{i,i+1}$ ,  $\mathbf{J}_{i,i+1}$ ,  $\mathbf{E}_{i,i+1}$ ,  $\mathbf{S}_{i,i+1}$ , the dimension of DLA is  $d_k^2$ .*

*Proof.*

$$[\mathbf{J}_{i,i+1}, \mathbf{E}_{i+1,i+2}] = -i\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z \quad (11)$$

$$[\mathbf{R}_{i,i+1}, \mathbf{E}_{i+1,i+2}] = i\mathbf{J}_{i,i+1} \otimes \sigma_{i+2}^z \quad (12)$$

Then it can generate all  $\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z$  and  $\mathbf{J}_{i,i+1} \otimes \sigma_{i+2}^z$ , and

$$[\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z, \mathbf{R}_{i+1,i+2}] = i\mathbf{J}_{i,i+2} \quad (13)$$

$$[\mathbf{J}_{i,i+1} \otimes \sigma_{i+2}^z, \mathbf{R}_{i+1,i+2}] = -i\mathbf{R}_{i,i+2} \quad (14)$$

Then it can generate all  $\mathbf{R}_{i,i+2}$  and  $\mathbf{J}_{i,i+2}$ , and so that all  $\mathbf{R}_{ij}$ ,  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$  ( $[\mathbf{R}_{ij}, \mathbf{J}_{ij}] = -2i\mathbf{S}_{ij}$ ). With  $[\mathbf{J}_{ik}, [\mathbf{J}_{ij}, \mathbf{R}_{jk}]] = 2\mathbf{S}_{ik} \otimes \sigma_j^z$ , all the  $\mathbf{S}_{ij} \otimes \sigma_k^z$  can be generated too. We notice that

$$\mathbf{S}_{ij} \otimes \sigma_k^z + \mathbf{S}_{jk} \otimes \sigma_i^z = -\mathbf{E}_{ij} + \mathbf{E}_{jk} \quad (15)$$

Therefore, all  $\mathbf{E}_{ij}$  can be generated by  $\mathbf{S}_{i+1,i} \otimes \sigma_j^z + \mathbf{S}_{i,j} \otimes \sigma_{i+1}^z + \mathbf{E}_{i,i+1} = \mathbf{E}_{ij}$ , so the DLA can be specified to type I with the dimension  $d_k^2$ .  $\square$

### Sketch of the Proof of Corollary 3

Notice that NN connectivity allows for fewer generators than FC, which indicates the DLA dimension for NN is either smaller than or equal to that for FC. Consequently, we only need to verify the five configurations in DLA type I (including the one derived in Lemma 2). In this section, we focus on the unique configuration ( $j, e \neq 0, r = s = 0$ ) where a reduction in DLA dimension occurs, leaving the derivation for the remaining configurations to Supplementary Note B.2.

As shown in Equation 11, the commutator of  $\mathbf{J}_{i,i+1}$  and  $\mathbf{E}_{i,i+1}$  is  $\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z$ , from which we can successively derive the following results:

$$\begin{aligned} [\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z, \mathbf{J}_{i+1,i+2}] &= -\mathbf{R}_{i+2,i} \\ [\mathbf{R}_{i,i+2}, \mathbf{E}_{i+2,i+3}] &= i\mathbf{J}_{i,i+2} \otimes \sigma_{i+3}^z \\ [\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z, \mathbf{R}_{i,i+2}] &= i\mathbf{J}_{i,i+1}. \end{aligned} \quad (16)$$

We are unable to generate  $\mathbf{R}_{i,i+1}$  from the given generators, which also precludes the generation of  $\mathbf{S}_{i,i+1}$  as well. As a result, the DLA dimension for this configuration with NN connectivity is given by

$$\dim(\text{DLA}) = \begin{cases} d_k^2 & \text{if } k < n/2 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases} \quad (17)$$

Therefore for any  $n$  and  $k$ , a two-qubit HWP gate is universal with nearest neighbor connectivity if and only if the coefficients satisfy the following conditions:

$$(1) \ e \neq 0, j \neq 0, r \neq 0; \quad (2) \ e \neq 0, j \neq 0, s \neq 0; \quad (3) \ e \neq 0, r \neq 0, s \neq 0. \quad (18)$$

#### Sketch of the Proof of Theorem 4

Consider the partial derivative of the cost function  $C$  with respect to the  $l$ -th parameter  $\theta_l$ . We decompose the ansatz into two sections:  $\mathbf{U}_-$  representing the unitary matrix of the circuit before gate  $\mathbf{H}_l$ , and  $\mathbf{U}_+$  denotes the unitary matrix after gate  $l$ . The variance of the partial derivative can be expressed as:

$$Var_\theta[\partial_l C] = \int_{\mathbf{U}_+} d\mathbf{U}_+ \int_{\mathbf{U}_-} d\mathbf{U}_- (\partial_l C(\theta))^2, \quad (19)$$

where  $\partial_l C(\theta) = \partial_l \left( \text{Tr}(\mathbf{U}_- \rho \mathbf{U}_-^\dagger O_+) \right) = i \text{Tr}(\mathbf{U}_- \rho \mathbf{U}_-^\dagger [\mathbf{H}_l, O_+])$ , with  $\rho$  as the input state and  $O$  as the observable for measurement. Thus, the variance becomes:

$$Var_\theta[\partial_l C] = \int_{\mathbf{U}_+} d\mathbf{U}_+ \int_{\mathbf{U}_-} d\mathbf{U}_- \left( i \text{Tr}(\mathbf{U}_- \rho \mathbf{U}_-^\dagger [\mathbf{H}_l, O_+]) \right)^2, \quad (20)$$

where the initial state is typically a pure state in the  $d_k$ -dimensional HWP subspace. Therefore, by setting  $\text{Tr}(\rho) = 1$  and  $\text{Tr}(\rho^2) = 1$ , we derive:

$$\begin{aligned} Var_\theta[\partial_l C] &= -\frac{1}{d_k(d_k + 1)} \int_{\mathbf{U}_+} d\mathbf{U}_+ \text{Tr}([H_l, O_+]^2) \\ &= -\frac{2 \text{Tr}(\mathbf{H}_l^2)}{d_k(d_k + 1)} \left( \frac{\text{Tr}^2(O) - d_k \text{Tr}(O^2)}{d_k^2 - 1} \right). \end{aligned} \quad (21)$$

Without loss of generality, we set the observable  $O$  as  $Z_0$  since other observables yield the same magnitude. This gives  $\text{Tr}(O) = \frac{d_k(n-2k)}{n}$  and  $\text{Tr}(O^2) = d_k$ . Substitute these into Equation 21, we obtain:

$$\begin{aligned} Var_\theta[\partial_l C] &= -\frac{2}{d_k(d_k + 1)} \times \frac{2k(n-k)d_k}{n(n-1)} \times \left( \frac{\frac{d_k^2(n-2k)^2}{n^2} - d_k^2}{d_k^2 - 1} \right) \\ &= \frac{16k^2(n-k)^2 d_k^2}{(d_k + 1)n^3(n-1)(d_k^2 - 1)} \approx \frac{16k^2(n-k)^2}{n^4 d_k}. \end{aligned} \quad (22)$$

Upon further analysis, we find that if  $k = 1$ , then  $Var_\theta[\partial_l C] \approx \frac{16}{n^3}$ . Conversely, when  $k = \frac{n}{2}$ ,  $Var_\theta[\partial_l C] \approx \left(\frac{n}{n/2}\right)^{-1}$ , which is approximate to exponentially small. This result supports the conjecture that the trainability of the circuit is closely tied to  $d_k$ , where a smaller  $d_k$  leads to improved trainability. For detailed derivation, see Supplementary Note C.



### Implementation Details

Experiments were conducted on a machine with 2TB memory, 100 cores Intel Xeon Platinum 8480+ CPU, and 8 GPUs (Nvidia H100), with a Python simulator simulating the ansätze.

**BS Ansatz:** The BS gate we utilized to construct the ansatz is one of the two-qubit HWP operators derived from the theoretical results. The BS gate follows the condition (1) in Corollary 3, so the Hamiltonian for BS gate is thus

$$\mathbf{H}_{BS} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1+i}{2\sqrt{2}} & 0 \\ 0 & \frac{1-i}{2\sqrt{2}} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (23)$$

with  $\mathbf{H}_{BS} = \mathbf{H}_{BS}^2$ . We obtain the unitary matrix of the BS gate by Taylor expansion:

$$\mathbf{U}_{BS}(\theta) = e^{i\theta\mathbf{H}_{BS}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{(e^{i\theta}+1)}{2} & \frac{(1+i)(e^{i\theta}-1)}{2\sqrt{2}} & 0 \\ 0 & \frac{(1-i)(e^{i\theta}-1)}{2\sqrt{2}} & \frac{(e^{i\theta}+1)}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (24)$$

where  $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ . A possible decomposition of the BS gate is illustrated in Fig. 5, each BS gate requires three CNOT gates to implement. We also show two BS ansatz implementations with different physical qubit connectivity.

Since the fully connected physical qubit topology is hard to realize on existing quantum hardware, we lay more emphasis on the NN connectivity ansatz and the corresponding expressivity and trainability. For directed two-qubit HWP operators, which indicates  $\mathbf{U}_{ij} \neq \mathbf{U}_{ji}$  with  $i, j$  as the qubit index, we can alternatively insert layers containing reversed HWP operators as in Fig. 5b to reach the maximum expressivity.

**Molecular Electronic Structures:** We begin by estimating the ground state of molecular systems, focusing on molecular Hamiltonians within the Born-Oppenheimer approximation and expressed in second quantization. These Hamiltonians, obtained from the Python package PySCF [42], are transformed into qubit form using the minimal basis set STO-3G and the Jordan-Wigner transformation. The Hamiltonian in second quantization is given by:

$$\mathcal{H} = \sum_{i,j} h_{ij} \hat{a}_j^\dagger \hat{a}_i + \frac{1}{2} \sum_{i,j,p,q} g_{ijpq} \hat{a}_j^\dagger \hat{a}_q^\dagger \hat{a}_p \hat{a}_i, \quad (25)$$

where  $\hat{a}_i = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$  denotes the annihilation operator on qubit  $i$ , corresponding to the  $i$ -th molecular orbital in the active space. The coefficients  $h_{ij}$  and  $g_{ijpq}$  denote the one- and two-electron integrals, respectively. A well-established solving approach is the VQE method, with the chemically inspired unitary coupled cluster (UCC) ansatz [6, 30, 43]. The coupled cluster operator, in second quantization, is defined as  $\hat{T} = \hat{T}_1 +$

$\hat{T}_2 + \dots + \hat{T}_v$  with the single and double excitation operators as

$$\hat{T}_1 = \sum_{i,j} \hat{a}_j^\dagger \hat{a}_i, \quad \hat{T}_2 = \sum_{i,j,p,q} \hat{a}_j^\dagger \hat{a}_q^\dagger \hat{a}_p \hat{a}_i. \quad (26)$$

For a given reference initial state  $|\psi_0\rangle$ , the UCC ansatz wave function is given by  $|\psi\rangle = e^{\hat{T}-\hat{T}^\dagger} |\psi_0\rangle$  [43], where  $T - T^\dagger$  is an anti-Hermitian operator which makes it suitable for quantum computers since the exponential of an anti-Hermitian operator is a unitary operator. The Hamiltonian for the single excitation term is

$$\mathcal{H}_{single} = \frac{1}{i}(\hat{T} - \hat{T}^\dagger) = \frac{1}{i} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (27)$$

which is indeed the GR we utilize as the baseline in unitary approximation. The Hamiltonian for the double excitation operator, on the other hand, is more complex, involving a  $16 \times 16$  matrix. This increased complexity significantly challenges implementation on NISQ devices.

