UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE CHIMICHE

CORSO DI LAUREA IN CHIMICA

VQE ALGORITHM FOR CHEMISTRY SIMULATION

Relatore: Prof. Antonino Polimeno

Laureando: Niccolò Galiazzo

Anno accademico: 2021/2022

Contents

Chapter 1

Quantum computing

1.1 Introduction

The variational quantum *eigensolver* (VQE) [1] is a method that uses a hybrid quantum-classical computational approach to find the eigenvalues of a Hamiltonian. VQE has been proposed as an alternative to fully quantum algorithms, which require hardware not yet available and has been successfully applied to solve the electronic Schrödinger equation for a variety of small molecules. The scalability of this method is limited by two factors: the complexity of the quantum circuits and the complexity of the classical optimization problem. Both of these factors are affected by the choice of the variational ansatz used to represent the trial wave function.

The purpose of this work is to discuss the outlines of the VQE method, within the general framework of quantum computing. Therefore, I shall firstly present the core principles of quantum computing to solve computational chemistry problems [2]. I shall then discuss briefly which limiting factors exist for our present computational technology and comment on the circuit models that can be used to describe both classic and quantum algorithms.[3, 4, 5] Finally I will focus on the rules that govern the action of these circuits, after introducing basic quantum chemistry concepts like the density operator and its matrix representation.[6] An historical view of the implementation of quantum computing algorithms will be also provided, together with a comment on the potential usefulness of quantum computing to solve chemical computaitional problems.[7] Lastly some perspectives on the state of the art in quantum computing will be provided.[8, 9, 10, 11, 12]

1.2 Density Matrix

Quantum computing is a type of computation that uses quantum states to perform calculations, on devices named *quantum computers*. In quantum computing, a *qubit* is the basic unit of information, i.e. a two-state (or two-level) quantum-mechanical system, one of the simplest quantum systems displaying the peculiarity of quantum mechanics. Operation are performed on qubits in *circuits* (see above) through the application of *operators*.

In quantum computation and quantum communication, there are many practical scenarios in which the state of our qubits cannot be written down as linear combinations in a given basis, but instead must be expressed in terms of ensembles (statistical mixtures) of multiple states, each with an associated probability of occurrence. Simply put, the density matrix is an alternative way of expressing quantum states. However, unlike the state-vector representation, this formalism allows us to use the same mathematical language to describe both the more simple quantum pure states, as well as the quantum mixed states that consist of ensembles of pure states.

1.2.1 Pure state

Pure states are those for which we can precisely define their quantum state at every point in time. In general, we know that in the conventional state vector notation, an $n-$qubit pure state can be expressed as

$$|\Psi\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} \tag{1.1}$$

where $N = 2^n$. An alternative way to express this pure quantum state is in the form of a matrix. This can be done by using the density operator representation, which is defined as

$$\rho \equiv |\Psi\rangle \langle\Psi| \tag{1.2}$$

Here, the term $|\Psi\rangle\langle\Psi|$ represents the outer product of the state Ψ with itself

$$\rho = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} [\alpha_{\mathbf{0}}^*, \alpha_{\mathbf{1}}^*, ..., \alpha_{\mathbf{n}}^*] = \begin{bmatrix} |\alpha_0|^2 & \alpha_0\alpha_1^* & \cdots & \alpha_0\alpha_N^* \\ \alpha_1\alpha_0^* & |\alpha_1|^2 & \cdots & \alpha_1\alpha_N^* \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_N\alpha_0^* & \alpha_N\alpha_1^* & \cdots & |\alpha_N|^2 \end{bmatrix} \tag{1.3}$$

1.2.2 Mixed state

Mixed states are those that consist of a statistical ensembles of different quantum states. This means that, unlike pure states, mixed states cannot be represented as linear superpositions of normalized state vectors. In general, a mixed state consisting of an ensemble of n pure states can be expressed in the form of a list of outcome elements

$$\{|\Psi\rangle\}_{j=1}^n = \{|\Psi_1\rangle, |\Psi_2\rangle, ..., |\Psi_n\rangle\} \tag{1.4}$$

where each item has a corresponding probability of occurrence given by

$$\{p_j\}_{j=1}^n = \{p_1, p_2, ..., p_n\} \tag{1.5}$$

Here p_j correspond to the *classical* probability of the system being in state $|\Psi_j\rangle$, and the total number of possible states, n, need not be equal to the dimension of the underlying Hilbert space. A mixed state, consisting of several possible outcome pure state $|\Psi_j\rangle$, each with probability of occurrence p_j, is defined as a density matrix of the form

$$\rho \equiv \sum_j p_j |\Psi_j\rangle\langle\Psi_j| \tag{1.6}$$

It is easy to see that this general definition of the density matrix also holds for pure state, for which we will only have one $|\Psi_j\rangle$ term with $p_j = 1$. A very natural question to ask at this point is: how do mixed states evolve under unitary operations? It can be shown that if an initial arbitrary state $|\Psi_j\rangle$ with probability p_j evolves into the state $\hat{U}|\Psi_j\rangle$ after unitary evolution, then the evolution of a density matrix, consisting of an ensemble of normalized states $\{|\Psi_j\rangle\}_{j=1}^n$ with probabilities $\{p_j\}_{j=1}^n$ is given by

$$\rho = \sum_j p_j |\Psi_j\rangle\langle\Psi_j| \to \rho' = \sum_j p_j \hat{U}|\Psi_j\rangle\langle\Psi_j|\hat{U}^\dagger = \hat{U}\rho\hat{U}^\dagger \tag{1.7}$$

Here, $\hat{U}\dagger$ is the conjugate transpose of the operator $\hat{U}$.

