Iterative qubit coupled cluster method: A systematic approach to the full-CI limit in quantum chemistry calculations on NISQ devices

Ilya G. Ryabinkin and Scott N. Genin

OTI Lumionics Inc., 100 College St. #351, Toronto, Ontario M5G 1L5, Canada
(Dated: June 27, 2019)

Noisy intermediate-scale quantum (NISQ) devices is the current realm of quantum computing. To exploit their limited powers to full extent it is imperative to develop suitable algorithms. We present the iterative qubit coupled cluster (iQCC) method aimed to perform realistic quantum chemistry calculations on such devices. The iQCC method can use as few as a single generator of entanglement (a single “Pauli word”) yet is able to reach the full configurational interaction (FCI) limit. The iQCC method is also endowed with a few techniques that allow for relatively large problems to be addressed. We show the capacity of our method on the electronic structure calculations for the LiH (4 qubits), H$_2$O (6 qubits) and N$_2$ (14 qubits) molecules.

I. INTRODUCTION

Demonstration of superior power of quantum computations is a very popular theme today. However, multiple limitations of contemporary quantum architectures makes such demonstration—evidence of quantum supremacy—very difficult. Most of the existing universal quantum computers are limited to the relatively small – typically, a couple of dozens – number of qubits, cannot entangle them with arbitrary connectivity, and their entangling gates, especially the multi-qubit ones, have insufficient fidelity, which implies short decoherence time: a quantum superposition that is believed to be at the heart of “quantumness” can exist only for limited time. Low fidelity also restricts the number of gates that could be applied to steer a superposition in time. Such a situation is dubbed in literature as the “era of noisy intermediate-scale quantum (NISQ) devices [1].” Consequently, most of the real applications of quantum computations, for example, for quantum cryptography, are currently impossible. However, certain applications, such as quantum chemistry calculations, can be made meaningful having merely 30–50 qubits and a few hundreds of gates available [2], given a proper method tailored to both a specific application and quantum hardware requirements. The goal of this paper is to propose such a method. Our development relies on our previous methodology [3, 4] as well as numerous proposals and advancements that already existed or occurred since then in the literature [5–13].

The central problem of quantum chemistry—the electronic structure problem—is to describe properties, such as equilibrium structures, optical spectra, atomization and reaction energies, etc. of atoms, molecules, and solids by solving the electronic Schrödinger equation [14]. Computational complexity of its exact solution is, unfortunately, exponential with respect to the size of a system; numerous approximate methods have been proposed to tackle this problem, but there is always significant trade-offs between simulation accuracy and the computational demands of a given quantum chemistry method. For example, the coupled-cluster singles and doubles and non-iterative triples (CCSD(T)) method [14], known as the gold standard in quantum chemistry, can reach chemical accuracy (≤ 1 kcal mol$^{-1}$) only for closed-shell molecules near equilibrium configurations and still possess unfavourable scaling ~ $N^7$, making it impractical for large problems, which are often of scientific or commercial interest. Moreover, even this computationally intensive method cannot accurately describe chemical reactions or materials containing transition metals.

Following the seminal ideas of R. Feynman [15–17], modeling of quantum mechanical systems, that is, atoms, molecules, and solids, is considered as an ideal application of quantum computations. The first method for simulating electronic spectra of molecular Hamiltonians, the quantum phase estimation (QPE) algorithm, has been proposed in late nineties and later been applied to a hydrogen molecule [18–20]. However, the number of gates and coherence requirements of QPE quickly ruled out any possibility to simulate larger molecules on current NISQ devices. This lead to development of a more resource-efficient method, the variational quantum eigensolver (VQE) [5, 6], which replaced the long coherenst evolution required by QPE by many short coherent evolutions. Since its inception, higher practical utility of VQE has been demonstrated by realistic simulations of molecules larger than H$_2$, namely, LiH, BeH$_2$, and H$_2$O [10, 21, 22].

We emphasize that VQE is an umbrella term, and various forms are possible, and, in fact, have already been investigated in literature. In the following sections we briefly recapitulate steps preceding application of VQE to quantum chemistry problems, and analyze different flavours of the VQE approach in terms of their origins, computational complexity, ability to approach the exact answer systematically, and the ability to utilize given (possibly limited) quantum resources to position our developments properly. We do not consider, however, any scenario involving noisy quantum hardware. We also assume that shallower quantum circuits containing smaller number of gates, and shorter Hamiltonians are universally preferable, as well as lesser amount of work incurred on either quantum or classical computer.

We announce the following properties of our new iterative qubit coupled cluster (iQCC) method: 1) it is a
systematic procedure built upon the qubit coupled cluster (QCC) predecessor \cite{3} that approaches the exact [the full configurational interaction (FCI)] limit in an iterative fashion by repeating the QCC calculations followed by “dressing” of intermediate Hamiltonians with the generated predecessors \cite{3}. It shows the linear (geometric) convergence rate, it allows for flexible partitioning of computational efforts between quantum and classical devices—down to a single generator of entanglement that is translated to just a few entangling gates constituting a quantum circuit.

Together with the development of the iQCC method we introduce a novel efficient screening procedure for selecting generators of entanglement for the QCC procedure, which contributes significantly to numerical efficiency of the method. We compare the new screening procedure with several alternatives that are related to the existing VQE-style approaches, namely those that are based on unitary coupled cluster (UCC) parametrization (see Sec. II.D) and “hardware-efficient” Ansatz by Kandala \textit{et al.} \cite{10}.