In Figure 2, we conduct numerical experiments on five molecules. For  $H_2$ ,  $LiH$ ,  $F_2$ , and  $BeH_2$ , the data points across varying bond lengths are obtained by iterating over the coefficients in the set  $\{0.7, 0.9, 0.95, 1, 1.05, 1.1, 1.3, 1.6, 2\}$  time the minimum-energy bond length. This setting focuses on both the performance near the minimum-energy point and at extended bond lengths. The maximum number of iterations is capped at 10,000, and all the results are a minimum of 10 random seeds, with the error bar indicating the range of 10 seeds. Since the number of layers is kept identical across all bond lengths, the results show that increasing bond length leads to greater variance in the BS ansatz, suggesting that the problem becomes more challenging as bond lengths increase. For the water molecule, we explore the potential energy surface with respect to both bond length and bond angle. The bond length coefficients are sampled from the set  $\{0.7, 0.8, 0.9, 0.95, 1, 1.05, 1.1, 1.2, 1.3, 1.5\}$ , differing from the other molecules to achieve a more evenly distributed bond lengths. The bond angles are selected from  $\{40^\circ, 60^\circ, 80^\circ, 90^\circ, 100^\circ, 104.5^\circ, 110^\circ, 120^\circ, 150^\circ, 180^\circ\}$ .

**The Fermi-Hubbard Model:** The Fermi-Hubbard model is a fundamental model in condensed matter physics used to study the behavior of interacting fermions on a lattice. The electrons move between discrete sites on a lattice, and each site can be occupied by up to two electrons. The Hamiltonian of the Fermi-Hubbard model can be written in the following form:

$$H_{HF} = -t \sum_{i,j} \sum_{\sigma} t_{ij} (a_{i\sigma}^\dagger a_{j\sigma} + a_{j\sigma}^\dagger a_{i\sigma}) + U \sum_i a_{i\uparrow}^\dagger a_{i\uparrow} a_{i\downarrow}^\dagger a_{i\downarrow}, \quad (28)$$

where  $i, j$  denotes adjacent sites, and  $\sigma \in \{\uparrow, \downarrow\}$  denotes the spin. Each site contains two fermionic modes corresponding to spin-up and spin-down states. Unlike molecular Hamiltonians, the Fermi-Hubbard Hamiltonian is more complex as it includes both

intra-site interactions and inter-spin sector couplings. The first term in eq. 6 is the hopping term, also known as the XY-interaction [18, 19]

$$\mathcal{H}_{XY} = a_i^\dagger a_j + a_j^\dagger a_i = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \frac{1}{2}(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y), \quad (29)$$

which is similar to the GR representing the creation and annihilation of the electrons. The second term in Equation 6 is the onsite term, adding a phase to the state  $|11\rangle \langle 11|$ :

$$\mathcal{H}_{\text{onsite}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (30)$$

which represents the repulsive Coulomb interaction when two electrons with opposite spins occupy the same site. [36] utilized both terms to construct a Hamiltonian variational ansatz (the EHV ansatz) by applying NN-connected XY-interaction and onsite layers alternately to achieve the desired accuracy.

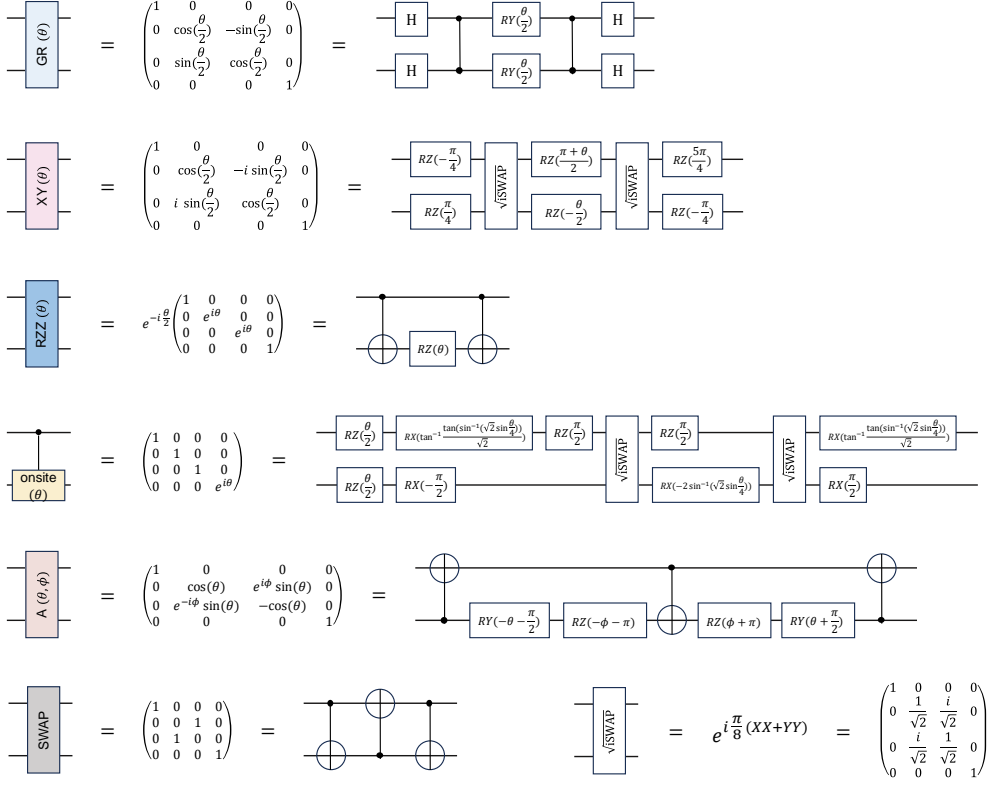
In Figure 3, we present numerical results for both 1D ( $1 \times 4$  and  $1 \times 6$ , where all sites lie along a line) and 2D ( $2 \times 3$ , arranging on a rectangular lattice) Fermi-Hubbard models. Although the  $1 \times 6$  and  $2 \times 3$  configurations have the same number of sites, the  $2 \times 3$  model allows for additional hopping interactions (both vertical and horizontal), making it fundamentally different from the 1D configuration. In this experiment, we focus more on the performance of the ansätze when underparameterized. We report results for 5 layers, 30 layers, and a sufficient number of layers to reach overparameterization. The results demonstrate that the NN-connected BS ansatz is efficient across all layers, consistently outperforming the Hamiltonian variational ansatz. It is worth noting that the EHV ansatz requires interactions between qubits on the same site and adjacent sites within the same spin sector, which is impractical on NISQ devices without introducing qubit swaps. These additional SWAP gates contribute to increased circuit compilation overhead, further highlighting the advantages of the proposed BS ansatz for near-term quantum hardware.

## Author Contributions

G.Y. and K.P. initiated the project, developed the theoretical framework and implemented the source code for the simulation. R.W. derived the universal conditions. K.P. and M.R. conducted numerical experiments on the Unitary approximation task. R.W. conducted numerical experiments on the Fermi-Hubbard model. M.R., H.C. and X.W. conducted numerical experiments on the molecular electronic structure simulation. G.Y. proposed experiments, interpreted the results, and wrote the paper with input from all authors. J.Y. wrote the paper and supervised this project.

## Competing interests

The authors declare no competing interests.



**Fig. 4** Two-qubit HWP gates in related works with their unitary matrices and possible decomposition. The two-qubit elementary gates utilized in the decomposition involve CX, CZ, and  $\sqrt{i}$ SWAP gates (unitary matrix in the bottom right). From top to bottom, we have Givens Rotations [15, 16], XY-interaction [18, 19], RZZ [5], onsite gate [36], A gate [20], and non-parameterized SWAP gate.

	$\mathbf{R}_{ij}$	$\mathbf{J}_{ij}$	$\mathbf{E}_{ij}$	$\mathbf{S}_{ij}$		$\mathbf{R}_{ij}$	$\mathbf{J}_{ij}$	$\mathbf{E}_{ij}$	$\mathbf{S}_{ij}$
$\mathbf{R}_{ij}$	0	$\mathbf{S}_{ij}$	0	$\mathbf{J}_{ij}$	$\mathbf{R}_{jk}$	$\mathbf{J}_{ik} \otimes \sigma_j^z$	$\mathbf{R}_{ik} \otimes \sigma_j^z$	$\mathbf{J}_{jk} \otimes \sigma_i^z$	$\mathbf{J}_{jk}$
$\mathbf{J}_{ij}$		0	0	$\mathbf{R}_{ij}$	$\mathbf{J}_{jk}$		$\mathbf{J}_{ik} \otimes \sigma_j^z$	$\mathbf{R}_{jk} \otimes \sigma_i^z$	$\mathbf{R}_{jk}$
$\mathbf{E}_{ij}$			0	0	$\mathbf{E}_{jk}$			0	0
$\mathbf{S}_{ij}$				0	$\mathbf{S}_{jk}$				0

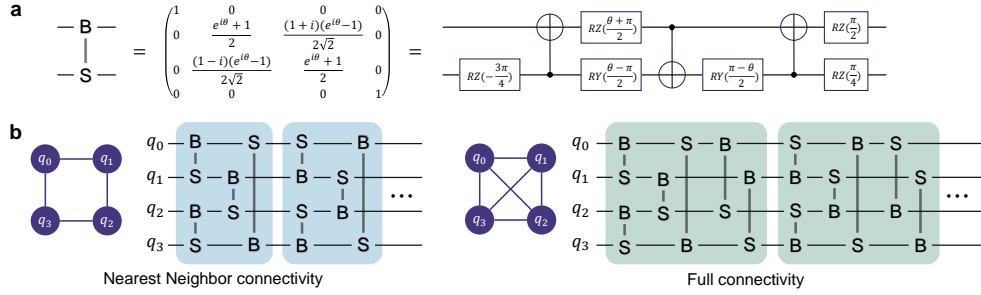
**Table 1** The commutation results of the four basis matrices (coefficients are omitted for simplicity). **Left:** commutator on same qubits; **Right:** commutator on different qubits.

	$\mathbf{R}_{ij} \otimes \sigma_k^z$	$\mathbf{J}_{ij} \otimes \sigma_k^z$	$\mathbf{E}_{ij} \otimes \sigma_k^z$	$\mathbf{S}_{ij} \otimes \sigma_k^z$		$\mathbf{R}_{ij} \otimes \sigma_k^z$	$\mathbf{J}_{ij} \otimes \sigma_k^z$	$\mathbf{E}_{ij} \otimes \sigma_k^z$	$\mathbf{S}_{ij} \otimes \sigma_k^z$
$\mathbf{R}_{jk}$	$\mathbf{J}_{ik}$	$\mathbf{R}_{ik}$	$\mathbf{J}_{jk}$	$\mathbf{J}_{jk} \otimes \sigma_i^z$	$\mathbf{R}_{jk} \otimes \sigma_i^z$	$\mathbf{J}_{ik} \otimes \sigma_j^z$	$\mathbf{R}_{ik} \otimes \sigma_j^z$	$\mathbf{J}_{jk} \otimes \sigma_i^z$	$\mathbf{J}_{jk}$
$\mathbf{J}_{jk}$		$\mathbf{J}_{ik}$	$\mathbf{R}_{jk}$	$\mathbf{R}_{jk} \otimes \sigma_i^z$	$\mathbf{J}_{jk} \otimes \sigma_i^z$		$\mathbf{J}_{ik} \otimes \sigma_j^z$	$\mathbf{R}_{jk} \otimes \sigma_i^z$	$\mathbf{R}_{jk}$
$\mathbf{E}_{jk}$			0	0	$\mathbf{S}_{jk} \otimes \sigma_i^z$			0	0
$\mathbf{S}_{jk}$				0	$\mathbf{E}_{jk} \otimes \sigma_i^z$				0

**Table 2** The commutation results of the four basis matrices tensored with Pauli-Z operator (coefficients are omitted for simplicity). **Left:** only one with Pauli-Z; **Right:** both tensored with Pauli-Z.