1.2.3 Matrix elements

We understand that in the case of pure state in the state vector representation, each vector element corresponds to a probability amplitude. But what what do the elements of the density matrix represent? Consider, once again, a general mixed state ρ consisting of an ensemble of pure states $\{|\Psi_j\rangle\}_{j=1}^{n}$, each with probability of occurrence $\{p_j\}_{j=1}^{n}$:

$$\rho = \sum_j p_j \left|\Psi_j\right\rangle \left\langle\Psi_j\right| \tag{1.8}$$

We know that each individual pure state $|\Psi\rangle_j$ can be written as a linear superposition of elements forming a complete orthonormal basis $|\phi_u\rangle_{u=1}^{m}$:

$$\left|\Psi_j\right\rangle = \sum_u \alpha_u^{(j)} \left|\phi_u\right\rangle \tag{1.9}$$

We can then replace this expression into our definition of our general mixed state, and get a density matrix in terms of the orthonormal basis elements:

$$\rho = \sum_{u,v} (\sum_j p_j \alpha_u^{(j)} (\alpha_v^{(j)})^*) \left|\phi_u\right\rangle \left\langle\phi_v\right| = \sum_{u,v} \rho_{uv} \left|\phi_u\right\rangle \left\langle\phi_v\right| \tag{1.10}$$

where ρ_{uv} are the individual matrix elements in the $\{|\phi_u\rangle\}_{u=1}^{m}$ basis. Written in matrix form, ρ is given by:

$$\rho = \begin{bmatrix} \rho_{11} & \rho_{112} & \cdots & \rho_{1m} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{mm} \end{bmatrix} \tag{1.11}$$

It's worth noticing that, in ρ_{uv}, the diagonal terms ρ_{kk} actually correspond to the probability of finding the system in a particular basis state $|\phi_k\rangle$:

$$\rho_{kk} = \sum_j p_j \alpha_k^{(j)} (\alpha_k^{(j)})^* = \sum_j p_j \left|\alpha_k^{(j)}\right|^2 \tag{1.12}$$

Here, $\left|\alpha_k^{(j)}\right|^2$ corresponds to probability of finding the basis state $|\phi_k\rangle$ within a given $|\Psi_j\rangle$ state, so summing over all p_j values gives us the total probability

of the whole system being in state $|\phi_k\rangle$. On the other hand, the off-diagonal terms of the matrix are a measure of the *coherence* between the different basis states of the system. In other words, they can be used to quantify how a pure superposition state could evolve (*de-cohere*) into a mixed state. Lastly, it can be shown from the definition of ρ_{uv} that $\rho^*_{uv} = \rho_{uv}$, which basically means the density matrix is Hermitian.

1.2.4 Trace and Positivity

A density matrix that represents a valid pure or mixed state must satisfy two conditions:

- Its trace must be equal to one
- The matrix must be positive-semidefinite

Remembering that the trace of a matric (denoted as Tr) is the sum of its diagonal terms, we have:

$$Tr(\rho) = \sum_k \rho_{kk} = \sum_k \sum_j p_j \left|\alpha_k^{(j)}\right|^2 = \sum_j p_j \sum_k \left|\alpha_k^{(j)}\right|^2 = 1 \tag{1.13}$$

This follows as all basis states are normalized, and all probabilities must add to 1:

$$\sum_k \left|\alpha_k^{(j)}\right|^2 = 1 \; and \; \sum_j p_j = 1 \tag{1.14}$$

and for an arbitrary state $|\Phi_q\rangle$ that is part of the state space

$$\langle\Psi_q| \rho |\Psi_q\rangle = \sum_j p_j \langle\Psi_q|\Psi_j\rangle \langle\Psi_j|\Psi_q\rangle = \sum_j p_j |\langle\Psi_q|\Psi_j\rangle|^2 \geq 0 \tag{1.15}$$

A very useful property of the density matrix is that when taking the trace Tr of its square ρ^2, we obtain a scalar value γ that is good measure of the purity of the state the matrix represent. For normalized states, this value is always less than or equal to 1, with equality occurring for the case of pure state.

1.2.5 Reduced density matrix

Another advantage of working with the density matrix notation is that, when dealing with composite systems, it provides a practical way to extract the state of each subsystem, even if they are entangled. This is done in the form of what os known as the reduced density matrix. Consider a quantum system composed of subsystem A and B, and fully described by the density matrix ρ_{AB}. The reduced density matrix of subsystem A is then given by:

$$\rho_A = Tr_b(\rho_{AB}) \tag{1.16}$$

Here, Tr_B is an operation known as the partial trace, which is defined as:

$$Tr_B(|\xi_u\rangle \langle xi_v| \otimes |\chi_u\rangle \langle \chi_v|) \equiv |\xi_u\rangle \langle \xi_v| \, Tr(|\chi_u\rangle \langle \chi_v|) \tag{1.17}$$

$|xi_u\rangle$ and $|\xi_v\rangle$ are arbitrary state in the subspace of A, and $|\chi_u\rangle$ and $|\chi_v\rangle$ arbitrary states in the subspace of B. Tr is the standard trace operation, which for two arbitrary states $Tr(|\chi_u\rangle \langle \chi_v|) = \langle \chi_v|\chi_u\rangle$. Similarly, we can calculate the reduced density matrix of subsystem B using the partial trace over A:

$$Tr_A(|\xi_u\rangle \langle \xi_v| \otimes |\chi_u\rangle \langle \chi_v|) \equiv Tr(|\xi_u\rangle \langle \xi_v|) \, |\chi_u\rangle \langle \chi_v| \tag{1.18}$$

1.3 Qubit and quantum circuit model

A quantum bit(*qubit*) is a two-level quantum system, described by a two-dimensional complex Hilbert space. In this space, one may choose a pair of normalized and mutually orthogonal quantum states,

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{1.19}$$

to represent the values 0 and 1 of a classical bit. These two states form a *computational basis*. From the superposition principle, any state of the qubit may be written as

$$|\Psi\rangle = \alpha \, |0\rangle + \beta \, |1\rangle \, , \tag{1.20}$$

where the amplitudes α and β are complex numbers, constrained by the normalization condition

$$|\alpha|^2 + |\beta|^2 = 1. \tag{1.21}$$

Since state vectors are defined only up to a global phase of no physical significance, one may choose α real and positive (except for the basis state $|1\rangle$, in which $\alpha = 0$, and one may take $\beta = 1$). Thus, the generic state of a qubit may be written as

$$|\Psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle = \begin{bmatrix}\cos\frac{\theta}{2}\\ e^{i\phi}\sin\frac{\theta}{2}\end{bmatrix} \tag{1.22}$$

with $(0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi)$. Therefore, unlike the classical bit, which can only be set equal to 0 or 1, the qubit resides in a vector space, parametrized by the continuous variables α and β (or θ and ϕ). Thus, a continuum of stats is allowed. However, there is a catch: to extract this information we must perform a measurement and quantum mechanics tells us that from the measurement of the polarization state σ_n of a qubit along any axis n, we obtain only a single bit of information ($\sigma_n = +1$ or $\sigma_n = -1$).