We support our developments by calculations of the total electronic energies of LiH, H$_2$O, and N$_2$ molecules by the iQCC method. The latter two molecules are commonly used as examples of the strong correlation problem \cite{23–26}. When their bonds are stretched (symmetrically for the H$_2$O molecule), the traditional methods which use the single-determinant (Hartree–Fock) solution as a reference, such as coupled-cluster singles and doubles (CCSD) or CCSD(T), fail, and one has to resort to exponentially difficult multiconfigurational approaches, like multiconfigurational self-consistent field (MCSCF) or even FCI. Having in mind applications of the iQCC method for solving the strong correlation problem, we assess its capabilities relative to the FCI solution, or, to be more precise, the FCI solution for a given active set [complete active space (CAS)] of orbitals.

II. THEORY

A. The electronic structure problem

The electronic structure problem seeks the solution of the time-independent electronic Schrödinger equation,

$$\hat{H}_e(\vec{r}|\vec{R})\Psi_i(\vec{r}|\vec{R}) = E_i(\vec{R})\Psi_i(\vec{r}|\vec{R}).$$

(1)

\(\hat{H}_e\) is the electronic Hamiltonian of a molecule in first quantization. It depends on a set of electronic coordinates \(\vec{r} = (r_1, \ldots, r_N)\) as variables and nuclear positions \(\vec{R} = (R_1, \ldots, R_N)\) as parameters. Solutions of Eq. (1), \(E_i(\vec{R})\) and \(\Psi_i(\vec{r}|\vec{R})\), are known as potential energy surfaces (PESs) and electronic wave functions respectively. To account for indistinguishability and fermionic nature of electrons, the differential operator \(\hat{H}_e\) is replaced by its second-quantized counterpart

$$\hat{H}_e = \sum_{ij} \hat{a}^\dagger_i \hat{a}_j + \frac{1}{2} \sum_{ijkl} \langle ij \mid kl \rangle \hat{a}^\dagger_i \hat{a}^\dagger_j \hat{a}_k \hat{a}_l.$$  

(2)

which acts on vectors in a finite-dimensional Fock space. Here \(\hat{a}^\dagger_i\) (\(\hat{a}_i\)) are fermionic creation (annihilation) operators, and

$$\langle ij \mid kl \rangle = \int \psi_i^*(\vec{x}_1) \psi_j^*(\vec{x}_2) \frac{1}{r_{12}} \psi_k(\vec{x}_1) \psi_l(\vec{x}_2) d\vec{x}_1 d\vec{x}_2$$

(3)

are molecular one- and two-electrons integrals, respectively. \(\{\psi_i(\vec{x})\}_{i=1}^{N_a}\) are the spin-orbitals, that depend on a joined (spatial plus spin) coordinate of an electron, \(\vec{x} = (\vec{r}, \sigma)\), and constitute a spin-orbital basis of the size \(N_a\). Typically, spin-orbitals are themselves constructed as linear expansions over an auxiliary basis set of atomic-centered functions known as atomic orbitals. The size of the one-electron basis determines the size of a matrix representation of \(\hat{H}_e\), which is \(2^{N_a} \times 2^{N_a}\). Diagonalization of this matrix gives the “exact solution” of the electronic structure problem in a chosen one-electron basis, also known as the FCI solution. It can only be improved by enlarging the one-electron basis set. Exponential growth of the Fock-space Hamiltonian matrix with \(N_a\) makes the problem of finding the exact solution intractable for any but the smallest molecules. However, the FCI electronic energies, when available, are used as references in method development. Traditionally, a given method is considered accurate if it deviates from the FCI solution less than 1 m\(E_h\) (close to “more chemical” definition as 1 kcal mol\(^{-1}\) \(\approx 1.6 mE_h\), quoted above), which is approximately the average thermal energy \(k_b T\) at room temperature—a typical energy scale for chemical phenomena.

B. Qubitization

To be suitable for quantum computations, the Fock-space Hamiltonian (2) must be converted to a qubit form, in which fermionic creation (annihilation) operators are mapped to operators acting on qubits. This can be accomplished by using one of the fermion-to-qubit transformations, such as the Jordan–Wigner (JW) \cite{2, 27} or Bravyi–Kitaev (BK) \cite{28–32}. The qubit form of the Hamiltonian is:

$$\hat{H} = \sum_l \hat{H}_l = \sum_l C_l \hat{P}_l,$$

(5)

where \(C_l\) are numerical coefficients deduced from one- and two-electron integrals \(\langle ij \mid kl \rangle\), and \(\hat{P}_l\) are products

$$\hat{P}_l = \cdots \hat{p}^{(l)}_1 \hat{p}^{(l)}_0,$$  

(6)
A trial wave function in the UCC form is
\[ |\Psi(\tau)\rangle = \hat{U}(\tau) |0\rangle, \]
where \( \hat{U}(\tau) \) must be realized as a quantum circuit. \( j \) denotes the current step of the algorithm.

2. Measuring the expectation value of each term of the Hamiltonian (5) and summing them to the total energy estimate for a given set of parameters \( \tau(j) \),
\[ E(\tau(j)) = \langle \Psi(\tau(j)) | \hat{H} | \Psi(\tau(j)) \rangle = \sum_I G_I \langle \Psi(\tau(j)) | \hat{P}_I | \Psi(\tau(j)) \rangle, \]
(8)

3. Feeding \( E(\tau(j)) \) to a suitable minimization algorithm running on a classical computer to determine next-step parameters \( \tau(j+1) \) which lower the energy.

A nice VQE flowchart can be found in Ref. 33. It is quite obvious that there are multiple ways to choose the unitary \( \hat{U}(\tau) \) and the reference state \( |0\rangle \) in Eq. (7). Below we consider a few popular choices.

