Ansatz Choice for Variational Quantum Algorithms for Quantum Chemistry in the NISQ era

Hasan Saygınel*
Department of Physics & Astronomy, University College London, London, WC1E 6BT, United Kingdom
(Dated: March 4, 2022)

Variational quantum algorithms are one of the promising near-term algorithms that aim to show a quantum advantage in the area of quantum chemistry. A state preparation circuit, known as an Ansatz, lies at the heart of the algorithm and determines the expressivity of the model it represents. The ansatz also determines the circuit properties such as the number of non-local gates in the circuit and the circuit depth, which are particularly important considerations when using Noisy Intermediate-Scale Quantum (NISQ) processors that have limited qubit count, sparse connectivity as well as short coherence times. Following a brief discussion of recently proposed metrics for the properties of a 'good' ansatz, this paper discusses two types of ansätze in the literature; chemically-motivated & hardware efficient ansatz. Near-term and future outlook in this field is also discussed at the end, where current research is directing towards hardware-efficient ansätze where physical properties such as symmetries in quantum numbers are encoded in the circuit such that only the relevant parts of the Hilbert space are explored.

I. INTRODUCTION

Variational quantum algorithms (VQAs) [1–3] fall under the family of hybrid quantum-classical algorithms where the computational tasks are allocated between quantum and classical devices based on their advantage. The classical computer is used for storage, classical pre and post-processing as well as for optimisation. The quantum processor computes expectation values of a Hermitian operator, $\mathcal{H}$, for different parameters of a parameterised quantum state $|\psi(\boldsymbol{\theta})\rangle$ where $\boldsymbol{\theta}$ represents a vector of variational parameters. The variational state is prepared using a state preparation circuit known as an ansatz via the application of a unitary quantum circuit $|\psi(\boldsymbol{\theta})\rangle = V(\boldsymbol{\theta})\,|0\rangle^{\otimes n}$ on the n-fold tensor product of state $|0\rangle$. The expectation value can be expressed as:

$$\langle\mathcal{H}\rangle_{V(\theta)} = \langle 0|V(\boldsymbol{\theta})^{\dagger}\mathcal{H}V(\boldsymbol{\theta})|0\rangle \tag{1}$$

where $|0\rangle$ represents an n-fold tensor product of state $|0\rangle$.

A. VQAs for Quantum Chemistry Problems

The first proposed and illustrated VQA was the Variational Quantum Eigensolver (VQE) [4] which achieves ground state estimation of a Hamiltonian based on the Rayleigh-Ritz variational principle that bounds the ground state energy. VQE can be used to find the ground state of molecules (which has already been demonstrated for small molecules [5–7]) that forms the basis of other chemistry calculations such as reaction rates. More recently, VQAs for Hamiltonian simulation have also been proposed [8]. This paper mainly focuses on the VQE algorithm; details of which are given below.

* hasan.sayginel.17@ucl.ac.uk

Classical preprocessing. The first step is to define the problem Hamiltonian for an electronic wave function. Working in the second quantisation formalism, the orbital occupation number basis is used, where the elements of the basis are given as $|n_0, n_1, \ldots, n_{M-1}\rangle$ for a molecule with M spin orbitals and $n_i = 0, 1$ indicate the occupation of each electronic orbital [9]. In this basis, the Hamiltonian is represented as

$$\mathcal{H}_{fm} = \sum_{p,q} h_{pq}\hat{a}_p^{\dagger}\hat{a}_q + \frac{1}{2}\sum_{p,q,r,s} h_{pqrs}\hat{a}_p^{\dagger}\hat{a}_q^{\dagger}\hat{a}_r\hat{a}_s \tag{2}$$