1.3.1 Bloch sphere

The Bloch sphere representation is useful in thinking about qubits since it provides a geometric picture of the qubit and of the transformations that one can operate on the state of a qubit. Owing to the normalization condition $(|\alpha|^2 + |\beta|^2 = 1$, the qubit's state can be represented by a point on a sphere of unit radius, called the *Bloch sphere*. This sphere can be embedded in a three-dimensional space of Cartesian coordinates ($x = \cos\phi\sin\theta$, $y = \sin\phi\sin\theta$, $z = \cos\theta$). Thus, the state $\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle$ can be written as

$$|\Psi\rangle = \begin{bmatrix}\sqrt{\frac{1+z}{2}}\\ \frac{x+iy}{\sqrt{2(1+z)}}\end{bmatrix}. \tag{1.23}$$

By definition, a Bloch vector is a vector whose components (x, y, z) single out a point on the Bloch sphere. Therefore, each Bloch vector must satisfy the normalization condition $x^2 + y^2 + z^2 = 1$.

1.3.2 Bloch ball

The above construct can be generalized by applying the density-operator formalism. There is a one-to-one correspondance between the density matrices

for a qubit and the points on the unit ball $0 \leq |\mathbf{r}| \leq 1$, which is known as the Bloch ball. The vector $\mathbf{r}$ is known as the Bloch vector for a generic state of the qubit. In the case of a pure state, the density matrix has his determinant equal to 0, which in turn implies $|\mathbf{r}| = 1$. One can conclude that pure states are located on the boundary of the Bloch ball.

1.3.3 Circuit model of quantum computation

As seen above we defined a computation as the implementation, by a suitable configuration of Boolean logic gates, of the mapping $f : \{0,1\}^n \to \{0,1\}^m$. In the circuit model, the n input wires are place-holders for members of the set $\{0,1\}^n$ and in a typical electronic circuit, the diagrammatic *wire* represents a physical wire whose state or bit assignment is determined by a voltage presence. In a quantum circuit, each wire lead represent a qubit ket. We assume that the "wire" in a quantum circuit diagram represents a qubit in a definite quantum state. That is not altered until the wire enters a quantum gate or a measurement device. By convention, time is assumed to run from left to right in the diagram, and the wire lead on the far left-hand side of a diagram denotes the initial qubit state. The qubit state is processed by a quantum gate and the wire lead exiting the gate from its right represents the output state of that qubit. Typically, quantum gates are shown as labeled boxes or solid and empty nodes. The initial qubit ket could be in $|0\rangle$ or $|1\rangle$ bases state, or in a linear combination of the two. Because Hilbert space operators induce maps between vectors in Hilbert space, they serve as quantum analogs of classical gates. One crucial difference between a classical Boolean logic gate and a quantum logic gate is that the ladder is *reversible*. Because $adj(U) = U^{-1}$ unitary operators are quantum logic candidates. Notably the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ are unitary operators and serve as elementary quantum gates. The Hadamard quantum logic gate is widely used in quantum circuits; its representation is

$$\mathbf{H} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{1.24}$$

With the computational basis $|0\rangle, |1\rangle$ as inputs, the Hadamard gate outputs the linear superposition states

$$\begin{aligned} |-\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{aligned} \tag{1.25}$$

Now we generalize the quantum gate concept for application in multi-qubit circuits. The computational basis states for a two-qubit system are $|00\rangle$,$|01\rangle$,$|10\rangle$,$|11\rangle$. Quantum gates are operators in this Hilbert space and, as is the case for a single qubit, they must be unitary. It can be shown that

$$\begin{aligned} (\mathbf{U_a} \otimes \mathbf{U_b})^\dagger &= \mathbf{U_a}^\dagger \otimes \mathbf{U_b}^\dagger \\ (\mathbf{U_a} \otimes \mathbf{U_b})^\dagger (\mathbf{U_a} \otimes \mathbf{U_b}) &= (\mathbf{U_a}^\dagger \mathbf{U_a}) \otimes (\mathbf{U_b}^\dagger \mathbf{U_b}) = \mathbf{I} \otimes \mathbf{I} \end{aligned} \tag{1.26}$$

Therefore

$$\mathbf{I} \otimes \sigma_\mathbf{x}, \sigma_\mathbf{x} \otimes \sigma_z, \mathbf{I} \otimes \mathbf{I} \tag{1.27}$$

are just a few possible two-qubit quantum gate candidates. Some multi-qubit gates cannot be expressed as the direct product of qubits. Of these, one of the most important is the controlled-not, or *CNOT* gate. The state of the first qubit, x_c, is not altered under the action of the CNOT gate (control qubit). If the state $x_c = 0$, the state in the lower wire is also unaffected, and the gate acts as an identity operator. However if the state of the first qubit $x_c = 1$, the lower bit *flips* as in a classical NOT gate. Symbolically, the gate, denoted by operator U_C, is described by rule $U_C |x_1 x_0\rangle \rightarrow |y_1 y_0\rangle$, where $y_1 = x_1$ and $y_0 = x_0 \oplus x_1$. $\mathbf{U_C}$ can also be expressed as $\mathbf{U_C} = |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10| + |11\rangle\langle 11|$ or

$$\mathbf{U_C} = \frac{1}{2}\mathbf{I} \otimes \mathbf{I} + \frac{1}{2}\sigma_\mathbf{z} \otimes \mathbf{I} + \frac{1}{2}\mathbf{I} \otimes \sigma_\mathbf{x} - \frac{1}{2}\sigma_\mathbf{z} \otimes \sigma_\mathbf{x} = \mathbf{I}_4 \tag{1.28}$$

Therefore, the CNOT gate is a sum of direct product operators that cannot be factored, in fact the result of the CNOT is that the two-qubit system is in an entangled state. The usefulness of the circuit model in classical computation is due to the fact that a sequence of elementary operations (NOT, OR, AND,...) allows one to build up arbitrarily complex computations. A similar result exists for quantum computation, that is, any unitary operation in the Hilbert space of n qubits can be decomposed into one-qubit and two-qubit CNOT gates.

Chapter 2

VQE algorithm

2.1 A description of the VQE algorithm

The VQE is grounded in the variational principle (and more precisely in the Rayliegh-Ritz functional), which optimizes an upper bond for the lowest possible expectation value of an observable with respect to a trial wavefunction. Namely, providing a Hamiltonian $\hat{H}$, and a trial wavefunction $|\Psi\rangle$, the ground state energy associated with this Hamiltonian, E_0, is bounded by:

$$E_0 \leq \frac{\langle\Psi|\,\hat{H}\,|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{2.1}$$