D. Unitary coupled cluster (UCC) parametrization of VQE

The UCC parametrization [34–36] was the first proposed to be used in the VQE formalism. To date, it is also the mostly elaborated approach [5, 7, 22, 33, 37, 38]. A trial wave function in the UCC form is
\[ |\Psi(\tau)\rangle = e^{\hat{T} - \hat{T}^{\dagger}} |\Phi_0\rangle, \]
where \( |\Phi_0\rangle \) is a fixed reference, commonly taken as an \( n \)-electron Slater determinant constructed from the set of canonical Hartree–Fock molecular orbitals (MOs),
\[
\hat{T} = \hat{T}_1 + \hat{T}_2 + \ldots, \\
\hat{T}_1 = \sum_{ia} t_{ia}^a \hat{a}_i^a \hat{a}_i^\dagger, \\
\hat{T}_2 = \sum_{i j a b} t_{ijab}^{ab} \hat{a}_i^a \hat{a}_i^\dagger \hat{a}_j^b \hat{a}_j^\dagger, \\
\ldots
\]
are cluster operators, and \( \{t_{ia}^a, t_{ijab}^{ab}, \ldots \} \equiv \tau \) are the corresponding amplitudes to be determined. \( \hat{E}_{ij}^{ab} \) are elementary excitation operators promoting electrons from spin-orbitals occupied in \( |\Phi_0\rangle \) labelled with \( i, j, \ldots \) to virtual (empty) spin-orbitals labelled with \( a, b, \ldots \).

\( \hat{T}_k \) are general mutually commuting fermionic \( k \)-fold (single, double, etc.) excitation operators. Without Hermitian conjugates, \( \hat{T}_k \). Eq. (9) can be recognized as the “traditional coupled-cluster Ansatz”, which in its CCSD(T) form is known as the “gold standard” of quantum chemistry. For weakly correlated systems amplitudes \( \tau \) decay quickly with the excitation rank \( k \) [39], allowing for highly accurate treatment at low \( k \), typically 2–3. Perhaps due to this behaviour, the unitary coupled cluster singles and doubles (UCCSD) parametrization has become very popular in VQE calculations. Apart from that, the UCC form guarantees the variational upper bound for the ground-state energy [see, however, Ref. 36].

Unfortunately, \( \hat{T}_k \) and \( \hat{T}_k^{\dagger} \) operators do not commute, which leads to a non-terminating Baker–Campbell–Hausdorff series [34] and makes the UCC Ansatz exponentially difficult for a classical computer. UCC cannot be directly implemented on a quantum computer either because of the sum of operators in exponent. Instead, one has to resort to Trotterization [33, 40], for example:
\[ e^{\sum_{k=1}^\infty \tau_k \hat{E}_k} \approx \left( \prod_{k=1}^\infty e^{\frac{\tau_k}{\rho} \hat{E}_k} \right)^\rho, \]
(13)
where \( \rho \) is the level of Trotterization, usually \( \rho = 1 \) or 2. While this approximation is likely to retain the property of a rapid fall-off of higher-rank cluster amplitudes, the energy estimates are no longer identical to the genuine UCC value.

This is not the only consequence of the “fitting” of the UCC form to quantum circuitry. The elementary excitation operators \( \hat{E}_{ij}^{ab} \) must be converted to a qubit form by means of any of fermion-to-qubit transformations mentioned above. This leads to their representation as sums of Pauli words, and after an additional Trotterization step, a product of exponents of individual Pauli words emerges:
\[ \hat{U}(\tau) = \prod_k \exp(\mathrm{i} t_k \hat{P}_k), \]
(14)
where \( \tau = (t_1, t_2, \ldots) \) are new amplitudes. Their explicit relations to the original \( \tau \) become rather complicated and depend on the type of the chosen transformation, the excitation rank of fermionic operators included in the UCC form, the basis set size, the arrangement of spin indices of spin-orbitals, etc.

Finally, additional difficulties arise due to the lack of one-to-one correspondence between fermionic excitation rank and the rank of Pauli words. Each elementary fermionic excitation operator of a fixed rank is a sum of Pauli words of different ranks. At this point the connection between fermionic and qubit pictures is largely
lost, and amplitudes \{t_k\} in Eq. (14) must be considered as abstract parameters.

Thus, the UCC parametrization, while being system-specific, may not be optimal. It provides a list of Pauli words, which contribute considerably into correlation energy, but there is no systematic way to select the “best” ones, for example, those which lead to the fastest convergence of the Ansatz (14). Hardware requirements can be accounted for in a \textit{ad hoc} manner by choosing, for example, the lowest-rank words, which does not, however, guarantee systematic convergence neither to the genuine UCC energy nor to the FCI limit.

E. “Hardware-efficient” parametrization of Kandala \textit{et al.} [10]

In Ref. 10 n-qubit trial states were parametrized as:

\[
\Psi(\theta) = \prod_{q=1}^{n} \left[ U^{q,d}(\theta) \right] \times U_{\text{ENT}} \times \cdots \\
\times \prod_{q=1}^{n} \left[ U^{q,0}(\theta) \right] \times U_{\text{ENT}} \times \prod_{q=1}^{n} \left[ U^{q,0}(\theta) \right] \left| 0 \right>.
\]

(15)

Eq. (15) has a structure of an alternating sequence of products of individual qubit rotations \(U^{q,i}(\theta)\),

\[
U^{q,i}(\theta) = e^{i\theta_{q}^1} e^{i\theta_{q}^2} e^{i\theta_{q}^3}, \quad 0 \leq i \leq d \tag{16}
\]

where \(\theta_{q}^1, \theta_{q}^2, \) and \(\theta_{q}^3\) are the Euler angles of \(q\)-th qubit, interleaved with action of one of the “entanglers” \(U_{\text{ENT}}\),

\[
U_{\text{ENT}}(t) = \exp(-it\hat{T}),
\]

(17)

where \(\hat{T}\) is just one of the Pauli words with rank \(k \geq 2\). \(\left| 0 \right>\) is the initial state of a qubit register, for example, “all spins up”.