	I	II	III	IV	V	VI	VII	VIII
$\mathbf{R}_{ij}$	✓	✓					✓	✓
$\mathbf{J}_{ij}$	✓	✓						✓
$\mathbf{E}_{ij}$	✓	✓	✓	✓		✓	✓	✓
$\mathbf{S}_{ij}$	✓	✓	✓		✓	✓		✓
dim	$d_k^2$	$\begin{cases} d_k^2 - 1 & k < n/2 \\ \frac{d_k^2}{2} - 2 & k = n/2 \end{cases}$	$\frac{d_k(d_k-1)}{2}$	$\begin{cases} n & k=1 \\ \frac{n(n-1)}{2} & 1 < k < n/2 \\ \frac{(n-1)(n-2)}{2} & k = n/2 \end{cases}$	$n-1$	$\begin{cases} n & k=1 \\ \frac{n(n-1)}{2} & k > 1 \end{cases}$	$\begin{cases} d_k^2 & k < n/2 \\ d_k^2/2 - 1 & k = n/2 \end{cases}$	$d_k^2 - 1$

**Table 3** The classification of all fifteen configurations to eight distinct types of DLA and the corresponding DLA dimension, where the checkmark denotes the Hamiltonian of the HWP operator contains the corresponding basis matrix with a non-zero coefficient.



**Fig. 5 Implementation details of the BS ansatz.** **a** A possible circuit implementation of the proposed BS gate, where each BS gate requires three CNOT gates. **b** The implementation of the BS ansatz with both NN and FC connectivity. The physical qubit layout is depicted on the left. In the NN topology, implementing two-qubit gates between non-adjacent qubits, such as  $q_0$ - $q_2$  and  $q_1$ - $q_3$ , requires additional SWAP gates, which introduce overhead. To minimize the impact of qubit mapping and routing complexities, we omit these generators under NN connectivity. Consequently, FC connectivity allows for more generators per layer compared to NN connectivity. Additionally, certain HWP gates, such as the BS gate, are directional. To maximize expressivity, we alternate between layers containing BS gates and their reverse counterparts.

## Appendix A Preliminaries

### A.1 Quantum Computing and Quantum Machine Learning

In quantum computing, 'qubit' (abbreviation of 'quantum bit') is a key concept which is similar to a classical bit with a binary state. The two possible states for a qubit are the state  $|0\rangle$  and  $|1\rangle$ , which correspond to the state 0 and 1 for a classical bit respectively. We refer the readers to the textbook [22] for comprehension of quantum information and quantum computing. Here we give a brief introduction to the background.

A quantum state is commonly denoted in bracket notation. It is also common to form a linear combination of states, which we call a superposition:  $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ . Formally, a quantum system on  $n$  qubits is an  $n$ -fold tensor product Hilbert space  $\mathcal{H} = (\mathbb{C}^2)^{\otimes d}$  with dimension  $2^d$ . For any  $|\psi\rangle \in \mathcal{H}$ , the conjugate transpose  $\langle\psi| = |\psi\rangle^\dagger$ . The inner product  $\langle\psi|\psi\rangle = \|\psi\|_2^2$  denotes the square of the 2-norm of  $\psi$ . The outer product  $|\psi\rangle\langle\psi|$  is a rank 2 tensor. Computational basis states are given by  $|0\rangle = (1, 0)$ , and  $|1\rangle = (0, 1)$ . The composite basis states are defined by e.g.  $|01\rangle = |0\rangle \otimes |1\rangle = (0, 1, 0, 0)$ .

Analog to a classical computer, a quantum computer is built from a quantum circuit containing wires and elementary quantum gates to carry around and manipulate the quantum information. These gates can be parameterized quantum gates such as  $Rx(\theta)$ ,  $Ry(\theta)$ ,  $Rz(\theta)$  or basic gates as  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_z$ ,  $CNOT$ ,  $CZ$ . For an initial state  $|\psi_0\rangle$  and  $L$  layers of quantum circuit, the final state  $|\psi'\rangle$  can be denoted as

$$|\psi'\rangle = \prod_{l=1}^L \mathbf{U}_l |\psi_0\rangle. \quad (\text{A1})$$

Quantum computing is inherently susceptible to noise, which poses significant challenges for reliable computation. Extensive research has been devoted to quantum error mitigation [44, 45] and quantum error correction [46–49]. While these efforts are essential for universal quantum computing, some works have also focused on developing quantum ansätze that are self-protected or resilient to specific types of errors [50, 51]. Such approaches are particularly valuable for the near-term advancement of quantum processors. The HWP ansatz explored in this paper offers immunity to bit-flip errors, thereby significantly enhancing its practical utility by simplifying error mitigation and correction in near-term quantum devices.

### A.2 Hamming Distance and Hamming Weight

In this section, we recall the definition of Hamming distance and Hamming weight, which is constantly used in this paper. Considering two binary vectors  $\mathbf{a}$  and  $\mathbf{b}$  with  $\mathbf{a}, \mathbf{b} \in \{0, 1\}^N$ , where  $N$  is the dimension of the vectors. We can define the corresponding Hamming distance of these two vectors as follows:

---

**Algorithm 1** Computing the dimension of DLA

---

**Require:** Generator set  $\mathcal{G} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_P\}$ ,  $\mathbf{h}_i = \sigma(\mathbf{H}_i)$ .

```
let  $\mathbf{D} = \mathbf{h}_1$ ;
let  $r = \text{rank}(\mathbf{D})$ ;
for  $i = 2$  to  $P$  do
  if  $\text{rank}(\mathbf{D}, \mathbf{h}_i) > r$  then
     $\mathbf{D}.\text{append}(\mathbf{h}_i)$ ;
     $r = r + 1$ ;
  end if
end for
let  $r_{out} = 0$ ;
let  $r_n = \text{rank}(\mathbf{D})$ ;
while  $r_{out} \neq r_n$  and  $r \neq N^2$  do
  for  $l = r_{out} + 1$  to  $r_n$  do
    for  $j = 1$  to  $l - 1$  do
      let  $\mathbf{H}_{tmp} = [\sigma^{-1}(\mathbf{D}_{:,l}), \sigma^{-1}(\mathbf{D}_{:,j})]$ ,  $\mathbf{h}_{tmp} = \sigma(\mathbf{H}_{tmp})$ ;
      if  $\text{rank}(\mathbf{D}, \mathbf{h}_{tmp}) > r$  then
         $\mathbf{D}.\text{append}(\mathbf{h}_{tmp})$ ;
         $r = r + 1$ ;
      end if
    end for
  end for
  let  $r_{out} = r_n$ ;
  let  $r_n = \text{rank}(\mathbf{D})$ ;
end while
return  $r_{out}$ .
```

---

**Definition 2. Hamming distance:** The Hamming distance  $\mathcal{D}$  of two binary vectors  $\mathbf{a}$  and  $\mathbf{b}$  is:

$$\begin{aligned} \mathbf{c} &= \mathbf{a} \oplus \mathbf{b}, \\ \mathcal{D}(\mathbf{a}, \mathbf{b}) &= \sum_{i=1}^N \mathbf{c}_i, \end{aligned} \tag{A2}$$

where  $\oplus$  stands for exclusive OR operator.

With the definition of Hamming distance, we can further define the Hamming weight of a given binary vector  $\mathbf{a}$ .

**Definition 3. Hamming weight:** Let  $\mathbf{0} = 0^N$ . The Hamming weight  $HW$  of binary vector  $\mathbf{a}$  is:

$$HW(\mathbf{a}) = \mathcal{D}(\mathbf{a}, \mathbf{0}), \tag{A3}$$

### A.3 Detailed algorithm for computing the dimension of DLA

In this section, we provide an algorithm to calculate the DLA dimension. We define a transformation  $\sigma$  which satisfies  $\mathbf{h} = \sigma(\mathbf{H})$  ( $\mathbf{H} \in \mathbb{C}^{N \times N}$ ,  $\mathbf{h} \in \mathbb{C}^{N^2 \times 1}$ ), where  $\mathbf{h}$  is a column vector obtained by concatenating the columns of  $H$  vertically.  $\sigma^{-1}$  stands for



the inversed transformation which maps a column vector  $\in \mathbb{C}^{N^2 \times 1}$  back to a matrix  $\in \mathbb{C}^{N \times N}$ .  $\mathbf{D}_{:,j}$  denotes the  $j$ -th column of matrix  $\mathbf{D}$ , and  $\text{rank}(\mathbf{D}, \mathbf{h})$  is the rank of matrix  $\mathbf{D}$  appending column vector  $\mathbf{h}$  to the right.  $[\cdot, \cdot]$  denotes the commutator between two matrices with  $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$ . The pseudo code for the algorithm is shown in Alg. 1.

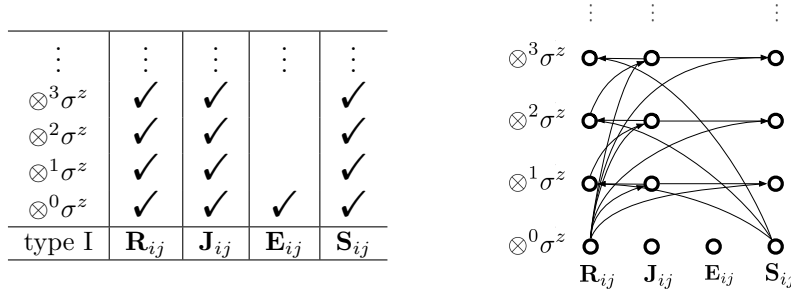
## Appendix B Detailed Proof for the Universality

### B.1 Proof of Theorem 1

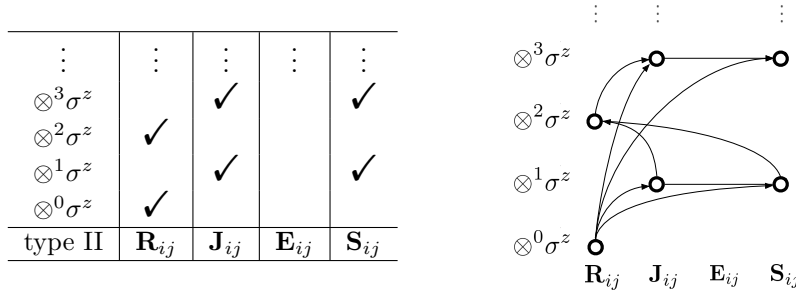
**Theorem.** For any  $n$  and  $k$ , a two-qubit HWP gate is universal with full connectivity (FC) if and only if the coefficients satisfy one of the following two conditions:

$$(1) e \neq 0, j \neq 0; \quad (2) e \neq 0, r \neq 0, s \neq 0. \quad (\text{B4})$$

*Proof.* We have derived the commutation results of the four basis matrices in the Method section 4. Now we provide the derivation of the eight distinct DLA types. The directed arrows in the figures indicate generative relations of the elements.

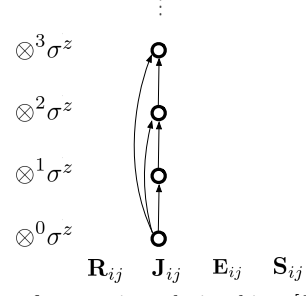


**Fig. B1** Generators in the DLA type I and one feasible set of generative relationships:  $[\mathbf{R}_{ij}, \mathbf{R}_{jk} \otimes^l \sigma_{m_l \neq i, j, k}^z] = i \mathbf{J}_{ki} \otimes \sigma_j^z \otimes^l \sigma_{m_l \neq i, j, k}^z = i \mathbf{J}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{J}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z, \mathbf{R}_{ki}] = 2i \mathbf{S}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{S}_{ki}, \mathbf{J}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z] = 2i \mathbf{R}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z$ .