The objective of the VQE is therefore to find a parametrization of $|\Psi\rangle$, such that the expectation value of the Hamiltonian is minimized. This expectation value forms an upper bound for the ground state energy, and in an ideal case should be indistinguishable from it to the level of accuracy desired. In mathematical terms, we aim to find an approximation to the eigenvector $|\Psi\rangle$ of the Hermitian operator $\hat{H}$ corresponding to the lowest eigenvalue, E_0. In order to translate this minimization task into a problem that can be executed on a quantum computer, one must start by defining a so-called *ansatz* wavefunction that can be implemented on a quantum device as a series of quantum gates. Given that we can only perform unitary operations or measurements on a quantum computer, we do this by using parametrized unitary operations. We hence express $|\Psi\rangle$ as the application of a generic parametrized unitary $U(\theta)$ to an initial state for N qubits, with θ denoting a set of parameters taking values in $]-\pi, pi]$. The qubit register is generally initialized as $|0\rangle^{\otimes N}$, written as $|0\rangle$ for simplicity, although low-depth

operations can be performed for alternative initializations before the unitary is applied. Noting that $|\Psi\rangle$ (as well as any $U(\theta)|\Psi\rangle$) is necessarily be a normalized wavefunction, we can now write the VQE optimization problem:

$$E_{VQE} = \min_\theta \langle 0| U^\dagger(\theta)\hat{H}U(\theta)|0\rangle \tag{2.2}$$

This equation is also referred to as the cost function of the VQE optimization problem, which is a term adopted from the machine learning and optimization literature. We can continue this description by writing the Hamiltonian in a form that is directly measurable on a quantum computer, as a weighted sum of spin operators. Observables suitable for direct measurement on a quantum device are tensor products of spin operators (Pauli operators) on the different qubits. We can define these as Pauli string: $\hat{P}_a \in \{I, X, Y, Z\}^{\otimes N}$, with N the number of qubits used to model the wavefunction. The Hamiltonian can be rewritten as

$$\hat{H} = \sum_a^P \omega_a \hat{P}_a \tag{2.3}$$

with ω_a a set of weights, and P the number of Pauli strings in the Hamiltonian. We can now rewrite E_{VQE} as

$$E_{VQE} = \min_\theta \sum_a^P \omega_a \langle 0| U^\dagger(\theta)\hat{P}_a U(\theta)|0\rangle \tag{2.4}$$

where the hybrid nature of the VQE becomes clearly apparent: each term $E_{P_a} = \langle 0| U^\dagger(\theta)\hat{P}_a U(\theta)|0\rangle$ corresponds to the expectation value of a Pauli string $\hat{P}_a$ and can be computed on a quantum device, while the summation and minimization $E_{VQE} = \min_\theta \sum_a^P \omega_a$ is computed on a conventional computer.

2.1.1 VQE stack

The VQE, as presented above, can be decomposed into a number of components, which all involve a number of significant choices which impact the design and overall cost of the algorithm. We refer to the layering of these different components as the VQE stack. Most choices made on specific elements of this stack have significant implications on the entire VQE process. We summarize the key components below and provide a brief introduction to each of them and how they fit together:

- *Hamiltonian construction and representation*: the first step in the VQE is to define the system for which we want to find the ground state. This is in general an *ab initio* molecular, a solid-state system, or a spin lattice model. For each of these, one starts with a specific geometry (or conformation) of the system, specifying the distance between each atom, or the geometry of the lattice. Constructing the Hamiltonian defines the quantum observable for the individual single-particle degree of freedom. Given the Hamiltonian defines the quantum observable for the total energy associated with an electronic wavefunction, the choice of basis is critical to define the space spanned by the wavefunction. Furthermore, the ground state energy is in principle invariant to a single-particle unitary transformations between these basis functions in which the system Hamiltonian is represented. This allows these degree of freedom defining the basis to be unitarily rotated amongst each other, with the total energy remaining invariant to this operation. These different representations could include, as examples, molecular orbitals from prior mean-field calculation, plane-wave functions, or local atomic functions, all representing the spatial distribution (or 'orbitals') for the single-particle Fock states, from which the many-body basis is formed. The choice of basis can have a significant impact on the accuracy and cost of the final result, as the type of basis and number of basis functions chosen both determine the size of the computation required and the accuracy of the representation. Following the Pauli exclusion principle the electronic wavefunction must be antisymmetric with respect to the exchange of two electrons. From a mathematical perspective, this means that we must decide whether we enforce this antisymmetry through the definition of the wavefunction or through the definition of the operators. These are referred to (for historical reasons) respectively as first and second quantization. In second quantization the Hamiltonian is expressed in terms of fermionic operators, also known as creation ($\hat{a_j^\dagger}$) and annihilation ($\hat{a_j}$) operators. These correspond to the action of adding, or removing an electron from a given basis function with integer index j, respectively (e.g. an orbital or a lattice site), ensuring appropriate fermionic antisymmetry with respect to permutation of any two particles.

- *Encoding of operators*: qubit registers on quantum computers can only measure observables expressed in a Pauli basis, due to the two-level

nature of spins: $\hat{P}_a \in \{I, X, Y, Z\}^{\otimes N}$, for N qubits. In first quantization operators can be directly translated into spin operators that can be measured on quantum computers, as they are not used to enforce antisymmetry of the wavefunction. In second quantization the Hamiltonian is expressed as a linear combination of fermionic operators which are defined to obey this antisymmetry relationship, unlike Pauli operators. The role of a fermionic to spin encoding is therefore to construct observables, from Pauli operators, which maintain this relationship. A transformation of fermionic operators to spin operators that meets this criterion was demonstrated a long time ago, and recent research has focused on improving on this initial work. The key factors determining the efficiency of an encoding are their Pauli weight (the maximum number of non-identity elements in a given spin operator). It is worth noting that for certain ansatz choices, in particular those defined in terms of fermionic operators, the encoding can have significant implications on gate depth and trainability.

- *Measurement strategy and grouping*: The next step in the VQE stack is to determine how measurements are distributed and organized to efficiently extract the required expectation values from the trial wave function. In general, to achieve a precision of ϵ on the expectation value of an operator, we are required to perform $O(\frac{1}{\epsilon^2})$ repetitions (usually denoted as shots) of the circuit execution, each completed with a measurement at the end. The objective of the measurement strategy is to make the number of repetitions as low as possible. Several techniques are available to achieve this, in particular, the use of efficient weighting of the number of measurements across the operators. This can be further optimized by using properties of the Lie algebra in which Pauli strings are defined. Via processing of the encoded Pauli strings to measure, it is possible to identify commuting groups of operators that can be measured jointly, and subsequently find the measurement bases in which all operators of a given group can be simultaneously measured. In order to perform this joint measurement, a short quantum circuit must therefore be designed and applied for each group, to rotate the measurement basis and to perform this joint measurement. Alternatively, because of information overlap between different Pauli strings, one can also try to reduce the number of measurements required using inference methods from fewer shots.