The number \(d\) was called the depth of the scheme. In the experimental setup of Ref. 10 the first set of rotations in \(U^{q,0}(\theta)\) was not implemented \((\theta_{q}^0 = 0, q = 1 \ldots n)\), and entanglers’ amplitudes were all kept fixed, so that the Ansatz (15) contained \(n(3d + 2)\) variational parameters in total.

It has to be emphasized that the set of \(T\) at each level was tailored solely to hardware requirements, to achieve maximum coherency and to satisfy connectivity limitations; hence, this form represents the other extreme of VQE; the set of entanglers carries no information about the system being optimized. Nevertheless, for sufficiently large \(d\)—whose values were in fact system-dependent—the chemical accuracy has been achieved, though on a simulator, not the real hardware [10, 41].

F. The qubit coupled cluster (QCC) method

The first implementation of the QCC method was presented by current authors in Ref. 3. The design principles were: 1) to suggest a systematically improvable and potentially exact form of a trial wave function that can be directly implemented as a quantum circuit, 2) to minimize circuit depth as much as possible. We recognized that no matter how the space of trial functions was devised, the final form is the product of exponents of Pauli words (entanglers), as in Eq. (14). However, it might be beneficial to separate independent-qubit transformations (parametrized by Euler angles) from their multi-qubit counterparts in the spirit of Kandala \textit{et al.} [10] approach, since their hardware properties in terms of availability, connectivity, stability, fidelity, coherence times, etc. are quite different. Thus, we chose a dual-tier parametrization of a trial wave function. At the first tier, called a qubit mean-field (QMF) level, a direct product of states of individual qubits,

\[
|\Omega\rangle = \prod_{i=1}^{n} |\Omega_i\rangle, \tag{18}
\]

is used to represent a trial wave function.

\[
|\Omega_i\rangle = \cos \left( \frac{\theta_i}{2} \right) |\alpha\rangle + \sin \left( \frac{\theta_i}{2} \right) |\beta\rangle \tag{19}
\]

is a spin-coherent state for the \(i\)-th qubit [42–45], in which \((\phi_i, \theta_i) \equiv \Omega_i\) are azimuthal and polar angles on the “Bloch sphere” of the \(i\)-th qubit, and \(|\alpha\rangle\) and \(|\beta\rangle\) are eigenstates of a \(\hat{z}\) operator with eigenvalues \(+1\) and \(-1\), respectively. Note that Bloch and Euler angles are closely related, and any coherent state can be generated, up to an insignificant global phase, using three Euler rotations [46].

Eq. (18) is the simplest variational Ansatz possible on a quantum computer. It can be interpreted as an independent-qubit approximation. The closest fermionic analog of the QMF form (18) is a Slater-determinant wave function of the conventional Hartree–Fock method [14], with the main difference that qubits are treated as distinguishable particles as opposed to indistinguishable electrons. Nevertheless, the analogy between direct-product states, Eq. (18), and Slater determinants is limited, and the QMF and Hartree–Fock methods are neither identical nor subsets of each other [47]. The QMF ground-state energy is defined as a minimum of the following energy function with respect to all Bloch angles \(\Omega = \{(\phi_i, \theta_i)\}_{i=1}^{n}\):

\[
E_{\text{QMF}} = \min_{\Omega} \langle \Omega | \hat{H} | \Omega \rangle. \tag{20}
\]

At the QMF energy minimum, \(\left| \Omega \right>\) is a state of the quantum register at the end of VQE run. \(\left| \Omega \right>\) has a simple relation to the Hartree–Fock spin-orbital occupation numbers \(n_i\) if the Jordan–Wigner fermion-to-qubit transformation was used to map fermionic degrees of freedom to qubits, namely:

\[
n_i = \left< \text{HF} | \hat{a}_i^\dagger \hat{a}_i | \text{HF} \right> = \left< \text{QMF} \left| \frac{1}{2} (1 - \hat{z}_i) \right| \text{QMF} \right> = \begin{cases} 0, & i \in \text{virtual}, \\
1, & i \in \text{occupied}, \end{cases} \tag{21}
\]
That is, the QMF solution features definite z-projections in the qubit “standard” basis; in the absence of qubit-flip noise these projections can be measured without uncertainty. Additionally, as follows from Eq. (19), in this case the qubit coherent states are real (up to an immaterial global phase), and the QMF energy does not depend on \( \phi_i \) angles.

At the second tier the QCC method introduces a multi-qubit unitary transformation

\[
U(t) = \prod_{k=1}^{N_{\text{ent}}} \exp(-it_k \hat{T}_k/2),
\]

where all \( \hat{T}_k \) have rank \( \geq 2 \) applied to the QMF wave function (18). The QCC energy can be written as

\[
E_{\text{QCC}} = \min_{\Omega, t} \langle \Omega | U(t)^\dagger \hat{H} U(t) | \Omega \rangle.
\]

Minimization in Eq. (23) gives the ground-state energy estimate for a chosen set of \( \hat{T}_k \). In the original non-iterative version of the QCC method [3] the number of them, \( N_{\text{ent}} \), is determined mainly by hardware limitations, but once it is fixed, the final answer—the ground-state energy estimate—cannot be further improved.