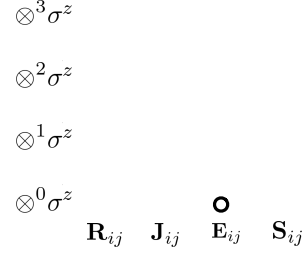
**Fig. B2** Generators in the DLA type II and a feasible set of generative relationships:  $[\mathbf{R}_{ij}, \mathbf{R}_{jk} \otimes^{2l} \sigma_{m_l \neq i, j, k}^z] = i \mathbf{J}_{ki} \otimes \sigma_j^z \otimes^{2l} \sigma_{m_l \neq i, j, k}^z = i \mathbf{J}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{J}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z, \mathbf{R}_{ki}] = 2i \mathbf{S}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{S}_{ki} \otimes \sigma_j^z, \mathbf{J}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z] = 2i \mathbf{R}_{ki} \otimes \sigma_j^z \otimes^{2l+1} \sigma_{m_l \neq i, j, k}^z = 2i \mathbf{R}_{ki} \otimes^{2l+2} \sigma_{m_l \neq i, j, k}^z$ .

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\otimes^3 \sigma^z$		✓		
$\otimes^2 \sigma^z$		✓		
$\otimes^1 \sigma^z$		✓		
$\otimes^0 \sigma^z$		✓		
type III	$\mathbf{R}_{ij}$	$\mathbf{J}_{ij}$	$\mathbf{E}_{ij}$	$\mathbf{S}_{ij}$



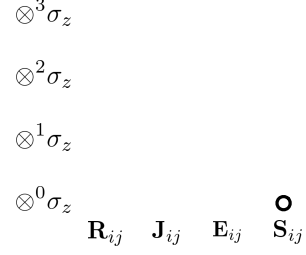
**Fig. B3** Generators in the DLA type III and a feasible set of generative relationships:  $[\mathbf{J}_{ij}, \mathbf{J}_{jk} \otimes^l \sigma_{m_l \neq i, j, k}^z] = i \mathbf{J}_{ik} \otimes \sigma_j^z \otimes^l \sigma_{m_l \neq i, j, k}^z = i \mathbf{J}_{ik} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z$ .

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\otimes^3 \sigma^z$				
$\otimes^2 \sigma^z$				
$\otimes^1 \sigma^z$				
$\otimes^0 \sigma^z$			✓	
type IV	$\mathbf{R}_{ij}$	$\mathbf{J}_{ij}$	$\mathbf{E}_{ij}$	$\mathbf{S}_{ij}$



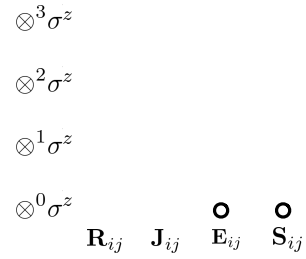
**Fig. B4** Generators in the DLA type IV .

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\otimes^3 \sigma^z$				
$\otimes^2 \sigma^z$				
$\otimes^1 \sigma^z$				
$\otimes^0 \sigma^z$				✓
type V	$\mathbf{R}_{ij}$	$\mathbf{J}_{ij}$	$\mathbf{E}_{ij}$	$\mathbf{S}_{ij}$

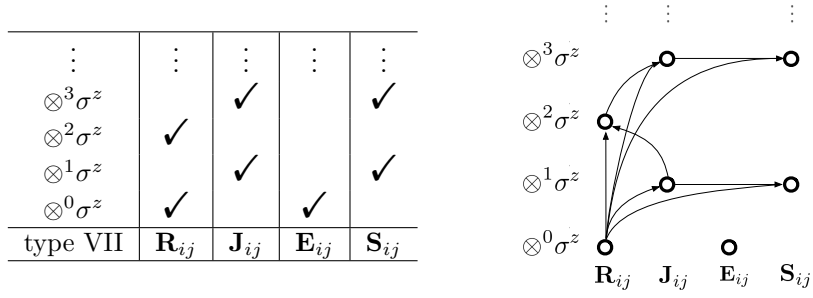


**Fig. B5** Generators in the DLA type V .

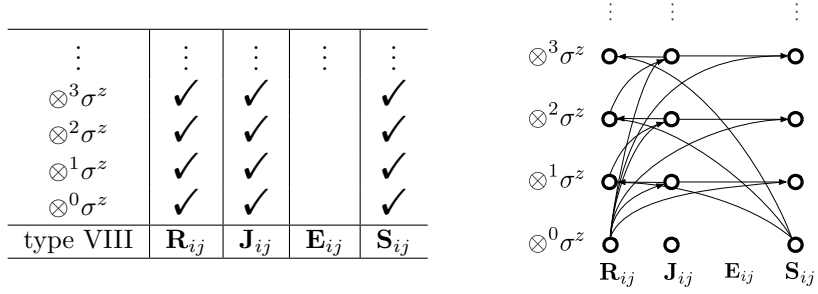
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\otimes^3 \sigma^z$				
$\otimes^2 \sigma^z$				
$\otimes^1 \sigma^z$				
$\otimes^0 \sigma^z$			✓	✓
type VI	$\mathbf{R}_{ij}$	$\mathbf{J}_{ij}$	$\mathbf{E}_{ij}$	$\mathbf{S}_{ij}$



**Fig. B6** Generators in the DLA type VI.



**Fig. B7** Generators in the DLA type VII and a feasible set of generative relationships:  $[\mathbf{R}_{ij}, \mathbf{R}_{jk} \otimes^{2l} \sigma_{m_l \neq i, j, k}^z] = i \mathbf{J}_{ki} \otimes \sigma_j^z \otimes^{2l} \sigma_{m_l \neq i, j, k}^z = i \mathbf{J}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{J}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z, \mathbf{R}_{ik}] = 2i \mathbf{S}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{S}_{ki} \otimes \sigma_j^z, \mathbf{J}_{ki} \otimes^{2l+1} \sigma_{m_l \neq i, k}^z] = 2i \mathbf{R}_{ki} \otimes \sigma_j^z \otimes^{2l+1} \sigma_{m_l \neq i, j, k}^z = 2i \mathbf{R}_{ki} \otimes^{2l+2} \sigma_{m_l \neq i, j, k}^z$ .



**Fig. B8** Generators in the DLA type VIII and a feasible set of generative relationships:  $[\mathbf{R}_{ij}, \mathbf{R}_{jk} \otimes^l \sigma_{m_l \neq i, j, k}^z] = i \mathbf{J}_{ki} \otimes \sigma_j^z \otimes^l \sigma_{m_l \neq i, j, k}^z = i \mathbf{J}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{J}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z, \mathbf{R}_{ik}] = 2i \mathbf{S}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z$ ,  $[\mathbf{S}_{ki}, \mathbf{J}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z] = 2i \mathbf{R}_{ki} \otimes^{l+1} \sigma_{m_l \neq i, k}^z$ .

The dimension of these eight DLA types is demonstrated in Table B.1. Notice that  $k \leq n/2$  in the Table since the dimension of DLA  $\{n, k\}$  is the same as  $\{n, n-k\}$ .

	I	II	III	IV	V	VI	VII	VIII
$\mathbf{R}_{ij}$	✓	✓					✓	✓
$\mathbf{J}_{ij}$	✓		✓					✓
$\mathbf{E}_{ij}$	✓			✓		✓	✓	
$\mathbf{S}_{ij}$	✓				✓	✓		✓
dim	$d_k^2$	$\begin{cases} d_k^2 - 1 & \text{if } k < n/2 \\ \frac{d_k^2}{2} - 2 & \text{if } k = n/2 \end{cases}$	$\frac{d_k(d_k-1)}{2}$	$\begin{cases} n & \text{if } k = 1 \\ \frac{n(n-1)}{2} & \text{if } 1 < k < n/2 \\ \frac{(n-1)(n-2)}{2} & \text{if } k = n/2 \end{cases}$	$n-1$	$\begin{cases} n(n-1) & \text{if } k = 1 \\ \frac{n(n-1)}{2} & \text{if } k > 1 \end{cases}$	$\begin{cases} d_k^2 & \text{if } k < n/2 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases}$	$d_k^2 - 1$

**Table B1** Eight distinct DLA types and the corresponding dimension. The checkmark denotes whether the basis matrix is in the generator set  $\mathcal{G}$ .

We then derive all fifteen configurations (excluding the all-zero one) and illustrate how they can be regulated to these eight DLA types. Recall the definition of the

Hamiltonian of the HWP gate acting on the  $i$ -th and  $j$ -th qubits:

$$\mathbf{H}_{HWij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & a & b & 0 \\ 0 & \bar{b} & c & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (\text{B5})$$

and it can be decomposed as a linear combination of the four basis matrices:

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + j\mathbf{J}_{ij} + e\mathbf{E}_{ij} + s\mathbf{S}_{ij}, r, j, e, s \in \mathbb{R}, \quad (\text{B6})$$

where  $r = \text{Re } b, j = \text{Im } b, e = \frac{a+c}{2}, s = \frac{a-c}{2}$ . We notice that

$$\mathbf{H}_{HWji} = r\mathbf{R}_{ij} - j\mathbf{J}_{ij} + e\mathbf{E}_{ij} - s\mathbf{S}_{ij} \quad (\text{B7})$$

So that

$$[\mathbf{H}_{HWij}, \mathbf{H}_{HWji}] = 4ir(\mathbf{I}\mathbf{S}_{ij} - s\mathbf{J}_{ij}) \quad (\text{B8})$$

$$\mathbf{H}_{HWij} + \mathbf{H}_{HWji} = 2e\mathbf{E}_{ij} + 2r\mathbf{R}_{ij} \quad (\text{B9})$$

$$\mathbf{H}_{HWij} - \mathbf{H}_{HWji} = 2j\mathbf{J}_{ij} + 2s\mathbf{S}_{ij} \quad (\text{B10})$$

1.  $r, j, e, s \neq 0$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + j\mathbf{J}_{ij} + e\mathbf{E}_{ij} + s\mathbf{S}_{ij} \quad (\text{B11})$$

The determinant of the coefficients of  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$  in Eq.B8 and Eq.B10 is

$$\begin{vmatrix} -4irs & 4irj \\ 2j & 2s \end{vmatrix} = -8ir(j^2 + s^2) \neq 0 \quad (\text{B12})$$

Then the set of generators contains  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$ . With  $[\mathbf{J}_{ij}, \mathbf{E}_{ij}] = -2i\mathbf{R}_{ij}$  and Eq. B11,  $\mathbf{R}_{ij}$  and  $\mathbf{E}_{ij}$  are contained too. So DLA of this case can be specified to type I and

$$\dim(\text{DLA}) = d_k^2 \quad (\text{B13})$$

2.  $r, j, e \neq 0, s = 0$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + j\mathbf{J}_{ij} + e\mathbf{E}_{ij} \quad (\text{B14})$$

Same as  $r, j, e, s \neq 0$  1, DLA of this case can be specified to type I and

$$\dim(\text{DLA}) = d_k^2 \quad (\text{B15})$$

3.  $r, j, s \neq 0, e = 0$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + j\mathbf{J}_{ij} + s\mathbf{S}_{ij} \quad (\text{B16})$$

Similar to  $r, j, e, s \neq 0$  1, the set of generators contains  $\mathbf{R}_{ij}$ ,  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$  but without  $\mathbf{E}_{ij}$ , so DLA of this case can be specified to type VIII and

$$\dim(\text{DLA}) = d_k^2 - 1 \quad (\text{B17})$$

4.  $r, j \neq 0, e = s = 0$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + j\mathbf{J}_{ij} \quad (\text{B18})$$

Same as  $r, j, s \neq 0, e = 0$ , DLA of this case can be specified to type VIII and

$$\dim(\text{DLA}) = d_k^2 - 1 \quad (\text{B19})$$

5.  $r, e, s \neq 0, j = 0$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + e\mathbf{E}_{ij} + s\mathbf{S}_{ij} \quad (\text{B20})$$