- *Ansatz and state preparation*: once the Hamiltonian has been prepared such that its expectation value can be measured on a quantum device, we can turn to the preparation of the trial wavefunction. In order to do this, one must decide on a structure for the parametrized quantum circuit, denoted as ansatz. It is used to produce the trial state, with which the Hamiltonian can be measured. Upon successful optimization of the ansatz parameters, the trial state becomes a model for the ground state wavefunction of the system studied. A wide range of ansatz are possible, and the appropriate choice depends on the problem beng addressed. The key aspects of the ansatz are its expressibility and trainability. The expressibility defines the ability of the ansatz to span a large class of states in the Hilbert space, defining the maximum accuracy its approximation of relevant low-energy states can achieve (assuming all parameters can be perfectly optimized). Its trainability describes the practical ability of the ansatz to be optimized using techniques tractable on quantum devices (related to the total number of parameters, their linear dependence, the structure of the optimization surface, and to the related concept of barren plateaus, which can arise when gradients almost vanish thereby preventing optimization). A good ansatz must be sufficiently expressive to guarantee that it can appropriately approximate the ground state wavefunction, however, it must not be so expressive that it renders the search for the target state intractable. Another important aspect of the ansatz choice is the scaling and complexity of its circuit depth with system size. This is particularly important for near-term application of the VQE, as it determines in great part the noise resilience of the method employed.

- *Parameter optimization*: The parameters of the ansatz used need to be updated iteratively until convergence. In general, this requires sampling the expectation value of the Hamiltonian several times for a given parameter set in the ansatz in order to define an update rule for the parameters (i.e. the updated value of the parameters is a function of the expectation value measured). The choice of optimization is critical for at least three main reasons:

  - it directly impacts the number of measurements required to complete an optimization step, as e.g. computing the numerical gradient of a quantum circuit can require value estimation of the

Hamiltonian with respect to several slightly modified wavefunctions

  - certain optimizers have been designed to alleviate specific optimization issues, such as the barren plateau problem
  - it directly impacts the number of iterations required to reach convergence

- *Error mitigation*: quantum noise is one of the main hurdles in the viability of the VQE, given that the method is to be used without error correction schemes on NISQ devices. Error mitigation aims to reduce the impact of quantum noise through post-processing of the measurement data (or occasionally through post-processing of the trial wavefunction ahead of measurements).

There are many restrictions of quantum computing that this approach does not take into account, and they suggested two more stringent conditions. The first one is that VQE must demonstrate similar or higher accuracy than any conventional method, but with lower computational time-to-solution. This condition takes into account possible limitation due to hardware runtime, potentially resulting in a large pre-factor for VQE computation. The pre-factor refers to the multiplier applied to a scaling rule to obtain the actual runtime of the method. If the VQE has better asymptotic scaling than conventional method, but a large pre-factor, this means an advantage could only be achieved in the asymptotic regime of very large systems. This would make it difficult to demonstrate quantum advantage for practical moderately sized system. The second condition exposed, which is also the most stringent form of quantum advantage for the VQE, is to achieve at least as good accuracy, and with faster compute time, for a system of sufficient complexity to accurately model a real problem of physical and chemical relevance. This involves demonstrations on systems, where the approximation error in defining the specific Hamiltonian for the original problem is smaller magnitude than its solution using the VQE.

One of the key challenges possibly holding back VQE is the very large amount of samples that are required to accurately compute the relevant values of the algorithm. There are two main aspects to manage for efficiently sampling these expectation values: the number of terms in the Hamiltonian cost functions, the number of shots required to sample an expectation value

at a certain level of accuracy. It is worth noting that the level of accuracy required changes throughout the optimization process. This could be because optimizer gradients must be estimated precisely enough to be distinguished from one another when the optimization landscape flattens.

2.1.2 Shot numbers in VQE

Generalized mappings for molecular Hamiltonian result in $P \sim O(n^4)$ distinct Pauli strings to estimate. With this in mind, let us consider the number of shots required to achieve a given precision. In any sampling experiment, the standard error is equal to $\epsilon = \frac{\sigma}{\sqrt{S}}$, where σ is the population standard deviation, and S is the experimental sample size, in our case, the number of shots. This means that the number of times an experiment needs to be repeated to achieve a given expected error ϵ goes as $O(\frac{1}{\epsilon^2}$. More specifically, when measurements are distributed optimally among the different Pauli strings, such that the variance is minimized with respect to a given precision ϵ, the number of measurements required is upper-bounded by

$$S \leq \left(\frac{\sum_a^P \omega_a}{\epsilon} \right)^2 \tag{2.5}$$

where ω_a are the weights of the Pauli strings in the Hamiltonian. As a result, for a given level of accuracy for each Pauli string measured independently, the overall scaling of the number of shots required for an energy estimation is $O(\frac{N^4}{\epsilon^2}$. In the context of quantum chemistry, successful computing methods are expected to produce results within a precision of $\epsilon = 1.6mE_H$ to the target. When results obtained numerically are within this level of precision to experimental results, the simulation is deemed to reach chemical accuracy. This metric can be used as a bound for target precision in the VQE context. One should be cautious however not to assume too much of a relationship between this number and the number of shots required to perform VQE. That is because the key bottleneck of VQE optimization is not the estimatin of the wavefunction itself but the estimation of gradients and in particular the difference between these gradients. This difference may be orders of magnitude smaller than the chemical precision threshold, requiring many more measurements. While polynomial in scaling, it has been pointed out on several occasions that the number of shots required to accurately compute a VQE optimization process rapidly becomes unmanageable, suggesting the method

might be unable to compete with its conventional computing counterparts. As such, a significant amount of effort has been devoted to finding solutions that reduce the pre-factor for the number of shots required.

2.1.3 Weighted distribution of measurements

In this methods our aim is to focus on measuring more precisely the operators which contribute most to the total variance of the expectation value estimated. With a given shot budget, one can improve the overall precision of measurement by distributing these shots towards specific operators. A straightforward manner to distribute these shots is to simply weight them respect to the Pauli string weight ($|\omega_A|$) in the Hamiltonian. If we look at the equation:

$$S = P \sum_{a}^{P} \frac{\omega_a^2 \mathrm{Var}[\hat{P}_a]}{\epsilon^2} \tag{2.6}$$

we can see that reducing the number of shots on strings contributing less to the total energy estimate (with lower $|\omega_a|$ value) and adding these to strings that contribute the most reduces total variance, as long as $\mathrm{Var}[\hat{P}_a]$ are similar for all. It can be shown that $S_a \propto |\omega_a| \sqrt{\mathrm{Var}[\hat{P}_a]}$, is optimal, and it is possible to demonstrate numerically that when considering random states, variations in $|\omega_a|$ tend to be higher than variation in $\mathrm{Var}[\hat{P}_a]$, resulting in the weight pro-rata distribution (where $S_a \propto |\omega_a|$) of measurements outperforming the uniform distribution in most cases. To address cases in which the number of shots is limited, it is proposed to perform measurements on Pauli strings randomly, with probabilities proportional to $|\omega_a| \sqrt{\mathrm{Var}[\hat{P}_a]}$, thereby allowing unbiased estimates even with a low total shot number.