where the subscript 'fm' represents a fermionic Hamiltonian, 'p, q, r, s' label single electron orbitals, a, $a^{\dagger}$ are the creation and annihilation operators and h_{pq}, h_{pqrs} are the one and two-Coulomb integrals respectively. This representation has to be mapped into a qubit representation to be implemented on a quantum processor. There are two common encoding techniques that achieves this, namely Jordan-Wigner encoding & Bravyi-Kitaev encoding [10]. In the Jordan-Wigner representation, the wavefunction can be mapped onto qubits in a straightforward way as

$$|n_0, n_1, \ldots, n_{M-1}\rangle \rightarrow |q_0, q_1 \ldots, q_{M-1}\rangle \tag{3}$$

where each qubit represents 1 spin-orbital. Therefore, a circuit with M qubits are needed for a molecule with M orbitals. $\mathcal{H}_{fm}$ also needs to be mapped into an operator that acts on qubits. Both Jordan-Wigner & Bravyi-Kitaev encoding map $\mathcal{H}_{fm}$ into a linear combination of Pauli strings of arbitrary length N, where $N \leq M$.

$$\mathcal{H}_{qu} = \sum_i c_i h_i \tag{4}$$

where h_i are individual terms of $\mathcal{H}_{qu}$. h_i are N-qubit tensor products of Pauli operators $\sigma_n \in \{X, Y, Z, I\}$.

$$h_i = \bigotimes_{n=0}^{N-1} \sigma_n \tag{5}$$

and σ_n can be implemented on quantum hardware.

Variational circuit. The variational circuit $V(\theta)$ contains both fixed quantum gates such as CNOTs and parameterised gates such as qubit rotations that can be optimised. The choice of an ansatz for $V(\theta)$ is arbitrary but should relate to the problem. See Section II.

Measurement. The quantum circuit should be measured for each h_i. In practice, this corresponds to multiple runs of the quantum circuit with different measurements each time. Due to the uncertainty principle, non-commuting observables cannot be measured simultaneously. However, if one has access to multiple quantum processors, multiple identical states can be prepared so that multiple non-commuting measurements can be taken in parallel. The outcome of each measurement for each h_i can be represented as

$$\langle M_i \rangle = \langle 0|V(\boldsymbol{\theta})^\dagger h_i V(\boldsymbol{\theta})|0\rangle \tag{6}$$

Classical postprocessing & optimisation. The measurement values from the quantum computer are used to estimate the overall expectation value of the Hamiltonian $\mathcal{H}_{qu}$. In VQE, the aim is to minimise $\langle \mathcal{H}_{qu} \rangle$ with respect to $\boldsymbol{\theta}$. To this end, a loss function of the following form is defined to minimise the expectation value which corresponds to the energy of the state.

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_i c_i \langle 0|V(\boldsymbol{\theta})^\dagger h_i V(\boldsymbol{\theta})|0\rangle \tag{7}$$

Each measurement circuit is run multiple times and the parameters are updated at each iteration via a classical optimisation method such as gradient descent.

In the following, the challenges associated with practically implementing VQE are discussed and associated with proposed qualitative and quantitative metrics when it comes to the Ansatz choice for $V(\boldsymbol{\theta})$.

B. Challenges with VQE

1. The first challenge arises due to equal scaling of the number of qubits with the number of orbitals in a molecule in the Jordan-Wigner representation. In the NISQ era, there is a strict bound on the number of qubits available and therefore this already prevents large molecules from being simulated. A practical way to get around this is to consider only the active orbitals [2].

2. The number of terms in the qubit version of the electronic structure Hamiltonian, $\mathcal{H}_{qu}$, scales as $O(M^4)$ where M is the number of spin orbitals as before and therefore there are many observables h_i that need to be evaluated. This bottleneck demonstrates the fact that quantum circuit complexity is offloaded to a lot of different measurements motivating efficient measurement techniques [3].

3. In general, it is unclear a priori what constitutes a good Ansatz for a problem and in near-term computing the performance of an Ansatz also depends on the quantum processor for which multiple qubit technologies exist. NISQ devices have varying native gate sets, topologies and gate error rates. This requires compilation which may introduce too much overhead for theoretically-devised Ansätze meaning the algorithm cannot be executed within the coherence time of qubits.