**G. A concept of the direct interaction space (DIS)**

A central feature of the QCC method, which distinguishes it from other approaches, is the procedure how to choose \( \hat{T}_k \) for Eq. (22). In Ref. 3 we proposed to use all possible rank \( \geq 2 \) entanglers with non-zero contribution to the gradient of the QMF energy. Namely, for a single entangler \( \hat{T} \) we can derive from Eqs. (22) and (23):

\[
\frac{dE[\hat{T}]}{dt} = \frac{dE[t; \hat{T}]}{dt} \bigg|_{t=0} = \frac{d}{dt} \left. \min_{\Omega} \langle \Omega | e^{-i\hat{T}t/2} \hat{H} e^{-i\hat{T}t/2} | \Omega \rangle \right|_{t=0} = \langle \Omega_{\text{min}} | -\frac{i}{2} [\hat{H}, \hat{T}] | \Omega_{\text{min}} \rangle,
\]

where \( \Omega_{\text{min}} \) is the set of Bloch angles at the QMF energy minimum. To be included into Eq. (22) \( \hat{T} \) must satisfy

\[
\left| \frac{dE[\hat{T}]}{dt} \right| > 0.
\]

If it were possible to include all operators satisfying the condition (25) in Eq. (22), the QCC method would become a single step of the steepest descent method with the exact line search [48, Chap. 3]. Here and below we shall refer to the collection of such operators as the direct interaction space (DIS). It is not a space in its proper meaning: it is not either closed with respect to taking linear combinations or equipped with any structure or a bilinear form, so that we use this terminology only for the sake of brevity. Elements of the DIS connect the starting mean-field solution with several stationary points which are candidates for the ground-state minimum.

Note that generators \( \hat{T} \) satisfying the condition (25) are special to the VQE form. As evident from Eq. (24), non-zero gradients can only be produced by purely imaginary operators, in other words those that have purely imaginary diagonal elements. Since in the absence of magnetic fields the molecular Hamiltonian has only real matrix elements (assuming real MO basis), this means that \( \hat{T} \) must be purely imaginary, and as it follows from the matrix representation of Pauli operators, such \( \hat{T} \) must contain the odd number of \( \gamma_i \) operators. This simple observation can be used in the first place to discard many potential generators. Secondly—what is more important—such operators do not appear in the qubit Hamiltonian (5). This fact explains why, for example, the QPE algorithm requires long propagation times: it uses the evolution operator, \( e^{-i\hat{H}t} \), as a generator of entanglement, but none of the terms in the Trotter approximation of \( e^{-i\hat{H}t} \) contribute to the first-order changes in energy [40, 49] leading to long induction time before a system “feels” the changes. Thus, the first-order gradient condition is essential for superior efficiency of VQE-style schemes over QPE.

The screening procedure proposed in Ref. 3 was analytic: it could decide if a given \( \hat{T} \) should be taken or rejected based on the condition (25). However, there are \( 4^n - 3n - 1 \) possible rank \( \geq 2 \) Pauli words; obviously, such a large space cannot be sieved efficiently. It is beneficial, therefore, to devise a constructive procedure to determine operators belonging to the DIS.

**Construction of the direct interaction space**

Let an operator \( \hat{T} \) belong to the DIS for the Hamiltonian \( \hat{H} \), in other words, \( \hat{T} \) satisfies Eq. (25). Expanding a commutator in the parent expression, Eq. (24), by linearity as

\[
[H, \hat{T}] = \sum_I C_I [\hat{P}_I, \hat{T}] = \sum_I C_I \left( [\hat{P}_I, \hat{T}] - \text{h.c.} \right),
\]

where “h.c.” stands for the Hermitian conjugate of \( \hat{P}_I \hat{T} \), we recognize that a necessary condition for having a non-zero gradient in Eq. (25) is to have at least some of commutators in Eq. (26) to be non-vanishing. Thus, we seek for such \( \hat{T} \) that: 1) the product \( \hat{P} \hat{T} \) has a total purely imaginary phase \( \pm i \) for a given \( \hat{P} = \cdots \hat{p}_1 \hat{p}_n \), where \( \hat{p}_i \) is one of the non-trivial Pauli single-qubit operators, 2) \( \langle \Omega_{\text{min}} | \hat{P} \hat{T} | \Omega_{\text{min}} \rangle \neq 0 \). These conditions can be translated into the following qubit-driven algorithm for constructing all possible \( \hat{T} \) as products of Pauli elementary operators: \( \hat{T} = \cdots \hat{t}_1 \hat{t}_0 \).

1. For each \( \hat{p}_i \) in \( \hat{P} \) find triplets (\( \hat{t}_i, \hat{w}_i, f_i \)), where \( \hat{t}_i \) and \( \hat{w}_i \) are the Pauli elementary operators (including
the identity operator), and \( f_i \) is a phase \( (f_i = \pm 1 \text{ or } \pm i) \), satisfying the equation
\[
\hat{p}_i \hat{t}_i = f_i \hat{w}_i. \tag{27}
\]

2. Evaluate the expectation value of \( \hat{w}_i \) on the reference state \( |\Omega_{\text{min}}\rangle \)
\[
d_i = \langle \Omega_{\text{min}} | \hat{w}_i | \Omega_{\text{min}} \rangle = \begin{cases} 
\cos \phi_i \sin \theta_i, & \hat{w}_i = \hat{x}_i \\
\sin \phi_i \sin \theta_i, & \hat{w}_i = \hat{y}_i \\
\cos \theta_i, & \hat{w}_i = \hat{z}_i, \\
1, & \hat{w}_i = \hat{\epsilon} 
\end{cases} \tag{28}
\]
where \( \Omega_i = (\phi_i, \theta_i) \) are the optimized Bloch angles for the \( i \)-th qubit in \( \Omega_{\text{min}} \); see Eqs. (18) and (19).

3. Save all triplets \((t_i, d_i, f_i)\) for each \( \hat{p}_i \) in a table, which has as many rows as the number of non-trivial elementary Pauli operators in \( \hat{P} \).

4. Traverse each entry of the table starting from the least significant non-trivial qubit in \( \hat{P} \) and build corresponding operators as \( \hat{T}_J = \cdots \hat{t}^1_{i_0} \hat{t}^0_{i_0} \) compute the corresponding total phases \( \hat{F}_J = \cdots \cdots \hat{f}^1_{k_0} \hat{f}^0_{k_0} \) and total coefficients, \( \hat{D}_J = \cdots \cdots \hat{d}^1_{k_1} \hat{d}^0_{k_0} \), where \( J = (k_0, k_1, \ldots) \) is a multi-index that enumerates a particular path in the table. Keep only the solutions with purely imaginary total phases \( \hat{F}_J \) and non-zero contributions \( \hat{D}_J \). We shall refer to these solutions as “inner solutions”.