2.1.4 Term truncation

Another approach to consider is to remove from measurement scope terms that have contributions significantly below the error tolerance threshold ϵ. This method has been shown to significantly reduce the cost of quantum chemistry calculations with negligible impact on accuracy. To implement this, one must observe that the contribution of any Pauli observable to the final energy estimate is bounded by the absolute value of its associated weight: $\left| \langle \Psi | \, \omega_a \hat{P}_a \, | \Psi \rangle \right| \leq |\omega_a|$. By ordering these contributions in ascending order,

one can construct a partial sum of the $k \leq P$ smallest contributor:

$$e_k = \sum_{a}^{k} |\omega_a| \tag{2.7}$$

From there, one can choose a constant $C \in [0, 1[$, and include in the partial sum the terms up to an index k that verify: $e_k \leq C_\epsilon$. This method introduces a bias in total energy estimation, and as such the key to implement successfully is to pick a constant C such that truncation bias is lower than the mean square error reduction from measurements added to the remaining terms. McClean et al. [4] present an adjusted estimate for the number of shots required to achieve ϵ:

$$S = (P - k) \sum_{a=k+1}^{P} \frac{\omega^2 \mathrm{Var}[\hat{P}_a]}{(1 - C^2)\epsilon^2} \tag{2.8}$$

If the expected number of shots is lower than without term truncation, then the method provides an improvement regarding the precision to measurement cost ratio.

2.2 Ansatz

Ansatz selection is a central part of the VQE stack. Choosing the right ansatz is essential to guarantee that the final solution is close to the global minimum. To achieve this, it is essential to maximize the chances of the ansatz covering parts of the Hilbert space that contain the solution (i.e. a state that is sufficiently close to the desired state which globally minimizes the expectation value of the Hamiltonian). The span of possible states an ansatz can reach is referred to as its expressibility. However, optimizing a general state could easily become intractable due to the number of parameters, the number of iterations required for convergence, or the number of shots required to achieve sufficient gradient accuracy to continue the optimization. Whether an ansatz can be optimized in a tractable manner is referred to as its trainability. In practice, it is better to choose an ansatz spanning a smaller subspace, but remaining trainable. Designing an ansatz for a given number of qubits hence involves finding an optimal trade-off between expressibility and trainability. The expressibility of an ansatz describes the uniformity of its span across the

unitary space. One can measure the expressibility of an ansatz by assessing the distance between the distributions of the unitaries that can be generated by the said ansatz, and maximally uniform distribution of unitaries in the corresponding Hilbert space, also known s the Haar measure. A given ansatz is called a t-design if it is indistinguishable from the Haar measure up to the t^{th} moment. A 2-design ansatz can produce any possible state in the Hilbert space considered, from any input state: it is maximally expressive. More formally, one can define as $\mathbb{U}$ the set of unitaries accessible by an ansatz, and $\mathbf{U}(N)$ the complete unitary group in which the ansatz is expressed (with N the number of qubits it spans), such that $\mathbb{U} \subseteq \mathbf{U}(N)$. The following super operator, representing the second order difference between the Haar measure on $\mathbf{U}(N)$ and the uniformity distribution of $\mathbb{U}$ can be constructed:

$$\mathbf{A}_{\mathbb{U}}(\cdot) := \int_{\mathbf{U}(n)} d_{\mu}(V) V^{\otimes 2}(\cdot)(V^{\dagger})^{\otimes 2} - \int_{\mathbb{U}} dU U^{\otimes 2}(\cdot)(U^{\dagger})^{\otimes 2} \tag{2.9}$$

with $d_{\mu}(V)$ the volume element of the Haar measure, and dU the uniform distribution over $\mathbb{U}$, $V \in \mathbf{U}_N$ and $U \in \mathbb{U}$. If $\mathbf{A}_{\mathbb{U}}(\hat{O}) \to 0$, then the ansatz producing $\mathbb{U}$ approaches a 2-design and offers therefore maximal expressibility. From this super-operator, one can compute a metric for expressibility of an ansatz as

$$\begin{aligned} \varepsilon_{\mathbb{U}}^{\rho} &:= \left\| \mathbf{A}_{\mathbb{U}}(\rho^{\otimes 2}) \right\|_2 \\ \varepsilon_{\mathbb{U}}^{\hat{P}} &:= \left\| \mathbf{A}_{\mathbb{U}}(\hat{P}^{\otimes 2}) \right\|_2 \end{aligned} \tag{2.10}$$

As such, the expressibility of an ansatz can be expressed with respect to an initial input state (ρ in the formulas above), or with respect to a measurement operator ($\hat{P}$). Following the equations above, one can interpret that if $\varepsilon = 0$ the ansatz is maximally expressive, while expressibility decreases as ε increases. the trainability of an ansatz refers to the ability to find the best set of parameters of the ansatz by (iteratively) optimizing the ansatz with respect to expectation values of the Hamiltonian in a tractable time. More specifically, an ansatz is considered trainable if its expected gradient vanishes at most polynomially as a function of the problem's features (e.g. system size, circuit depth). On the other hand, if the gradient vanishes exponentially, it is said to suffer from the barren plateau problem.

2.2.1 Barren plateau problem

A key issue that is inherent to all types of variational quantum algorithms is the risk of vanishing gradients, either during training or as a result of a random initialization. This refers to the risk of the expected cost function gradients vanishing exponentially as a function of specific properties of the optimization problems. McClean et al.[4] provide the first formal characterization of this barren plateau problem, and show that cost function gradients are vanishing exponentially in the number of qubits in the quantum register, providing random initialization of the circuit parameters. Even though this problem is akin to the vanishing gradient problem in machine learning, it has two striking differences that make it significantly more impactful on the prospects of variational quantum algorithms:

- The estimation of the gradients on a quantum device is essentially stochastic. Any observable can only be measured to a certain precision, increasing as the inverse square root of the number of shots. If gradients are exponentially approaching zero, it means that distinguishing between a positive and a negative gradient becomes increasingly difficult. Failing to establish the sign of the gradient reliably transforms the optimization into a random walk, overall requiring ana exponential number of shots to continue optimization.