4. Optimisation is another challenge because poor variable initialisation with a general Ansatz may lead to barren plateaus which are flat regions in the cost landscape with exponentially vanishing gradient and variance. Furthermore, optimisation difficulty can be associated with the number of variational gates in the circuit [11].

C. Ansatz Performance Metrics

The current best way to evaluate the performance of an Ansatz for a particular problem such as VQE is to numerically benchmark it for the use-case with a particular quantum processor because the performance depends both on how well the ansatz explores the Hilbert space and hardware limitations of the processor.

However, [12] suggested some qualifiers for choosing an appropriate Ansatz; (1) gates required to model the target quantum *approximately* should ideally scale polynomially, (2) the gates should be easily physically realisable (mostly rely on the native gate sets of NISQ processors), (3) multi-qubit gates should have compatible locality with the topology of the processor to minimise the SWAP gates, (4) robust convergence behaviour during optimisation (mitigating effects of barren plateaus), (5) ansatz should be informed by symmetries of the quantum state (it should relate to theory to some extent), (6) at sufficient (*exponential*) depth the approximation should converge to the exact representation of the target quantum state.

In addition, two relevant quantitative metrics were introduced by [13] to assess the modelling power of an Ansatz in an abstract sense; *expressibility* and *entangling capability* [3].

Expressibility. Defined as the deviation of states generated by an Ansatz $|\psi_{\boldsymbol{\theta}}\rangle$ from a uniform sample of the full Hilbert space known as the Haar measure. $|\psi_{\boldsymbol{\theta}}\rangle$ is sampled for randomly chosen parameters $\boldsymbol{\theta}$ and the Hilbert-Schmidt norm is calculated as follows.

$$A^{(t)} = || \int_{Haar} (|\psi\rangle \langle\psi|)^{\otimes t} \, d\psi - \int_{\boldsymbol{\theta}} (|\psi_{\boldsymbol{\theta}}\rangle \langle\psi_{\boldsymbol{\theta}}|)^{\otimes t} \, d\psi_{\boldsymbol{\theta}} ||^2_{HS} \quad (8)$$

where $|\psi\rangle$ is a state distributed according to the Haar measure and $||M||_{HS} = \sqrt{\mathrm{Tr}(M^\dagger M)}$ is the Hilbert Schmidt norm. Smaller $A^{(t)}$ represents a more expressive Ansatz.

Entangling capability. This measure describes the power of the Ansatz to generate entangled states using the Meyer-Wallach Q measure [14]. For this measure, a linear map $\iota_j(e)$ is defined which acts on the computational basis $\iota_j(b) = |b_1 \dots b_n\rangle = \delta_{bb_j} \left| b_1 \dots \tilde{b}_j \dots b_n \right\rangle$ where $b_j \in \{0, 1\}$ and $\tilde{b}_j$ denotes the absence of the j-th qubit.

$$Q(|\psi\rangle) = \frac{4}{n} \sum_{j=1}^{M} D(\iota_j(0) \, |\psi\rangle \, , \iota_j(1) \, |\psi\rangle) \quad (9)$$

where D is the generalised distance defined by the coefficients of the two states $|u\rangle = \sum u_i \, |e_i\rangle$ and $|v\rangle = \sum v_i \, |e_i\rangle$.

$$D(|u\rangle \, , |v\rangle) = \frac{1}{2} \sum_{i,j} |u_i v_j - u_j v_i|^2 \quad (10)$$

See [3] & [13] for more information.

II. VARIATIONAL CIRCUIT ANSATZ

There are two approaches to constructing trial solutions by designing an ansatz $V(\boldsymbol{\theta})$; chemically-motivated and hardware-efficient approaches. Chemically-motivated ansätze follow from quantum chemistry and classical computational methods whereas in hardware-efficent ansatz constructions the aim is to construct general unitary circuits that has access to the parts of the Hilbert space where the target solution lies.