Same as  $r, j, e, s \neq 0$  1, DLA of this case can be specified to type I and

$$\dim(\text{DLA}) = d_k^2 \quad (\text{B21})$$

6.  $r, e \neq 0, j = s = 0$

$$\mathbf{H}_{HWij} = e\mathbf{E}_{ij} + r\mathbf{R}_{ij} \quad (\text{B22})$$

$$[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}] = i r (e\mathbf{J}_{ij} \otimes \sigma_k^z + e\mathbf{J}_{jk} \otimes \sigma_i^z + r\mathbf{J}_{ki} \otimes \sigma_j^z) \quad (\text{B23})$$

(1)  $r \neq e, -2e$  We can alternate the cyclic subscripts to get

$$[\mathbf{H}_{HWjk}, \mathbf{H}_{HWki}] = i r (r\mathbf{J}_{ij} \otimes \sigma_k^z + e\mathbf{J}_{jk} \otimes \sigma_i^z + e\mathbf{J}_{ki} \otimes \sigma_j^z) \quad (\text{B24})$$

and

$$[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}] = i r (e\mathbf{J}_{ij} \otimes \sigma_k^z + r\mathbf{J}_{jk} \otimes \sigma_i^z + e\mathbf{J}_{ki} \otimes \sigma_j^z) \quad (\text{B25})$$

The determinant of the coefficients of  $\mathbf{J}_{ij}$ ,  $\mathbf{J}_{jk}$  and  $\mathbf{J}_{ki}$  in the formulas is

$$\begin{vmatrix} ire & ire & ir^2 \\ ir^2 & ire & ire \\ ire & ir^2 & ire \end{vmatrix} = i(e-r)^2 r^3 (2e+r) \neq 0 \quad (\text{B26})$$

Then the set of generators contains  $\mathbf{J}_{ij} \otimes \sigma_k^z$ . With

$$[\mathbf{J}_{ij} \otimes \sigma_k^z, \mathbf{H}_{HWij}] = r\mathbf{S}_{ij} \otimes \sigma_k^z \quad (\text{B27})$$

and

$$[\mathbf{S}_{ij} \otimes \sigma_k^z, \mathbf{J}_{ij} \otimes \sigma_k^z] = \mathbf{R}_{ij} \quad (\text{B28})$$

Then the set of generators contains  $\mathbf{R}_{ij}$  and  $\mathbf{E}_{ij}$  ( $\mathbf{H}_{HWij} = e\mathbf{E}_{ij} + r\mathbf{R}_{ij}$ ) but without  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type VII and

$$\dim(\text{DLA}) = \begin{cases} d_k^2 & \text{if } k < n/2 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases} \quad (\text{B29})$$

(2)  $r = e$

$$[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}] = -r^2 (\mathbf{J}_{ij} \otimes \sigma_k^z + \mathbf{J}_{ki} \otimes \sigma_j^z + \mathbf{J}_{jk} \otimes \sigma_i^z) \quad (\text{B30})$$

$$[\mathbf{H}_{HWik}, [\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}]] = 2r^3 \mathbf{S}_{ki} \otimes \sigma_j^z \quad (\text{B31})$$

$$[\mathbf{S}_{ij} \otimes \sigma_k^z, \mathbf{H}_{HWij}] = \mathbf{J}_{ij} \otimes \sigma_k^z \quad (\text{B32})$$

$$[\mathbf{S}_{ij} \otimes \sigma_k^z, \mathbf{J}_{ij} \otimes \sigma_k^z] = 2\mathbf{R}_{ij} \quad (\text{B33})$$

Then the set of generators contains  $\mathbf{R}_{ij}$  and  $\mathbf{E}_{ij}$  ( $H_{HWij} = e\mathbf{E}_{ij} + r\mathbf{R}_{ij}$ ) but without  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type VII and

$$\dim(\text{DLA}) = \begin{cases} d_k^2 & \text{if } k < n/2 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases} \quad (\text{B34})$$

$$(3)r = -2e$$

$$\mathbf{H}_{HWij} = e\mathbf{E}_{ij} - 2e\mathbf{R}_{ij} \quad (\text{B35})$$

$$[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}] = 2e^2 \mathbf{J}_{ij} \otimes \sigma_k^z + 2e^2 \mathbf{J}_{jk} \otimes \sigma_i^z - 4e^2 \mathbf{J}_{ki} \otimes \sigma_j^z \quad (\text{B36})$$

$$[\mathbf{H}_{HWki}, [\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}]] = 2e^2 (-8\mathbf{S}_{ki} \otimes \sigma_j^z + 3\mathbf{R}_{jk} - 3\mathbf{R}_{ij}) \quad (\text{B37})$$

$$[\mathbf{H}_{HWki}, [\mathbf{H}_{HWjk}, \mathbf{H}_{HWki}]] = 2e^3 (4\mathbf{S}_{ki} \otimes \sigma_j^z - 4\mathbf{R}_{ij} + 5\mathbf{R}_{jk}) \quad (\text{B38})$$

$$2[\mathbf{H}_{HWki}, [\mathbf{H}_{HWjk}, \mathbf{H}_{HWki}]] + [\mathbf{H}_{HWki}, [\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}]] = 2e^3 (-11\mathbf{R}_{ij} + 13\mathbf{R}_{jk}) \quad (\text{B39})$$

We can alternate the cyclic subscripts to get

$$2[\mathbf{H}_{HWij}, [\mathbf{H}_{HWki}, \mathbf{H}_{HWij}]] + [\mathbf{H}_{HWij}, [\mathbf{H}_{HWjk}, \mathbf{H}_{HWki}]] = 2e^3 (-11\mathbf{R}_{jk} + 13\mathbf{R}_{ki}) \quad (\text{B40})$$

$$2[\mathbf{H}_{HWjk}, [\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}]] + [\mathbf{H}_{HWjk}, [\mathbf{H}_{HWki}, \mathbf{H}_{HWij}]] = 2e^3 (-11\mathbf{R}_{ki} + 13\mathbf{R}_{ij}) \quad (\text{B41})$$

The determinant of the coefficients of  $\mathbf{R}_{ij}$ ,  $\mathbf{R}_{jk}$  and  $\mathbf{R}_{ki}$  in Eq. B39, Eq. B40 and Eq. B41 is

$$\begin{vmatrix} -11 & 13 & 0 \\ 0 & -11 & 13 \\ 13 & 0 & -11 \end{vmatrix} = 866 \neq 0 \quad (\text{B42})$$

Then the set of generators contains  $\mathbf{R}_{ij}$  and  $\mathbf{E}_{ij}$  ( $H_{HWij} = e\mathbf{E}_{ij} + r\mathbf{R}_{ij}$ ) but without  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type VII and

$$\dim(\text{DLA}) = \begin{cases} d_k^2 & \text{if } k < n/2 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases} \quad (\text{B43})$$

$$7. \ r, s \neq 0, j = e = 0$$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} + s\mathbf{S}_{ij} \quad (\text{B44})$$

Similar to  $r, j, e, s \neq 0$  1, the set of generators contains  $\mathbf{R}_{ij}$ ,  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$  but without  $\mathbf{E}_{ij}$ , so DLA of this case can be specified to type VIII and

$$\dim(\text{DLA}) = d_k^2 - 1 \quad (\text{B45})$$

$$8. \ r \neq 0, j = e = k = 0$$

$$\mathbf{H}_{HWij} = r\mathbf{R}_{ij} \quad (\text{B46})$$

Then the set of generators only contains  $\mathbf{R}_{ij}$  but without  $\mathbf{J}_{ij}$ ,  $\mathbf{E}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type II and

$$\dim(\text{DLA}) = \begin{cases} d_k^2 - 1 & \text{if } k < n/2 \\ \frac{d_k^2 - 1}{2} & \text{if } k = n/2 \end{cases} \quad (\text{B47})$$

9.  $j, e, s \neq 0, r = 0$

$$\mathbf{H}_{HWij} = j\mathbf{J}_{ij} + e\mathbf{E}_{ij} + s\mathbf{S}_{ij} \quad (\text{B48})$$

$$\mathbf{H}_{HWij} + \mathbf{H}_{HWji} = 2e\mathbf{E}_{ij} \quad (\text{B49})$$

Then the set of generators contains  $\mathbf{E}_{ij}$ , and

$$[\mathbf{H}_{HWij} + \mathbf{H}_{HWji}, \mathbf{H}_{HWjk} - \mathbf{H}_{HWkj}] = -4iej\mathbf{R}_{jk} \otimes \sigma_i^z \quad (\text{B50})$$

$$[\mathbf{E}_{jk}, \mathbf{R}_{ij} \otimes \sigma_k^z] = -i\mathbf{J}_{ij} \quad (\text{B51})$$

$$[\mathbf{R}_{ij} \otimes \sigma_k^z, \mathbf{J}_{jk}] = -i\mathbf{R}_{ki} \quad (\text{B52})$$

$$[\mathbf{R}_{ij}, \mathbf{J}_{ij}] = -2i\mathbf{S}_{ij} \quad (\text{B53})$$

Then the set of generators contains  $\mathbf{R}_{ij}$ ,  $\mathbf{J}_{ij}$ ,  $\mathbf{E}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type I and

$$\dim(\text{DLA}) = d_k^2 \quad (\text{B54})$$

10.  $j, e \neq 0, r = s = 0$

$$\mathbf{H}_{HWij} = j\mathbf{J}_{ij} + e\mathbf{E}_{ij} \quad (\text{B55})$$

Same as  $r \neq 0, j = e = k = 0$  8, DLA of this case can be specified to type I and

$$\dim(\text{DLA}) = d_k^2 \quad (\text{B56})$$

11.  $j, s \neq 0, r = e = 0$

$$\mathbf{H}_{HWij} = s\mathbf{S}_{ij} + j\mathbf{J}_{ij} \quad (\text{B57})$$

$$[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}] = ij(s\mathbf{R}_{ij} - j\mathbf{J}_{ki} \otimes \sigma_j^z - s\mathbf{R}_{jk}) \quad (\text{B58})$$

$$[[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}], \mathbf{H}_{HWij}] = -(j^3 + js^2)\mathbf{J}_{jk} + 2j^2s\mathbf{S}_{ij} - 2js^2\mathbf{J}_{ij} \quad (\text{B59})$$

$$[[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}], \mathbf{H}_{HWij}] - 2j^2\mathbf{H}_{HWij} = -(j^3 + js^2)\mathbf{J}_{jk} - (2j^3 + 2js^2)\mathbf{J}_{ij} \quad (\text{B60})$$

$$[[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}], \mathbf{H}_{HWjk}] = -2j^2s\mathbf{S}_{jk} + 2js^2\mathbf{J}_{jk} + (j^3 + js^2)\mathbf{J}_{ij} \quad (\text{B61})$$

$$[[\mathbf{H}_{HWij}, \mathbf{H}_{HWjk}], \mathbf{H}_{HWjk}] + 2j^2\mathbf{H}_{HWjk} = (2js^2 + 2j^3)\mathbf{J}_{jk} + (j^3 + js^2)\mathbf{J}_{ij} \quad (\text{B62})$$

The determinant of the coefficients of  $\mathbf{J}_{ij}$  and  $\mathbf{J}_{jk}$  in Eq. B60 and Eq. B62 is

$$\begin{vmatrix} -(j^3 + js^2) & -(2j^3 + 2js^2) \\ 2js^2 + 2j^3 & j^3 + js^2 \end{vmatrix} = -j^2(j^2 + s^2)^2 + 4j^2(j^2 + s^2)^2 = 3j^2(j^2 + s^2)^2 > 0 \quad (\text{B63})$$