- The barren plateau problem is dependent on the number of qubits (while the problem is dependent on the number of layers for the vanishing gradient problem). Additional research also shows that it can be linked to other factors specific to quantum circuits, including expressibility of the ansatz, degree of entanglement of the wavefunction, non-locality of the wavefunction, or quantum noise.

It is worth to briefly discuss the typical cost function landscape for single parameters in the variational quantum eigensolver. Another problem that affects this landscape is that of 'narrow gorges'. It refers to the fact that the local minimum (well definied by the region starting from the end of a barren plateau and going towards a local minimum) contracts exponentially in the number of qubits. Interestingly, it was shown that these two problems are equivalent. An alternative way to present the barren plateau problem is that it implies the expectation value of an observable with respect to a random state concentrates exponentially around the mean value of that obsevable,

rendering intractable optimization away from the mean. In the context of the VQE, the barren plateau problem can be formally characterized as follows. Consider a VQE optimization problem with cost function:

$$E(\theta) = \langle \Psi(\theta) | \hat{H} | \Psi(\theta) \rangle \tag{2.11}$$

with $\hat{H}$ the molecular Hamiltonian operator, and $|\Psi(\theta)\rangle$ the parametrized wave function with a vector θ of parameters. This cost function exhibits a barren plateau if, for any $\theta_i \in \theta$ and for any $\epsilon > 0$ there is $b > 1$ such that:

$$\Pr(|\partial_{\theta_i} E(\theta)| \geq \epsilon) \leq O\left(\frac{1}{b^N}\right) \tag{2.12}$$

which is an immediate consequence of Chebyshev's inequality and the result from the above (for the expectation value and variance). This means that the probability of a gradient being above a certain threshold (which could be arbitrarily small), can always be upper-bound by a number that decreases exponentially in the system size n. It is however important to note that while defined with respect to a cost gradient, the barren plateau problem also affects gradient-free optimizers. It is easy to understand, as gradient-free optimizers usually rely on sampling the cost landscape of a specific parameters, if variance across the landscape is minuscule, then it becomes impossible to accurately progress through the optimization step.

2.2.2 The unitary coupled cluster (UCC) ansatz

The unitary coupled cluster ansatz is arguably the most studied ansatz for VQE. It figures in the initial VQE work by Peruzzo et al. [1] and has taken an important part in the literature since then. The UCC theory stems from adapting the Coupled Cluster (CC) theory. CC is a post Hartree-Fock method that aims at recovering a portion of electron correlation energy by evolving an initial wavefunction (usually the Hartree-Fock wavefunction) under the action of parametrized excitation operators. in general, this are single-electron excitations and double electron excitations CC Single and Double (CCSD), however, these can sometimes go to higher-order (e.g. UCCSDT). Only excitation operators allowing transitions from an occupied orbital to an unoccupied orbital are traditionally included in CC. The action of these operators on the initial state is performed through exponentiation

of part of the cluster operator T. For ν, the maximum allowed excitation we have:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_\nu \tag{2.13}$$

with for example the single and double excitation operators:

$$\begin{aligned} \hat{T}_1 &= \sum_{ia} t_i^a \hat{a_a^\dagger} \hat{a_i} \\ \hat{T}_2 &= \sum_{ijab} t_{ij}^{ab} \hat{a_a^\dagger} \hat{a_b^\dagger} \hat{a_j} \hat{a_i} \end{aligned} \tag{2.14}$$

Using the Hartree-Fock state as reference state, the CC ansatz wavefunction is given by

$$|\Psi\rangle = e^{\hat{T}} |\Psi_{HF}\rangle \tag{2.15}$$

The conventional resolution method for CC scales $O(m^2(N-m)^4)$, with m the number of electrons, and N the number of spin orbitals. CCSD is however in general not variational and has been reported to fail in numerous cases, in particular in system with strong correlation, with possible solutions to avoid these failures usually scaling exponentially in the system size. Another issue worth mentioning in the context of th VQE is that the operator $e^{\hat{T}}$ is not unitary, and therefore the CC ansatz cannot be implemented as a series of quantum gates. The UCC method was developed as a way to address these caveats. It is based on the fact that for any linear operator $\hat{T}$, the expression ($\hat{T} - \hat{T}^\dagger$ is an anti-Hermitian operator. The exponential of an anti-Hermitian operator is a unitary operator, and the difference between the cluster operator and its complex conjugate can be used as an evolution operator to form a unitary version of CC. Elements of the truncated cluster operator for UCC are identical as for CC

$$|\Psi\rangle = e^{\hat{T} - \hat{T}^\dagger} |\Psi_{HF}\rangle \tag{2.16}$$

The energy can then be evaluated using a variational approach based on the Ritz functional 2. Exact resolution of UCC is known to scale exponentially in the system size. It is therefore natural to bring this ansatz to Quantum Computation. By grouping the excitation terms in $\hat{T}$ with their corresponding conjugate in $\hat{T}^\dagger$, and noting them τ, we can obtain the parametrized version of the UCC:

$$U(\overrightarrow{t}) = e^{\sum_j t_j (\tau_j - \tau_j^\dagger)} \tag{2.17}$$

where t_j correspond to the amplitude weights, and with j spanning all the excitation operators included. Akin to our description of CC, UCC can accept several level of excitations (UCCSD, UCCSDT). The next step is to convert the UCC ansatz into a series of parametrized quantum gates which can directly be implemented on a Quantum Computer. A usual step in the process of building the UCC ansatz is to use a Trotter-Suzuki approximation to separate the summation into a product of fermionic terms exponentials such that

$$U(\overrightarrow{t}) \sim U_{Trotter}(\overrightarrow{t}) = (\prod e^{\frac{t_j}{\rho}(\tau_j - \tau_j^\dagger)})^\rho \tag{2.18}$$

The Trotter number ρ defines the precision of the approximation, but also the pre-factor impacting the overall depth of the quantum circuit required. Using the mapping described before we can re-write each of the fermionic operators as

$$\tau_j - \tau_j^\dagger = i \sum_k \hat{P_{k,j}} \tag{2.19}$$

where the subterm $\hat{P_{k,j}}$ is a product of Pauli operators. This conversion allow to re-write ***, ssuming $\rho = 1$, into

$$U_1(\overrightarrow{t}) = \prod_j \prod_k^{2^{2l}k^{-a}} e^{it_j \hat{P_k^j}} \tag{2.20}$$

This illustrates the basic framework surrounding UCC and its application to VQE. The required circuit depth for this version of UCC has been shown to scale polynomially in the system size. More specifically, with η the number of electrons and N the number of spin-orbitals, the circuit depth is shown to scale $O((N-m)^2 m)$ for each Trotter step. It is worth noting that, it was numerically shown that a single Trotter step is sufficient for accurate description of the ground state in simple molecular system, because variational optimization can absorb some of the Trotterization error. The UCC ansatz can also be used in the context of restricted active space methods. Finally, it has been shown that operators in the UCC ansatz contribute to the accuracy of the ansatz a different extend, and that some can be discarded to increase efficiency.