A. Chemically-motivated ansätze

Let us consider the general fermionic hamiltonian in Eq (4). This time-independent hamiltonian generates a unitary evolution ($\hbar = 1$) and can be expressed as an exponential of a sum of Pauli operators as given below.

$$U(t) = e^{-i\mathcal{H}_{qu}t} = e^{\sum_j -ic_j h_j t} \quad (11)$$

The unitary, $U(t)$, if implemented on a quantum computer would be able generate the eigenstates of H_{qu}, however it cannot be directly implemented as it has to be mapped into native operations available. h_i, on the other hand, are readily n-fold tensors of Pauli operators that can be easily implemented. The problem is that different h_i terms are usually non-commuting and as such $U(t)$ cannot be rewritten as a product of h_i terms in an exact way. In general for matrix algebra $e^{A+B} \neq e^A e^B$, where A and B are matrices. There exists an approximation known as Suzuki-Trotter decomposition in order to be able to write $U(t)$ as a product of Paulis while bounding the error. To first-order, this decomposition is expressed as

$$e^{\sum_j -ic_j h_j t} = \Big(\prod_{j=1}^{k} e^{\frac{-ic_j h_j t}{m}} \Big)^m + O\Big(\frac{t^2}{m}\Big) \quad (12)$$

where m are the Trotter steps. Increasing the trotter step reduces the error but also increases the circuit depth significantly. However, $m = 2$ is often sufficient to attain reasonable errors (i.e. to within chemical accuracy = 1.6 mHa or 4 kJ/mol) [3].

Unitary coupled cluster singles and doubles (UCCSD) ansatz

One of the commonly studied ansätze in the literature is the 'Unitary coupled cluster singles and doubles' ansatz which was inspired from its classical analogue; the non-unitary CCSD method. The UCCSD trial solution is generated from the Hartree-Fock wavefunction (calculated classically) by realising the following unitary operator on a quantum processor:

$$|\psi(\boldsymbol{\theta})\rangle = e^{T(\boldsymbol{\theta}) - T(\boldsymbol{\theta})^\dagger} \, |\Phi_0\rangle \quad (13)$$

$$T(\boldsymbol{\theta}) = T_1(\boldsymbol{\theta}) + T_2(\boldsymbol{\theta}) + \dots \quad (14)$$

where T refers to possible (multi-)electron excitations of a molecular system. In UCCSD, T is truncated at double excitations whereas the entire sum to infinity yields the 'full configuration interaction' (FCI) ansatz. In fermionic excitation representation, T_1 and T_2 are expressed as

$$T_1(\boldsymbol{\theta}) = \sum_{\substack{p \in \mathrm{virt} \\ r \in \mathrm{occ}}} \theta_{pr} \hat{a}_p^\dagger \hat{a}_r, \quad T_2(\boldsymbol{\theta}) = \sum_{\substack{p,q \in \mathrm{virt} \\ r,s \in \mathrm{occ}}} \theta_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{c}_s \quad (15)$$

where the creation operators $\hat{a}^\dagger$ act on virtual (empty) orbitals and annihilation operators $\hat{a}$ act on occupied (full) orbitals. After trotterisation, the fermionic unitary operator is given as

$$\begin{aligned} U(\boldsymbol{\theta}) = &\prod_{p>r} \exp\{\theta_{pr} \hat{a}_p^\dagger \hat{a}_r - \text{h.c.}\} \times \\ &\prod_{p>q>r>s} \exp\{\theta_{pqrs} (\hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{c}_s) - \text{h.c.}\} \end{aligned} \quad (16)$$

which results in the following unitary qubit operator following the Jordan-Wigner encoding that can be expressed in the form of a circuit.