5. Compute the final gradient value for each of the solutions \( \hat{T}_J \)
\[
G_J = \sum_{I} C_I D_{IJ}^f, \tag{29}
\]
where \( D_{IJ}^f \) is the weight of \( \hat{T}_J \) for the operator \( \hat{P}_I \) from Eq. (26). Some solutions may have \( G_J = 0 \) due to specific relations between coefficients \( C_I \). Such relations depend on a molecular configuration (e.g. are manifestation of molecular symmetry) and an atomic basis set. Zero weights of these solutions are, therefore, “accidental”; we took a pragmatic approach and simply dropped them since they apparently contradict the definition (25).

How many solutions \( \hat{T}_J \) for each \( \hat{P} \) can we possibly have? For every qubit index \( i \) there is always a trivial solution \( \hat{t}_i = \hat{1}, \hat{f}_i = 1, \hat{w}_i = \hat{\epsilon} \) but to have a total imaginary phase there must be at least one non-trivial solution with \( \hat{t}_i \neq \hat{1} \). On the other hand, if a mean-field state \( |\Omega_{\text{min}}\rangle \) is characterized by Bloch angles that are not multiples of \( \pi \), one can have up to 4 possibilities per each qubit, but even for a “perfect” mean-field solution with \( \theta_i = (0, \pi) \) (if we work in a basis of \( z_i \) eigenstates), there is always a second non-trivial solution, \( \hat{t}_i = f_i \hat{p}_i \hat{z}_i \) with an appropriate phase \( f_i \). That is, the minimal number of solutions is \( \sim 2^{\text{rank}(\hat{P}) - 2} \) assuming that the overall phase can be imaginary only for a half of the solutions, and one imaginary phase is required. At the first glance, the space of inner solutions for molecular Hamiltonians is combinatorially large, but in fact, its precise size depends on the type of the fermion-to-qubit transformation employed.

For example, if the Jordan–Wigner transformation was used, the maximum rank of \( \hat{P}_I \) in the qubit Hamiltonian (5) is \( \sim n \), and indeed, the number of solutions is \( \sim 2^n \). However, for the Bravyi–Kitaev mapping the maximum rank is \( \sim \log_2(n) \), which leads to only polynomial size \( \sim 2^{\log_2 n} = n \).

It might appear, therefore, that the Bravyi–Kitaev transformation is superior to any other in terms of size of the DIS. Unfortunately, this is not the case. Combinatorial complexity re-emerges if one includes “outer solutions”, which are explained below. Consider a separable Hamiltonian, which acts on 4 qubits:
\[
\hat{H} = \hat{H}_{\text{Bell}} + \hat{z}_2 \hat{z}_3 = (\hat{x}_1 \hat{x}_0 + \hat{z}_1 \hat{z}_0) + \hat{z}_2 \hat{z}_3. \tag{30}
\]

It has two non-interacting groups of qubits, (0–1) and (2–3). \( \hat{H}_{\text{Bell}} \) has a strongly correlated, in fact, maximally entangled, ground state—one of the celebrated Bell states:
\[
|\Psi^-\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \tag{31}
\]

The QMF method applied to \( \hat{H}_{\text{Bell}} \) gives the mean-field solution, \(|\uparrow\uparrow\rangle [50]\), which lies strictly above the exact one (31) in energy. According to the DIS construction algorithm, two terms of \( \hat{H}_{\text{Bell}} \) gives rise to \( 2 \times 2^{(2–2)} = 2 \) inner solutions. They are: \( \hat{T}_1 = \hat{x}_1 \hat{y}_0 \) and \( \hat{T}_2 = \hat{y}_1 \hat{x}_0 \). However, any of them can be multiplied by an arbitrary string containing \( \hat{z}_2 \) and \( \hat{z}_3 \) variables leaving the value of the gradient unaffected, so that, by definition, products belong to the DIS of the full Hamiltonian. Examples of such operators are \( \hat{T}_3 = \hat{z}_2 \hat{x}_1 \hat{y}_0 \) and \( \hat{T}_4 = \hat{z}_3 \hat{z}_2 \hat{x}_1 \hat{y}_0 \) (enumeration is arbitrary). More generally, each entangler \( \hat{T}_J \) from the set of inner solutions for a given \( \hat{P} \) can be multiplied by Pauli elementary operators acting on qubits that are not presented in \( \hat{P} \), and if their mean-field expectation values are not zero, the resulting operators will belong to the DIS. In fact, for each \( T \) there are at least \( 2^{(n-\text{rank}(\hat{P}))} \) such “outer” solutions that correspond to all possible \( z \)-projections of qubits that are not affected by \( \hat{P} \). Thus, the algorithm shown above has to be modified: to include all (inner + outer) solutions, \( p_i \) must run all qubit indices; those that do not enter \( \hat{P} \) explicitly must be replaced by identity operators \( \hat{1} \).

Estimating the size of the DIS for the whole Hamiltonian \( H \) is more involved. Different \( \hat{P}_I \) lead to different set of solutions, but not all of them are distinct; different fermion-to-qubit transformations may also affect the result. We conjecture that the upper bound for the size of the DIS will be proportional to the number of terms in the qubit Hamiltonian, \( \sim n^4 \), which is related to the number of one- and two-electron integrals in the parent fermionic Hamiltonian (2) times \( 2^n \), which is the number of inner + outer solutions. Such an estimate, \( n^4 \times 2^n \),
is consistent with the number of free parameters in the problem: the ground state vector is a $2^n$-element vector in the Hilbert space of $n$ qubits, which is uniquely defined (assuming no degeneracy) by the Hamiltonian, which, in turn, depends on $\sim n^4$ values of molecular integrals.