2.3 Optimization strategy

The VQE being in essence an optimization problem, it aims at heuristically constructing an approximation of an electronic wavefunction through iterative learning of ansatz parameters. For the algorithm to be viable, it must be that it can learn a good enough approximation within a tractable number of learning steps. It was already demonstrated that optimization of the variational quantum ansatz is NP-hard, meaning that there exist at least some problems in which finding an exact solution for the VQE problem is intractable. As such, efficient optimization strategies that provide a well-approximated solution within an acceptable number of iterations are essential for any variational algorithm to be put into practice. Compared to the conventional numerical optimization problem, however, optimizing the expectation value from a variational quantum ansatz faces additional challenges:

- sampling noise and gate noise on NISQ devices disturb the landscape of the objective function. Such noise can be detrimental to the convergence of optimization, and could limit the scope for quantum advantage
- While the precision of conventional numerical optimization is generally not consider a problem, the precision of the measured expectation value is limited by the sample shot number. The cost of optimization is heavily dependent on the precision required for optimization
- Related to the point above, the landscape of the expectation value of variational ansatz may cause the vanishing of gradient very easily as a result of the barren plateau problem.

There is also positive news for optimizing variational ansatz parameters. Studies from recent years show the landscape of expectation value has some analytical properties that are useful to extract information, such as evaluating gradient directly on quantum devices. In addition, the ansatz landscape can be efficiently approximated to accelerate the convergence. Utilizing such prior knowledge helps to develop efficient optimization strategies for variational quantum algorithms. The objective function of a variational algorithm is constructed conventionally based on the measurement outcome. Denote $\mathbb{O}(\theta) = (\hat{O}_1(\theta^{(1)}), \hat{O}_2(\theta^{(2)}), \ldots, \hat{O}_a(\theta^{(a)}))$ are the observables used to compose the objective function and a is the number of observables. The objective

function is given by

$$\mathbf{L}(\theta) = C(\mathbb{O}(\theta)) \tag{2.21}$$

where C is a function maps the observed expectation value to the objective function, and usually have the simple linear form

$$C(\mathbf{X}) = \sum_i c_i X_i \tag{2.22}$$

where c_i is constant value defined by the problem as the coefficient of each measurement expectation value, X_i is the i-th component of $\mathbf{X}$. Such linear form preserves the analytical properties and it is essential for using the analytical methods to directly calculate the gradient or implement analytical gradient free optimization strategy. A measurement expectation value is given by

$$< \hat{O}_k(\theta^{(k)}) >= \langle \Psi_0 | \, U^{(k)\dagger}(\theta^{(k)}) \hat{M}_k U^{(k)}(\theta^{(k)}) \, | \Psi_0 \rangle \tag{2.23}$$

where $|\Psi_0\rangle$ is the initial state on the quantum computer. $\hat{M}^{(k)}$ is a Hermitian measurement operator, usually chosen to be the tensor product of Pauli operators to match the physical measurement implementation of quantum hardwares. $U_k^{(k)}(\theta^{(k)})$ is the variational ansatz defined as

$$U^{(k)}(\theta^{(k)}) = \prod_j U_j^{(k)}(\theta_j^{(k)}) \tag{2.24}$$

and each U_j is a quantum gate, which is generalized as

$$U_j^{(k)}(\theta_j^{(k)}) = e^{i\theta_j^{(k)} P_j^{(k)}} \tag{2.25}$$

where $P_j^{(k)}$ is a Hermitian matrix, usually is a tensor product of Pauli matrices. It is sometimes convenient to utilize the superoperator formalism and consider the noise into the optimization process, the expectation value can be written as

$$\left\langle \hat{O}_k(\theta^{(k)}) \right\rangle = \mathrm{Tr}\left[\hat{M}_k \Phi^{(k)}(\theta^{(k))\rho_0} \right] \tag{2.26}$$

where ρ_0 is the initial density operator and $\Phi^{(k)}(\theta^{(k)})$ denote the transformation matrix, which is given by

$$\Phi^{(k)}(\theta^{(k)}) = \prod_j \Phi_j^{(k)}(\theta_j(k)) \tag{2.27}$$

For devices that support single shot readout, each sample the quantum device would yield a bit string s. For each string s a measurement value could be calculated with $M_i(s)$, and the expectation value is the average of each $M_k(s)$.

$$\left\langle \hat{O}_k(\theta^{(k)}) \right\rangle = \sum_j \text{Prob}(s(\theta^{(k)}) = b_j M_k b_j \tag{2.28}$$

where $b_j \in B$, B covers all possible single shot string (all binary number from 0 to 2^{n-1} of the measurement outcome. Due to the physical implementation from the quantum hardware, not all quantum computing system support single shot readout. Some systems can only yield expectation value by averaging the signal from the readout. For example, some NMR systems use an ensemble of molecules to implement quantum computing and cannot read the state of each single molecule. In practice, the quantum hardware system may directly yield an expectation value, and the measurement approaches vary from different physical systems.

Bibliography

[1] Alberto Peruzzo et al. In: *Nature Communications* 5 (1 Sept. 2014), p. 4213.

[2] Sanjeev Arora and Boaz Barak. Cambridge University Press, 2009.

[3] Ken M. Nakanishi, Keisuke Fujii, and Synge Todo. In: *Physical Review Research* 2 (4 Oct. 2020), p. 043158.

[4] Jarrod R McClean et al. In: *New Journal of Physics* 18 (2 Feb. 2016), p. 023023.

[5] A V Uvarov and J D Biamonte. In: *Journal of Physics A: Mathematical and Theoretical* 54 (24 June 2021), p. 245301.

[6] Joonho Lee et al. In: *Journal of Chemical Theory and Computation* 15 (1 Jan. 2019), pp. 311–324. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.8b01004.

[7] William J. Huggins et al. In: *npj Quantum Information* 7 (1 Dec. 2021), p. 23.

[8] Bernard Zygelman. Springer International Publishing, 2018.

[9] Giuliano Benenti et al. World Scientific, 2018.

[10] Amira Abbas et al. 2020. URL: http://community.qiskit.org/textbook.

[11] X. Bonet-Monroig et al. In: *Physical Review A* 98 (6 Dec. 2018), p. 062339.

[12] Zhang Jiang et al. In: *Quantum* 4 (June 2020), p. 276.