$$U(\boldsymbol{\theta}) = \prod_{p>r} \exp\left\{ \frac{i\theta_{pr}}{2} \bigotimes_{a=r+1}^{p-1} Z_a (Y_r X_p - \text{h.c.}) \right\} \times \prod_{p>q>r>s} \exp\left\{ \frac{i\theta_{pqrs}}{8} \bigotimes_{b=s+1}^{r-1} Z_b \bigotimes_{a=q+1}^{p-1} Z_a (X_s X_r Y_q X_q + \dots) \right\} \tag{17}$$

The appendix includes circuit diagrams of Eq (17). The realisation of the UCCSD ansatz to model single and double fermionic excitations leads to ladders of CNOT gates. NISQ processors have high multi-qubit gate error rates and as such UCCSD ansatz has a high cost in terms of circuit depth and accumulated error.

Variations of the UCCSD exist in literature such as k-upUCCSD ansatz [15].

B. Hardware-efficient ansatz

Hardware-efficient ansätze focus on the current architecture and limitations of NISQ devices. Firstly, the ansatz only includes single and two-qubit gates in correspondence with the gate set of the device. Particularly, the two-qubit gate is taken to be the native two-qubit operation available. The goal of hardware efficient approaches is to create an expressive enough circuit that includes the target solution while maintaining a suitable circuit depth. This type of ansätze can be generalised in the following way.

$$V(\boldsymbol{\theta}) = \prod_{k=1}^{L} U_k(\boldsymbol{\theta}_k) W_k \tag{18}$$

where L is a hyperparameter of the model called layer corresponding to the number of repetitions of unitary blocks, $U_k(\boldsymbol{\theta}_k)$ is a parameterised layer which takes as input the variational parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L)$, whereas W_k contains only fixed gates. Typically, U_k consists of single-qubit rotation gates and W_k of two-qubit entangling gates, such as in the *Alternating Layered Ansatz* [16]. In this ansatz, blocks of parameterised and entangling layers are repeated L times. It has been demonstrated that this ansatz models a partial Fourier series where the Fourier coefficients and the frequency spectrum are determined by the ordering of gates as well as the depth. This adds another tool to investigate the expressivity of this ansatz in addition to those described in section I C.

C. Comparison between chemical and hardware-efficient ansätze

The theory behind quantum chemistry informs us regarding the form of the solution and chemical ansätze exploit this by constructing unitaries to directly model fermionic excitations. Hardware-efficient approaches are abstract and focus on optimising implementations on NISQ processors. It is important to realise that the overall goal is not to explore the Hilbert space of an n-qubit system generally, but to rather target the portion of the Hilbert space which contains the solution. Focusing on a subspace around the solution aids with the search during optimisation. Furthermore, hardware-efficient ansätze have been shown to result in the barren plateau phenomenon at some instances, particularly when many parameters $\boldsymbol{\theta}$ are initialised randomly; increasing the difficulty of finding the global minimum [11].

It is believed by some researchers that the hardware-efficient approaches may not scale well for increasing numbers of qubits if they model an ever-increasing exponentially large Hilbert space. Recent approaches focused on mixing two types of ansätze to incorporate symmetries of quantum physics while remaining hardware-efficient for NISQ applications.

III. GRADIENTS OF QUANTUM CIRCUITS

The optimisation stage of the overall VQE algorithm deserves discussion because of its dependence on the ansatz choice. The loss function, defined in Eq (7) can be regarded as a scalar-valued objective function and the optimisation problem can be posed as the minimisation of $\mathcal{L}(\boldsymbol{\theta})$ with respect to the variational parameters $\boldsymbol{\theta}$ of the quantum circuit. There exists two types of classical optimisation algorithms; gradient-based and gradient-free approaches.

Gradient-free algorithms such as 'simultaneous perturbation stochastic approximation' (SPSA) and 'constrained optimization by linear approximation' (COBYLA) can be used readily while gradient-based algorithms require the computation of a quantum gradient (either analytically or numerically).