We also conjecture that the size of DIS is asymptotically close (i.e., optimal) to the estimated gate complexity [which, in turn, is proportional to the number of $T_k$ in Eq. (22)] of a unitary transformation connecting two arbitrary states in the Hilbert space of $n$ qubits. As proven by Nielsen and Chuang [51, Sec. 4.5.4], to construct a unitary $U$ that satisfies

\[ \| \langle \Psi | - U | 0 \rangle \| \leq \epsilon, \]  

where $\Psi$ is a target state, $| 0 \rangle$ is a starting state, and $\epsilon$ is the required accuracy, one needs $\sim 2^n \log (1/\epsilon)/\log n$ elementary gates, which matches asymptotically the size of the DIS.

Overall, our algorithm constructs $\sim 2^n$ generators of entanglement, which satisfy Eq. (25). This estimate must be compared against $\sim 4^n$ operators (all possible Pauli words) that have to be tested in the naïve version of screening.

H. The iterative qubit coupled cluster (iQCC) method

General idea: optimization followed by dressing

Simply rewriting the QCC energy expression (23) as

\[ E_{\text{QCC}} = \min_{\Omega} \{ \min_{\Omega} \langle \Omega | \tilde{H}(\Omega) | \Omega \rangle \}, \]  

one recognizes that the QCC energy can be written as the minimum of QMF energies for a unitarily-transformed, “dressed” Hamiltonian

\[ \tilde{H}(t) = U^\dagger(t)\hat{H}U(t), \]  

parametrized by the set of amplitudes $t$. $\tilde{H}(t)$ can be evaluated recursively as

\[ \tilde{H}^{(k)}(t_k, \ldots, t_1) = e^{it_k T_k/2} \tilde{H}^{(k-1)}(t_{k-1}, \ldots, t_1) e^{-it_k T_k/2} = \tilde{H}^{(k-1)} + \sin t_k \left( \frac{i}{2} \tilde{H}^{(k-1)}, T_k \right) + \frac{1}{2} (1 - \cos t_k) \left( \tilde{H}^{(k-1)} T_k \tilde{H}^{(k-1)} - \tilde{H}^{(k-1)} \right), \]  

where $k = 1, \ldots, N_{\text{ent}}$ and $\tilde{H}^{(0)} = \hat{H}$. This procedure produces $3^{N_{\text{ent}}}$ distinct operator terms and exposes exponential complexity of the QCC form. However, each iteration of dressing (with fixed amplitudes) has only polynomial complexity since it is determined by the number of terms in the Hamiltonian. Consider the dressing of the original Hamiltonian $\tilde{H}^{(0)} = \hat{H}$, which has $M \sim n^4$ terms, with a unitary generated by the first entangler $U_1 = \exp(i T_1 t_1)$, for a fixed value of $t_1$. How many terms compared to the number of terms in the original $\tilde{H}$ a new qubit operator $\hat{H}$ computed by Eq. (35) can have?

Thus, new algebraically independent terms may only arise from the second, commutator term in Eq. (35). However, the presence of the commutator typically decrease the length of the corresponding operator—on average its size is a half of the size of the original $\hat{H}$, so we assume the “average” change of the length $M^{(k)} \approx 3/2 M^{(k-1)}$ and the worst-case $M^{(k)} = 2 M^{(k-1)} - |\hat{H}, T|$ cannot have more terms than $\hat{H}$ itself.

The idea of dressing lends itself to an iterative form of QCC: After the first optimization of $t$ amplitudes, absorb their values in the dressed Hamiltonian by evaluating $\tilde{H}^{(1)}$ according to Eq. (35) on a classical computer and take it as a starting point for the next iteration of the QCC method. The length of the updated $\tilde{H}$ is expected to be $\sim (3/2)^{N_{\text{ent}}(0)} M$, where $N_{\text{ent}}(0)$ is the number of entanglers used to set up the first QCC calculation.

What are the benefits of the iterative QCC procedure over a single-step one? First and foremost, it allows for flexible control of quantum resources used to implement the QCC Ansatz. It can make use of even a single entangler: as long as the optimal amplitude is found and its value is “stored” in the dressed Hamiltonian, and if the ground-state energy calculated with the dressed Hamiltonian was lower than that with the original one (which should always be the case in the absence of noise due to variational principle), the procedure can be deemed successful. Dressing effectively nullifies the impact of noise introduced by a NISQ device, and robust error correction is no longer necessary. The price paid for that is the larger size of the Hamiltonian to be measured: if the exact dressing is performed, one has exponential (albeit milder than for QCC with respect to $N_{\text{ent}}$) growth, $\sim (3/2)^{N_{\text{ent}}}$ in the number of measurements. Fortunately, the exactness of the transformation is not always required: one is typically interested in accuracy to within a certain threshold, for example, to chemical accuracy, and it is possible to define a more compact Hamiltonian yet to be able to reach the target accuracy.
Compression of intermediate Hamiltonians

If limited accuracy is sought for the final result, it is natural to “compress” intermediate and final Hamiltonians in such a way as to guarantee that the spectral disturbance of the ground-state energy,

$$ |E_0(\hat{H}^{(k)}) - E_0(\hat{H})| \leq \epsilon, $$

is less than the desired accuracy $\epsilon$. One can target, for example, the chemical accuracy $\epsilon = 1 \text{mE}_h$.