Then the set of generators contains  $\mathbf{J}_{ij}$ . With  $\mathbf{H}_{HWij} = s\mathbf{S}_{ij} + j\mathbf{J}_{ij}$  and  $[\mathbf{S}_{ij}, \mathbf{J}_{ij}] = 2i\mathbf{R}_{ij}$ , the DLA also contains  $\mathbf{R}_{ij}$  and  $\mathbf{S}_{ij}$  but without  $\mathbf{E}_{ij}$ , so DLA of this case can be specified to type VIII and

$$\dim(\text{DLA}) = d_k^2 - 1 \quad (\text{B64})$$

12.  $j \neq 0, r = e = s = 0$

$$\mathbf{H}_{HWij} = j\mathbf{J}_{ij} \quad (\text{B65})$$

Then the set of generators only contains  $\mathbf{J}_{ij}$  but without  $\mathbf{R}_{ij}$ ,  $\mathbf{E}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type III and

$$\dim(\text{DLA}) = \frac{d_k(d_k - 1)}{2} \quad (\text{B66})$$

13.  $e, s \neq 0, r = j = 0$

$$\mathbf{H}_{HWij} = e\mathbf{E}_{ij} + s\mathbf{S}_{ij} \quad (\text{B67})$$

$$\mathbf{H}_{HWij} + \mathbf{H}_{HWji} = 2e\mathbf{E}_{ij} \quad (\text{B68})$$

$$\mathbf{H}_{HWij} - \mathbf{H}_{HWji} = 2s\mathbf{S}_{ij} \quad (\text{B69})$$

Then the set of generators contains  $\mathbf{E}_{ij}$ ,  $\mathbf{S}_{ij}$ , but without  $\mathbf{R}_{ij}$  and  $\mathbf{J}_{ij}$ , so DLA of this case can be specified to type VI and

$$\dim(\text{DLA}) = \begin{cases} n & \text{if } k = 1 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases} \quad (\text{B70})$$

14.  $e \neq 0, r = j = s = 0$

$$\mathbf{H}_{HWij} = e\mathbf{E}_{ij} \quad (\text{B71})$$

Then the set of generators only contains  $\mathbf{E}_{ij}$  but without  $\mathbf{R}_{ij}$ ,  $\mathbf{J}_{ij}$  and  $\mathbf{S}_{ij}$ , so DLA of this case can be specified to type IV and

$$\dim(\text{DLA}) = \begin{cases} n & \text{if } k = 1 \\ \frac{n(n-1)}{2} & \text{if } k > 1 \\ \frac{(n-1)(n-2)}{2} & \text{if } k = n/2 \end{cases} \quad (\text{B72})$$

15.  $s \neq 0, r = j = e = 0$

$$\mathbf{H}_{HWij} = s\mathbf{S}_{ij} \quad (\text{B73})$$

Then the set of generators only contains  $\mathbf{S}_{ij}$  but without  $\mathbf{R}_{ij}$ ,  $\mathbf{J}_{ij}$  and  $\mathbf{E}_{ij}$ , so DLA of this case can be specified to type V and

$$\dim(\text{DLA}) = n - 1 \quad (\text{B74})$$

To sum up, we can conclude the classification of the fifteen configurations to the eight DLA types as the following Table B.1. Therefore for any  $n$  and  $k$ , a two-qubit HWP gate is universal with full connectivity if and only if the coefficients satisfy one of the following two conditions:

$$(1) \ e \neq 0, j \neq 0; \quad (2) \ e \neq 0, r \neq 0, s \neq 0. \quad (\text{B75})$$

□





4.  $j, e, s \neq 0, r = 0$

$$\mathbf{H}_{HWij} = j\mathbf{J}_{ij} + e\mathbf{E}_{ij} + s\mathbf{S}_{ij} \quad (\text{B84})$$

$$\mathbf{H}_{HWi,i+1} + \mathbf{H}_{HWi+1,i} = 2e\mathbf{E}_{i,i+1} \quad (\text{B85})$$

$$\mathbf{H}_{HWi,i+1} - \mathbf{H}_{HWi+1,i} = 2(s\mathbf{S}_{i,i+1} + j\mathbf{J}_{i,i+1}) \quad (\text{B86})$$

$$[\mathbf{H}_{HWi,i+1} + \mathbf{H}_{HWi+1,i}, \mathbf{H}_{HWi+1,i+2} - \mathbf{H}_{HWi+2,i+1}] = -4iejE_{i+1,i+2} \otimes \sigma_{zi} \quad (\text{B87})$$

$$[\mathbf{E}_{i+1,i+2}, \mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z] = \mathbf{J}_{i,i+1} \quad (\text{B88})$$

Then the set of generators contains all  $\mathbf{J}_{i,i+1}$ ,  $\mathbf{E}_{i,i+1}$ ,  $\mathbf{S}_{i,i+1}$  and  $\mathbf{R}_{i,i+1}$  ( $[\mathbf{J}_{i,i+1}, \mathbf{S}_{i,i+1}] = -2i\mathbf{R}_{i,i+1}$ ). According to the Lemma 2 DLA of this case can be specified to type I and

$$\dim(\text{DLA}) = d_k^2 \quad (\text{B89})$$

5.  $j, e \neq 0, r = s = 0$

Similar to 4, the set of generators contains  $\mathbf{J}_{i,i+1}$  and  $\mathbf{E}_{i,i+1}$ , then

$$[\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z, \mathbf{J}_{i+1,i+2}] = -\mathbf{R}_{i+2,i} \quad (\text{B90})$$

Then it can generate all  $\mathbf{R}_{i,i+2}$ , and

$$[\mathbf{J}_{i,i+1}, \mathbf{E}_{i+1,i+2}] = -i\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z \quad (\text{B91})$$

$$[\mathbf{R}_{i,i+2}, \mathbf{E}_{i+2,i+3}] = i\mathbf{J}_{i,i+2} \otimes \sigma_{i+3}^z \quad (\text{B92})$$

$$[\mathbf{R}_{i,i+1} \otimes \sigma_{i+2}^z, \mathbf{R}_{i,i+2}] = i\mathbf{J}_{i,i+1} \quad (\text{B93})$$

The set of generators contains  $\mathbf{R}_{i,i+2}$  but can't generate new elements like  $\mathbf{R}_{i,i+1}$  and  $\mathbf{S}_{i,i+1}$ , so it's specified to a different type of DLA with

$$\dim(\text{DLA}) = \begin{cases} d_k^2 & \text{if } k < n/2 \\ d_k^2/2 - 1 & \text{if } k = n/2 \end{cases} \quad (\text{B94})$$

Therefore for any  $n$  and  $k$ , a two-qubit HWP gate is universal with nearest neighbor connectivity if and only if the coefficients satisfy the following conditions:

$$(1) \ e \neq 0, j \neq 0, r \neq 0; \quad (2) \ e \neq 0, j \neq 0, s \neq 0; \quad (3) \ e \neq 0, r \neq 0, s \neq 0. \quad (\text{B95})$$

□

## Appendix C Detailed Proof for the Trainability

**Theorem 7.** Consider an  $n$ -qubit quantum circuit operating in the subspace with Hamming weight  $k$ . The variance of the cost function partial derivative is  $\text{Var}_\theta[\partial_l C] \approx \frac{16k^2(n-k)^2}{n^4 d_k}$ .

*Proof.* Consider the partial derivative of the cost function  $C$  with respect to the parameters  $\theta$ . For some parameter  $\theta_l$  in the  $l$ -th HWP gate  $\mathbf{H}_l$ , we have:

$$\partial_l C(\theta) = \partial_l \left( \text{Tr} \left( \mathbf{U}(\theta) \rho \mathbf{U}(\theta)^\dagger O \right) \right), \quad (\text{C96})$$

where  $\rho$  is the input state,  $O$  is the observable to measure. We split the whole ansatz to two parts with  $\mathbf{U}_-$  denotes the unitary matrix of the circuit before  $\mathbf{H}_l$  and  $\mathbf{U}_+$  denotes the unitary matrix after gate  $l$ .

$$\begin{aligned} \partial_l C(\theta) &= \partial_l \left( \text{Tr} \left( \mathbf{U}_- \rho \mathbf{U}_+^\dagger O \right) \right) \\ &= i \text{Tr} \left( \mathbf{U}_- \rho \mathbf{U}_+^\dagger [\mathbf{H}_l, O] \right), \end{aligned} \quad (\text{C97})$$

where  $[\cdot, \cdot]$  denotes the commutator of two matrices. The variance of the partial derivative is thus given by

$$\begin{aligned} \text{Var}_\theta[\partial_l C] &= \int_{\mathbf{U}_+} d\mathbf{U}_+ \int_{\mathbf{U}_-} d\mathbf{U}_- (\partial_l C(\theta))^2 \\ &= \int_{\mathbf{U}_+} d\mathbf{U}_+ \int_{\mathbf{U}_-} d\mathbf{U}_- \left( i \text{Tr} \left( \mathbf{U}_- \rho \mathbf{U}_+^\dagger [\mathbf{H}_l, O] \right) \right)^2 \\ &= - \int_{\mathbf{U}_+} d\mathbf{U}_+ \left( \frac{\text{Tr}(\rho^2) \text{Tr}([H_l, O]^2)}{d_k^2 - 1} - \frac{\text{Tr}^2(\rho) \text{Tr}([H_l, O]^2)}{d_k(d_k^2 - 1)} \right) \quad (\text{C98}) \\ &= - \int_{\mathbf{U}_+} d\mathbf{U}_+ \left( \text{Tr}([\mathbf{H}_l, O]^2) \frac{d_k \times \text{Tr}(\rho^2) - \text{Tr}^2(\rho)}{d_k(d_k^2 - 1)} \right) \\ &= - \frac{d_k \times \text{Tr}(\rho^2) - \text{Tr}^2(\rho)}{d_k(d_k^2 - 1)} \int_{\mathbf{U}_+} d\mathbf{U}_+ \text{Tr}([\mathbf{H}_l, O]^2). \end{aligned}$$

As long as the initial state is the  $d_k$  subspace, we have  $\text{Tr}(\rho) = 1$  and  $\text{Tr}(\rho^2) = 1$ .

$$\begin{aligned} \text{Var}_\theta[\partial_l C] &= - \frac{1}{d_k(d_k + 1)} \int_{\mathbf{U}_+} d\mathbf{U}_+ \text{Tr}([H_l, O]^2) \\ &= - \frac{2}{d_k(d_k + 1)} \left( \text{Tr}(\mathbf{H}_l O_+ \mathbf{H}_l O_+) - \text{Tr}(\mathbf{H}_l \mathbf{H}_l O_+ O_+) \right) \\ &= - \frac{2}{d_k(d_k + 1)} \left( \frac{\text{Tr}(\mathbf{H}_l^2) \text{Tr}(O)}{d_k^2 - 1} - \frac{\text{Tr}(\mathbf{H}_l^2) \text{Tr}(O^2)}{d_k(d_k^2 - 1)} - \frac{\text{Tr}(\mathbf{H}_l^2) \text{Tr}(O^2)}{d_k} \right) \\ &= - \frac{2 \text{Tr}(\mathbf{H}_l^2)}{d_k(d_k + 1)} \left( \frac{\text{Tr}^2(O) - d_k \text{Tr}(O^2)}{d_k^2 - 1} \right), \end{aligned} \quad (\text{C99})$$

where  $\text{Tr}(\mathbf{H}_l^2) = 2 \binom{n-2}{k-1} = \frac{2k(n-k)}{n(n-1)} d_k$ . Without loss of generality, we set the observable  $O$  as  $Z_0$  since other observables will also hold with the same magnitude. Thus,  $\text{Tr}(O) =$

$\frac{d_k(n-2k)}{n}$  and  $\text{Tr}(O^2) = d_k$ . Substitute this back into Equation C99 and we get

$$\begin{aligned}
\text{Var}_\theta[\partial_l C] &= -\frac{2}{d_k(d_k+1)} \times \frac{2k(n-k)d_k}{n(n-1)} \times \left( \frac{\frac{d_k^2(n-2k)^2}{n^2} - d_k^2}{d_k^2 - 1} \right) \\
&= \frac{4k(n-k)}{(d_k+1)n(n-1)} \times \frac{d_k^2(n^2 - (n-2k)^2)}{(d_k^2 - 1)n^2} \\
&= \frac{4k(n-k)}{(d_k+1)n(n-1)} \times \frac{d_k^2(4nk - 4k^2)}{(d_k^2 - 1)n^2} \\
&= \frac{16k^2(n-k)^2 d_k^2}{(d_k+1)n^3(n-1)(d_k^2 - 1)} \approx \frac{16k^2(n-k)^2}{n^4 d_k}.
\end{aligned} \tag{C100}$$

□

We can further analyze that if the  $k$  is only 1, then  $\text{Var}_\theta[\partial_l C] \approx \frac{16}{n^3}$ . If the  $k = \frac{n}{2}$  on the other hand,  $\text{Var}_\theta[\partial_l C] \approx \left(\frac{n}{n/2}\right)^{-1}$ , which is approximate to exponentially small. This result is consistent with the conjecture that the trainability of the circuit is closely related to  $d_k$ , and smaller  $d_k$  will lead to better trainability.