Gradient-based algorithms are typically an extension of the vanilla gradient descent update rule expressed as

$$\theta_i{}^{(t+1)} = \theta_i^{(t)} - \eta^{(t)} \boldsymbol{\nabla}_{\theta_i} \mathcal{L}(\boldsymbol{\theta}) \tag{19}$$

where η is the learning rate and the gradient $\boldsymbol{\nabla}_{\theta_i} = (\partial_{\theta_1}, \dots, \partial_{\theta_m})^T$ is taken with respect to each parameter at every iteration. $\partial_{\theta_i} \mathcal{L}$ terms can be calculated numerically using a finite difference formula:

$$\partial_{\theta_i} \mathcal{L} = \frac{\mathcal{L}(\theta_i + \frac{1}{2}\Delta\mu) - \mathcal{L}(\theta_i - \frac{1}{2}\Delta\mu)}{\Delta\mu} \tag{20}$$

where $\Delta\mu$ is an infinitesimal shift in the parameter θ_i. In practice, noise in NISQ processors might be too large to accurately measure the circuit output for infinitesimally varied parameters. Computing quantum gradients using a finite shift is also possible based on the parameter-shift rule [17].

A. Parameter-shift rule

The parameter-shift rule allows for an exact method of computing the gradient of a parameterised gate by running the same quantum circuit twice with a finite shift in the gate parameter.

Definition 1 *(Parameter-shift rule). The parameter-shift rule states that if the generator G of the gate $\mathcal{G}(\theta_i)$ (with a single parameter θ_i) has two unique eigenvalues denoted as e_0 and e_1, then the derivative of the circuit expectation with respect to the gate parameter θ_i is given by the difference in expectation of two circuits with shifted parameters scaled by a parameter r given as $r = \frac{e_1 - e_0}{2}$.*

$$\partial_{\theta_i}\mathcal{L} = r[\mathcal{L}(\theta_i + \frac{\pi}{4r}) - \mathcal{L}(\theta_i - \frac{\pi}{4r})] \tag{21}$$

where r is the scaling parameter and $\frac{\pi}{4r}$ is a finite shift. A gate with a gate generator that does not meet the only 2 unique eigenvalues requirement can still be differentiated by first decomposing it into a sequence of parameter-shift rule differentiable gates and using the chain rule [18][19].

B. Usefulness of analytical gradients

Recently, the parameter-shift rule was used to derive analytic expressions for UCCSD type circuit gradients [20]. Furthermore, the realisation that gradients of gates can be computed without a modification to the quantum circuit structure has allowed automatic differentiation methods to be utilised in quantum algorithms (introduced first by Pennylane[9]).

IV. FUTURE WORK

Current directions of research are focusing on reducing the number of h_i measurements Eq (6) needed [21], creating adaptive methods of ansatz construction [22], combining hardware-efficient approaches and theory [12] as well as developing better NISQ-era protocols such as classical shadow tomography [23] and error mitigation [24]. Since the focus of this paper is ansatz structures, only the adaptive method approaches called Adapt-VQE and hardware-efficient symmetry preserving ansatz, 'QNP', are elaborated on below.

A. Adapt-VQE

Adapt-VQE algorithm consists of an ansatz that starts with an identity gate and concatenates new operators at each iteration of the algorithm. The new operators added are chosen from a pool of operators for which the quantum gradient is calculated. Note that the pool of operators used for Adapt-VQE is the same as in full configuration interaction (FCI) and as such, the asymptotic scaling is exact. The operator with the largest gradient is chosen at each iteration. See [22] for more information.