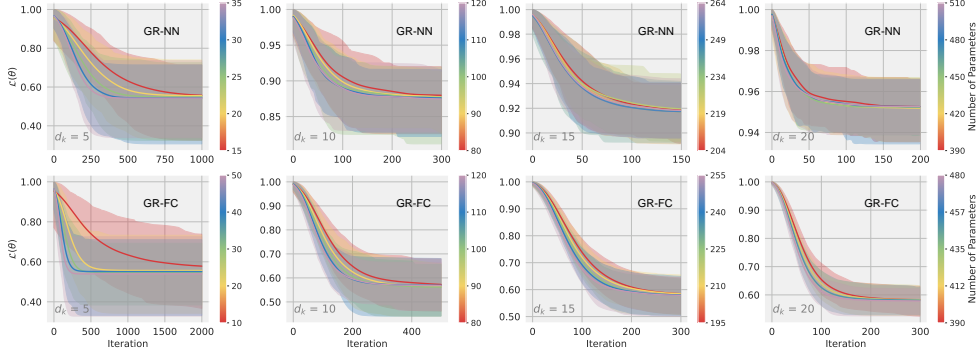
## Appendix D Implementation Details and additional Numerical Results

### D.1 Unitary Approximation

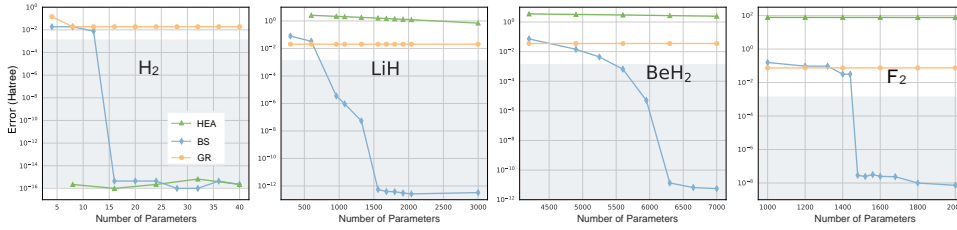
We further provide the training curves of the Givens Rotations, including all four cases ( $d_k = \{\binom{5}{1}, \binom{5}{2}, \binom{6}{2}, \binom{6}{3}\} = \{5, 10, 15, 20\}$ ) with both NN and FC connectivity. As illustrated in Figure D9, the ability to approximate an arbitrary unitary matrix is restricted by the DLA dimension and increasing the number of parameters as the BS ansatz does not improve the performance. The average value and the minimum value both converge to similar results over a different number of parameters indicating we could already be facing overparameterization for the Givens Rotations.

### D.2 Molecular Electronic Structures

For simulating molecular electronic structures, we first present the statistics of the molecules examined in our experiments (as shown in Table D3). The molecular Hamiltonians were obtained from OpenFermion without freezing any orbitals, a common technique used to reduce the problem size. All the numerical results in the main part (including the heat map in Fig. 2C) are a minimum of 10 random seeds since the results for UCCSD are sensitive to the initializing of the parameters. To further illustrate the effectiveness of the proposed BS ansatz, we utilized Givens Rotations, UCCSD, and HEA as the baseline methods. The implementation of the GR and HEA ansätze is illustrated in Figure D11, with the GR gate specifically depicted in Figure 4. The UCCSD ansatz is constructed using the MindQuantum Python package [52]. For the proposed BS ansatz, as well as all baseline methods, the Hartree-Fock (HF) state



**Fig. D9 Additional Results for Unitary Approximation.** The training curves for Givens Rotations. The solid line indicates the average loss, and the shadow area indicates the range of the results of 100 random unitary matrices.



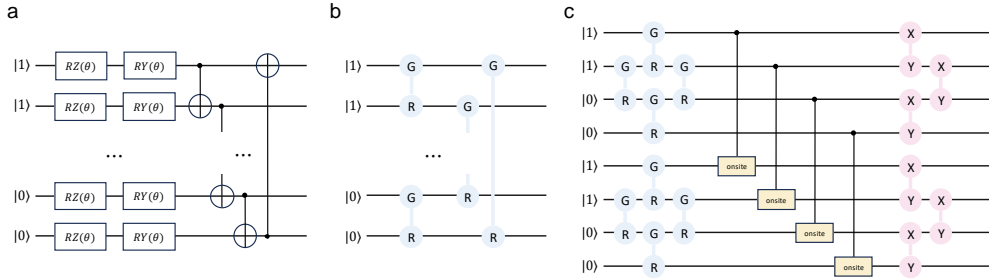
**Fig. D10 Additional Results for simulating molecular electronic structures.** The energy error w.r.t. the number of parameters with grey region shows results within chemical accuracy.

serves as the initial state. The exact ground state for comparison is obtained via the FCI method using PySCF [42].

In Figure D10, we further examine the impact of the number of parameters, which is treated as a hyperparameter in the main text. Both the HEA and GR ansätze exhibit a consistent trend, indicating that they have reached the maximum expressibility of the ansatz but remain unable to approximate the ground state with sufficient accuracy. We exclude the UCCSD ansatz from this comparison, as the ansatz is fixed for a given Hamiltonian. Instead, we provide the number of parameters and the total number of gates for both the UCCSD and BS ansätze in Table D3. Notably, our findings reveal that the number of parameters required to reach overparameterization for molecular electronic structures is significantly smaller than expected, diverging from the anticipated  $d_k^2$ . For an error margin close to chemical accuracy, fewer than  $2 \times d_k$  parameters are sufficient. This observation is particularly impactful, as it suggests that solving a VQE for Fermionic system simulation may require a number of parameters that scales linearly with  $d_k$ , offering valuable insights into the potential advantages of the VQE algorithm on intermediate-scale quantum processors.

Molecules	H <sub>2</sub>	LiH	H <sub>2</sub> O	BeH <sub>2</sub>	F <sub>2</sub>
$n$	4	12	14	14	20
$k$	2	4	10	6	18
$d_k$	6	495	1001	3003	190
#params for UCCSD	12	640	1000	1488	684
#gates for UCCSD	158	12612	21090	31414	17166
#params for BS	40	3000	2240	7000	4000
#gates for BS	326	24016	17944	56020	32038

**Table D3** Statistics of molecules.  $n$  and  $k$  are the numbers of orbitals and electrons, respectively. The number of gates is counted when decomposed to basic parameterized rotation gates and simple two-qubit gates as CNOT or CZ. Note that the number of parameters is more than enough for BS ansatz (see Figure D10) to ensure the stability of the performance. We can reduce the number of gates to a similar value to UCCSD and still reach much better results. All the other baseline methods (e.g., HEA, GR) have parameters that are identical to those of the BS ansatz.



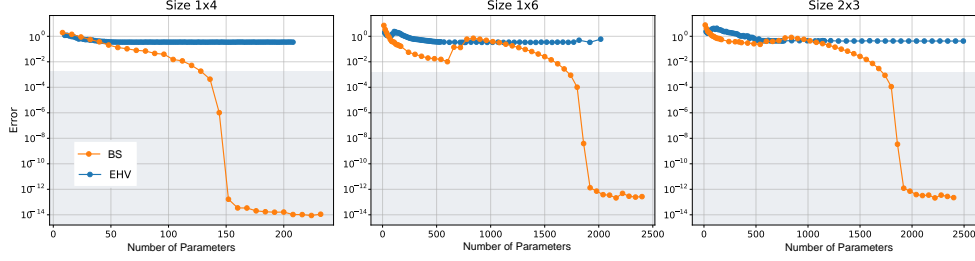
**Fig. D11 Ansatz of Baseline Methods.** **a** One layer of HEA ansatz utilizing parameterized rotation Y and Z gates and CNOT gates. The first  $k$  (number of electrons) qubits are initialized to  $|1\rangle$  as the HF state. **b** One layer of GR ansatz with  $n$  (number of orbitals) GR gates connecting neighboring qubits. **c** An example for EHV ansatz for  $1 \times 4$  sites Fermi-Hubbard model at half filling. Top four qubits for spin-ups and the rest for spin-downs. The GR gates are utilized for initialization, and the EHV ansatz requires repeatedly applying the onsite layer and XY-interaction layer.

## Appendix E The Fermi-Hubbard Model

We begin with a brief introduction to the Fermi-Hubbard model, focusing on how a variational quantum algorithm can be employed to solve it, with specific details on the implementation of the EHV ansatz [36]. The Hamiltonian of the Fermi-Hubbard model is given by:

$$H_{HF} = -t \sum_{i,j} \sum_{\sigma} t_{ij} (a_{i\sigma}^{\dagger} a_{j\sigma} + a_{j\sigma}^{\dagger} a_{i\sigma}) + U \sum_i a_{i\uparrow}^{\dagger} a_{i\uparrow} a_{i\downarrow}^{\dagger} a_{i\downarrow}. \quad (\text{E101})$$

For a Fermi-Hubbard model with a  $1 \times 4$  site configuration, there are four fermionic modes for spin-up and four for spin-down electrons. By applying the Jordan-Wigner transformation, each mode is mapped onto a qubit. As shown in Figure D11c, the first four qubits correspond to spin-up modes, while the remaining four correspond to spin-down modes, with qubits 1-5, 2-6, 3-7, 4-8 representing the two modes of a single site.



**Fig. E12 Additional Results for Fermi-Hubbard model.** The energy error w.r.t. the number of parameters with grey region shows results within chemical accuracy.

In a half-filled system, half of the sites are occupied, meaning each spin is represented by four qubits, with two in the  $|1\rangle$  state. Notably, hopping occurs only within the same spin, so the GR and XY interaction gates operate independently on either the top four or bottom four qubits. The EHV ansatz improves the initial state, which would otherwise be set to  $|11001100\rangle$ , by using one diamond-shaped layer of parameterized GR gates. The EHV ansatz is composed of alternating layers: one layer of onsite gates, coupling the two modes of each site, and one layer of NN XY-interaction gates within each spin. These layers are applied alternately to achieve the desired accuracy, with all gates in each layer sharing a single parameter. Notice that the EHV ansatz requires interactions between qubits on the same site and adjacent sites with the same spin. However, this configuration is impractical on a NISQ device without incorporating qubit swaps, regardless of the mapping approach. Thus, a full implementation of the EHV ansatz on quantum hardware necessitates the use of additional fermionic SWAP gates, which introduce a phase of  $-1$  on the  $|11\rangle \langle 11|$  state compared to standard SWAP gates. These additional gates increase the complexity, further highlighting the advantages of the proposed BS ansatz.