B. Quantum number preserving (QNP) Ansätze

Ansätze can be constructed in such a way that only hardware-efficient gates are used yet the unitaries formed by joint action of them are motivated by theory. Unitaries that map qubits to physically relevant parts of the Hilbert space are useful. Givens rotations, for example, preserve particle number symmetry. As an example, single excitation Givens rotation is given as:

$$\begin{aligned} G(\phi)\,|01\rangle &= \cos(\phi/2)\,|01\rangle + \sin(\phi/2)\,|10\rangle \\ G(\phi)\,|10\rangle &= \cos(\phi/2)\,|01\rangle + \sin(\phi/2)\,|10\rangle \end{aligned} \tag{22}$$

while leaving the subspaces $|00\rangle$ and $|11\rangle$ invariant. In matrix form $G(\theta)$ is expressed as:

$$G(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\phi/2) & -\sin(\phi/2) & 0 \\ 0 & \sin(\phi/2) & \cos(\phi/2) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{23}$$

$G(\phi)$ can be efficiently realised on a NISQ processor and the double excitation extension $G^{(2)}(\phi)$ is also efficiently realisable.

ACKNOWLEDGMENTS

I would like to thank Peter Coveney for supervising this project and Alexis Ralli & Tim Weaving for the insightful discussions.

Appendix A: Ansatz Circuit Diagrams

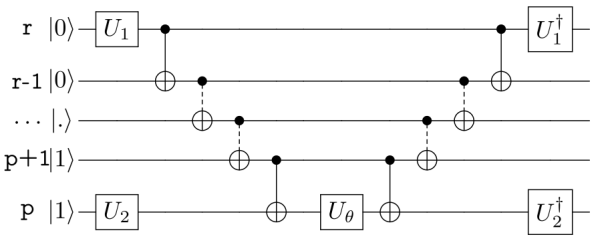

FIG. 1. UCCSD - Single Fermion Excitation. U-gates represent single qubit unitaries. (Taken from [25]).

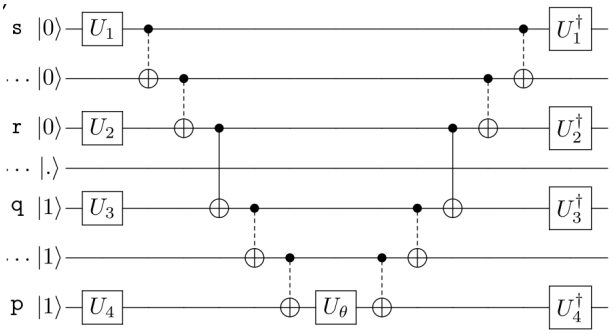

FIG. 2. UCCSD - Fermionic Double Excitation. U-gates represent single qubit unitaries. (Taken from [25].)

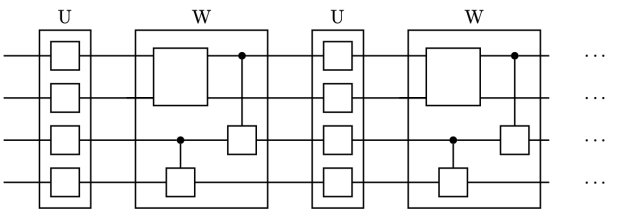

FIG. 3. General structure of hardware-efficient ansätze. U layer contains only single qubit gates whereas the W layer contains two-qubit entangling gates.

Appendix B: Parameter-Shift rule

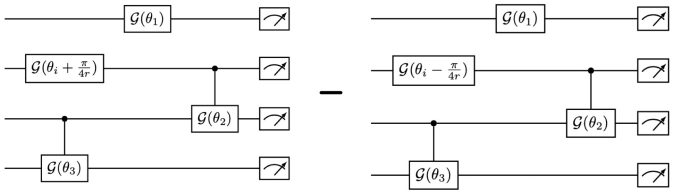

FIG. 4. Parameter shift rule. Same circuit is run twice with shifted parameters to calculate the gradient for 1 gate with respect to its variational parameter. This procudure is repeated for every gate for which a gradient needs to be calculated.

[1] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Science and Technology **4**, 043001 (2019).

[2] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, *Chemical Reviews*, Chemical Reviews **119**, 10856 (2019).

[3] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Reviews of Modern Physics **94**, 015004 (2022), arXiv:2101.08448 [quant-ph].

[4] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nature Communications **5**, 4213 (2014).

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, S. Demura, A. Dunsworth, D. Eppens, E. Farhi, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, S. Habegger, M. P. Harrigan, A. Ho, S. Hong, T. Huang, W. J. Huggins, L. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, C. Jones, D. Kafri, K. Kechedzhi, J. Kelly, S. Kim, P. V. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, P. Laptev, M. Lindmark, E. Lucero, O. Martin, J. M. Martinis, J. R. McClean, M. McEwen, A. Megrant, X. Mi, M. Mohseni, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, H. Neven, M. Yuezhen Niu, T. E. O'Brien, E. Ostby, A. Petukhov, H. Putterman, C. Quintana, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, D. Strain, K. J. Sung, M. Szalay, T. Y. Takeshita, A. Vainsencher, T. White, N. Wiebe, Z. J. Yao, P. Yeh, and A. Zalcman, Hartree-Fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020), arXiv:2004.04174 [quant-ph].

[6] Ground-state energy estimation of the water molecule on a trapped-ion quantum computer, npj Quantum Information **6**, 33 (2020).

[7] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, Doubling the Size of Quantum Simulators by Entanglement Forging, PRX Quantum **3**, 010309 (2022), arXiv:2104.10220 [quant-ph].

[8] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[9] PennyLane dev team, Quantum Chemistry Tutorials, `https://pennylane.ai/qml/demos/tutorial\_quantum\_chemistry.html` (2021), online; accessed 31 January 2022.

[10] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, A comparison of the bravyi–kitaev and jordan–wigner transformations for the quantum simulation of quantum chemistry, *Journal of Chemical Theory and Computation*, Journal of Chemical Theory and Computation **14**, 5617 (2018).

[11] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications **9**, 4812 (2018).

[12] G.-L. R. Anselmetti, D. Wierichs, C. Gogolin, and R. M. Parrish, Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems, New Journal of Physics **23**, 113010 (2021), arXiv:2104.05695 [quant-ph].

[13] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, arXiv e-prints , arXiv:1905.10876 (2019), arXiv:1905.10876 [quant-ph].

[14] D. A. Meyer and N. R. Wallach, Global entanglement in multiparticle systems, Journal of Mathematical Physics **43**, 4273 (2002), https://doi.org/10.1063/1.1497700.

[15] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, *Journal of Chemical Theory and Computation*, Journal of Chemical Theory and Computation **15**, 311 (2019).

[16] M. Schuld, R. Sweke, and J. J. Meyer, Effect of data encoding on the expressive power of variational quantum-machine-learning models, Phys. Rev. A **103**, 032430 (2021), arXiv:2008.08605 [quant-ph].

[17] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018), arXiv:1803.00745 [quant-ph].

[18] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).

[19] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. Sohaib Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, PennyLane: Automatic differentiation of hybrid quantum-classical computations, arXiv e-prints , arXiv:1811.04968 (2018), arXiv:1811.04968 [quant-ph].

[20] J. S. Kottmann, A. Anand, and A. Aspuru-Guzik, A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers, Chem. Sci. **12**, 3497 (2021).

[21] A. Ralli, P. J. Love, A. Tranter, and P. V. Coveney, Implementation of measurement reduction for the variational quantum eigensolver, Phys. Rev. Research **3**, 033195 (2021).

[22] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nature Communications **10**, 3007 (2019).

[23] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nature Physics **16**, 1050 (2020), arXiv:2002.08953 [quant-ph].

[24] Y. Guo and S. Yang, Quantum error mitigation via matrix product operators, arXiv e-prints , arXiv:2201.00752 (2022), arXiv:2201.00752 [quant-ph].

[25] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions, Phys. Rev. A **98**, 022322 (2018), arXiv:1805.04340 [quant-ph].