Similar to molecular electronic structures, we analyze the effect of the number of parameters on both the BS and EHV ansätze. All the numerical results for the Fermi-Hubbard model are the average of 5 random seeds. In [36], the authors report results with a single layer (two for the  $1 \times 4$  system) and suggest that "increased ansatz depth can lead to higher performance." However, as shown in Figure E12, increasing the EHV ansatz depth does not consistently improve performance, with none of the three instances achieving error within chemical accuracy. It seems the parameter-sharing strategy is severely damaging the expressivity of the EHV ansatz. In contrast, the BS ansatz successfully approximates the ground state with high precision. Notably, the results for the  $1 \times 6$  and  $2 \times 3$  systems show fluctuations in error as the number of parameters increases, likely due to the increased complexity in the training landscape. As with the molecular electronic structures, we observe that the number of parameters required to reach overparameterization scales linearly with  $d_k$  ( $d_k[1 \times 4] = 70$ ,  $d_k[1 \times 6, 2 \times 3] = 924$ ). This suggests that adding correlations does not increase the difficulty of solving the system with the BS ansatz. Future work will focus on extending the symmetry-preserving VQE solver for fermionic systems, aiming to provide deeper theoretical insights into this observation.

## References

- [1] Preskill, J.: Quantum computing in the nisq era and beyond. *Quantum* **2**, 79 (2018)
- [2] Bharti, K., Cervera-Lierta, A., Kyaw, T.H., Haug, T., Alperin-Lea, S., Anand, A., Degroote, M., Heimonen, H., Kottmann, J.S., Menke, T., *et al.*: Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics* **94**(1), 015004 (2022)
- [3] Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K., McClean, J.R., Mitarai, K., Yuan, X., Cincio, L., *et al.*: Variational quantum algorithms. *Nature Reviews Physics* **3**(9), 625–644 (2021)
- [4] Tilly, J., Chen, H., Cao, S., Picozzi, D., Setia, K., Li, Y., Grant, E., Wossnig, L., Rungger, I., Booth, G.H., *et al.*: The variational quantum eigensolver: a review of methods and best practices. *Physics Reports* **986**, 1–128 (2022)
- [5] Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028* (2014)
- [6] Romero, J., Babbush, R., McClean, J.R., Hempel, C., Love, P.J., Aspuru-Guzik, A.: Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology* **4**(1), 014008 (2018)
- [7] Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., Gambetta, J.M.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *nature* **549**(7671), 242–246 (2017)
- [8] McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. *Nature communications* **9**(1), 4812 (2018)
- [9] Ragone, M., Bakalov, B.N., Sauvage, F., Kemper, A.F., Ortiz Marrero, C., Larocca, M., Cerezo, M.: A lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nature Communications* **15**(1), 7172 (2024)
- [10] Sone, A., Volkoff, T., Cincio, L., Coles, P.J.: Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications* **12**(1), 1791 (2021)
- [11] Bravyi, S.B., Kitaev, A.Y.: Fermionic quantum computation. *Annals of Physics* **298**(1), 210–226 (2002)
- [12] Bergeal, N., Vijay, R., Manucharyan, V., Siddiqi, I., Schoelkopf, R., Girvin, S., Devoret, M.: Analog information processing at the quantum limit with a josephson ring modulator. *Nature Physics* **6**(4), 296–302 (2010)



- [13] Hubbard, J.: Electron correlations in narrow energy bands. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **276**(1365), 238–257 (1963)
- [14] Heisenberg, W.: Zur Theorie des Ferromagnetismus. Springer, ??? (1985)
- [15] Wecker, D., Hastings, M.B., Wiebe, N., Clark, B.K., Nayak, C., Troyer, M.: Solving strongly correlated electron models on a quantum computer. Physical Review A **92**(6), 062318 (2015)
- [16] Jiang, Z., Sung, K.J., Kechedzhi, K., Smelyanskiy, V.N., Boixo, S.: Quantum algorithms to simulate many-body physics of correlated fermions. Physical Review Applied **9**(4), 044036 (2018)
- [17] Hadfield, S., Wang, Z., O’gorman, B., Rieffel, E.G., Venturelli, D., Biswas, R.: From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. Algorithms **12**(2), 34 (2019)
- [18] Bacon, D., Kempe, J., DiVincenzo, D., Lidar, D., Whaley, K.: Encoded universality in physical implementations of a quantum computer. arXiv preprint quant-ph/0102140 (2001)
- [19] Terhal, B.M., DiVincenzo, D.P.: Classical simulation of noninteracting-fermion quantum circuits. Physical Review A **65**(3), 032325 (2002)
- [20] Gard, B.T., Zhu, L., Barron, G.S., Mayhall, N.J., Economou, S.E., Barnes, E.: Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. npj Quantum Information **6**(1), 10 (2020)
- [21] Wang, S., Fontana, E., Sharma, K., Sone, A., Cincio, L., Coles, P.J.: Noise-induced barren plateaus in variational quantum algorithms. Nature communications **12**(1), 6961 (2021)
- [22] Nielsen, M.A., Chuang, I.: Quantum computation and quantum information. American Association of Physics Teachers (2002)
- [23] Chong, F.T., Franklin, D., Martonosi, M.: Programming languages and compiler design for realistic quantum hardware. Nature **549**(7671), 180–187 (2017)
- [24] Khatri, S., LaRose, R., Poremba, A., Cincio, L., Sornborger, A.T., Coles, P.J.: Quantum-assisted quantum compiling. Quantum **3**, 140 (2019)
- [25] Guo, S., Sun, J., Qian, H., Gong, M., Zhang, Y., Chen, F., Ye, Y., Wu, Y., Cao, S., Liu, K., et al.: Experimental quantum computational chemistry with optimized unitary coupled cluster ansatz. Nature Physics, 1–7 (2024)
- [26] Ramakrishna, V., Salapaka, M.V., Dahleh, M., Rabitz, H., Peirce, A.: Controllability of molecular systems. Physical Review A **51**(2), 960 (1995)

- [27] Życzkowski, K., Kus, M.: Random unitary matrices. *Journal of Physics A: Mathematical and General* **27**(12), 4235 (1994)
- [28] Mezzadri, F.: How to generate random matrices from the classical compact groups. *arXiv preprint math-ph/0609050* (2006)
- [29] Larocca, M., Ju, N., García-Martín, D., Coles, P.J., Cerezo, M.: Theory of over-parametrization in quantum neural networks. *Nature Computational Science* **3**(6), 542–551 (2023)
- [30] Anand, A., Schleich, P., Alperin-Lea, S., Jensen, P.W.: Sukin sim, manuel díaz-tinoco, jakob s. kottmann, matthias degroote, artur f. izmaylov, and alán aspurguzik. a quantum computing view on unitary coupled cluster theory. *Chem. Soc. Rev* **51**(5), 1659–1684 (2022)
- [31] Knowles, P.J., Handy, N.C.: A new determinant-based full configuration interaction method. *Chemical physics letters* **111**(4-5), 315–321 (1984)
- [32] LeBlanc, J.P., Antipov, A.E., Becca, F., Bulik, I.W., Chan, G.K.-L., Chung, C.-M., Deng, Y., Ferrero, M., Henderson, T.M., Jiménez-Hoyos, C.A., *et al.*: Solutions of the two-dimensional hubbard model: benchmarks and results from a wide range of numerical algorithms. *Physical Review X* **5**(4), 041041 (2015)
- [33] Cade, C., Mineh, L., Montanaro, A., Stanisic, S.: Strategies for solving the fermi-hubbard model on near-term quantum computers. *Physical Review B* **102**(23), 235122 (2020)
- [34] Cai, Z.: Resource estimation for quantum variational simulations of the hubbard model. *Physical Review Applied* **14**(1), 014059 (2020)
- [35] Anselme Martin, B., Simon, P., Rančić, M.J.: Simulating strongly interacting hubbard chains with the variational hamiltonian ansatz on a quantum computer. *Physical Review Research* **4**(2), 023190 (2022)
- [36] Stanisic, S., Bosse, J.L., Gambetta, F.M., Santos, R.A., Mruzckiewicz, W., O’Brien, T.E., Ostby, E., Montanaro, A.: Observing ground-state properties of the fermi-hubbard model using a scalable algorithm on a quantum computer. *Nature communications* **13**(1), 5743 (2022)
- [37] Slater, J.C.: The theory of complex spectra. *Physical review* **34**(10), 1293 (1929)
- [38] Helgaker, T., Jorgensen, P., Olsen, J.: *Molecular Electronic-structure Theory*. John Wiley & Sons, ??? (2013)
- [39] Zeier, R., Schulte-Herbrüggen, T.: Symmetry principles in quantum systems theory. *Journal of mathematical physics* **52**(11) (2011)
- [40] d’Alessandro, D.: *Introduction to Quantum Control and Dynamics*. CRC press,

??? (2021)

- [41] Schirmer, S.G., Fu, H., Solomon, A.I.: Complete controllability of quantum systems. *Physical Review A* **63**(6), 063410 (2001)
- [42] Sun, Q., Berkelbach, T.C., Blunt, N.S., Booth, G.H., Guo, S., Li, Z., Liu, J., McClain, J.D., Sayfutyarova, E.R., Sharma, S., *et al.*: Pyscf: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8**(1), 1340 (2018)
- [43] Bartlett, R.J., Kucharski, S.A., Noga, J.: Alternative coupled-cluster ansätze ii. the unitary coupled-cluster method. *Chemical physics letters* **155**(1), 133–140 (1989)
- [44] Huang, H.-L., Xu, X.-Y., Guo, C., Tian, G., Wei, S.-J., Sun, X., Bao, W.-S., Long, G.-L.: Near-term quantum computing techniques: Variational quantum algorithms, error mitigation, circuit compilation, benchmarking and classical simulation. *Science China Physics, Mechanics & Astronomy* **66**(5), 250302 (2023)
- [45] Cai, Z., Babbush, R., Benjamin, S.C., Endo, S., Huggins, W.J., Li, Y., McClean, J.R., O’Brien, T.E.: Quantum error mitigation. *Reviews of Modern Physics* **95**(4), 045005 (2023)
- [46] Fowler, A.G., Mariantoni, M., Martinis, J.M., Cleland, A.N.: Surface codes: Towards practical large-scale quantum computation. *Physical Review A—Atomic, Molecular, and Optical Physics* **86**(3), 032324 (2012)
- [47] Kovalev, A.A., Pryadko, L.P.: Fault tolerance of quantum low-density parity check codes with sublinear distance scaling. *Physical Review A* **87**(2), 020304 (2013)
- [48] Breuckmann, N.P., Eberhardt, J.N.: Quantum low-density parity-check codes. *PRX Quantum* **2**(4), 040101 (2021)
- [49] Bausch, J., Senior, A.W., Heras, F.J., Edlich, T., Davies, A., Newman, M., Jones, C., Satzinger, K., Niu, M.Y., Blackwell, S., *et al.*: Learning high-accuracy error decoding for quantum processors. *Nature*, 1–7 (2024)
- [50] Viola, L., Fortunato, E.M., Pravia, M.A., Knill, E., Laflamme, R., Cory, D.G.: Experimental realization of noiseless subsystems for quantum information processing. *Science* **293**(5537), 2059–2063 (2001)
- [51] Pyshkin, P., Luo, D.-W., Wu, L.-A.: Self-protected adiabatic quantum computation. *Physical Review A* **106**(1), 012420 (2022)
- [52] Xu, X., Cui, J., Cui, Z., He, R., Li, Q., Li, X., Lin, Y., Liu, J., Liu, W., Lu, J., *et al.*: Mindspore quantum: a user-friendly, high-performance, and ai-compatible quantum computing framework. *arXiv preprint arXiv:2406.17248* (2024)