A compression procedure, which we propose, is based on the Weyl’s spectral perturbation theorem [53, 54]:

$$ \max_j |\lambda_j^0(\hat{H}) - \lambda_j^0(\hat{T})| \leq \|\hat{T} - \hat{H}\| \leq \|\hat{T} - \hat{\tilde{H}}\|_F, $$

where $\{\lambda_j^0\}$ are eigenvalues of the corresponding operator arranged in decreasing order, $\|\hat{A}\| = \sup_{\|\Psi\|=1} \|\hat{A}\Psi\|$ is the operator norm, which is for a normal (diagonalizable) operator equal to $\max_j \{\lambda_j\}$, and $\|\hat{A}\|_F = \sqrt{\text{tr} (\hat{A}^\dagger \hat{A})}$ is the Frobenius, also known as the Hilbert-Schmidt, norm of $\hat{A}$ [55]. While the operator norm is difficult to compute—it requires to know at least extremal eigenvalues—the Frobenius norm of the qubit Hamiltonian (5) is easy:

$$ \|\hat{T}\|_F = 2^{n/2} \sqrt{\sum_j |C_j|^2}, $$

since $\text{tr} (\hat{P}_j^\dagger \hat{P}_j) = 2^n \delta_{jj}$ where $\delta_{jj}$ is the Kronecker symbol and $n$ is the number of qubits. Thus, if all $|C_j|$ are sorted in descending order, we can define an approximate (compressed) Hamiltonian as

$$ \hat{\tilde{T}} = \sum_{j \leq j_0} C_j \hat{P}_j, $$

where $j_0$ is chosen to satisfy

$$ \sqrt{\sum_{j > j_0} |C_j|^2} \leq \frac{\epsilon}{2^{n/2}}. $$

At each iteration of the iQCC method one can replace the dressed Hamiltonian $\hat{H}^{(k)}$ with its compression version, $\hat{\tilde{T}}$, to use it as a starting operator for the next iteration. According to the inequality (38), this will disturb the spectrum by no more than $\epsilon$.

A more natural question, however, is how the error accumulates during iterations. It could be answered \textit{a posteriori} by computing the Frobenius norm of the difference between the original, $\hat{H}^{(0)} = \hat{H}$, and the final, $\hat{H}^{\text{fin}}$, Hamiltonians,

$$ \|\hat{H} - \hat{H}^{\text{fin}}\|_F = \sum_j |C_j|^2 + \sum_j |C_j^{\text{fin}}|^2 - 2 \text{tr} (\hat{H} \hat{H}^{\text{fin}}), $$

using Pauli polynomial algebra.

The suggested compression procedure is particularly well-suited for the use in the iQCC method. As we established in the previous Section, the main cause of growing of the dressed Hamiltonians is the commutator term in Eq. (35). However, its average value on the QMF wave function is precisely the value of the gradient contribution of the corresponding $T$, which is systematically reduced by the iQCC procedure. Thus, after initial rapid growth of the size of intermediate Hamiltonians, one could expect progressively stronger compression when the commutator contributions start systematically falling below the compression threshold $\epsilon$. We verify these expectations on examples in Sec. III.

Overview of the algorithm

A computational scheme behind the iQCC method is as follows. The initial Hamiltonian $\hat{H}^{(0)}$ is set to the qubit Hamiltonian

$$ \hat{H}^{(0)} = \hat{H}, $$

which is derived from its fermionic counterpart, Eq. (2), for a problem of interest \textit{via} one of the fermion-to-qubit mappings assuming particular grouping of spin-orbitals, for example, first all spin-up then all spin-down (“same spin” grouping). Spin-orbital ordering is important because it affects the existence of stationary qubits. The initial step of the iQCC procedure is the QMF energy minimization to determine $E^{(0)}_{\text{QMF}}$, and the initial set of Bloch angles, $\Omega^{(0)}$.

The $j$-th iteration of the iQCC procedure starts from screening of entanglers $T_k$ by Eq. (24) using the Hamiltonian from the previous step, $\hat{H}^{(j-1)}$, and Bloch angles $\Omega^{(j-1)}$. $N_{\text{ent}} \geq 1$ entanglers with largest absolute gradients are then inserted into the QCC form, Eq. (22), and the QCC energy expression, Eq. (23), is minimized with respect to entanglers’ amplitudes and Bloch angles to determine their updated values, $\{t_i^{(j)}\}_{i=1}^{N_{\text{ent}}}$ and $\Omega^{(j)}$, respectively. Chosen entanglers and the corresponding optimized amplitudes are used to dress the Hamiltonian $\hat{H}^{(j-1)}$ by Eq. (35). Optionally, the dressed Hamiltonian can be compressed by Eq. (40). The resulting operator becomes an updated Hamiltonian $\hat{H}^{(j)}$. The cycle is repeated until any of the following termination conditions are met: 1) the energy difference between two consecutive iterations is less that a threshold, 2) the largest energy gradient (computed as a part of the screening procedure at the beginning of the next iteration) drops below a threshold, 3) the number of iterations exceeds a predefined value. At exit, the last computed value of the QCC energy becomes the best ground-state energy estimate.
TABLE I. Electronic structure calculations details and parameters of the second-quantized and qubit Hamiltonians for molecules used in the study

| Property                                  | LiH              | H$_2$O$^+$       | N$_2$          |
|-------------------------------------------|------------------|------------------|----------------|
| Molecular configuration                    | $d$(LiH) = 1.7 Å | $d$(OH) = 1.25 Å | $d$(NN) = 2.118 Å |
| Assumed symmetry$^b$                       | $C_{2v}$         | $C_{2v}$         | $D_{2h}$       |
| Atomic basis set$^c$                       | STO-3G           | 6-31G            | cc-pVDZ        |
| Molecular orbital (MO) set$^d$             | 6                | 13               | 28             |
| Complete active space (CAS)                | $2e/3$orb $(2\alpha_1, 3\alpha_1, 4\alpha_1)$ | $4e/4$orb $(1\beta_1, 3\alpha_1, 4\alpha_1, 2\beta_1)$ | $10e/8$orb (full valence) |
| Fermion-to-qubit mapping                  | parity$^e$       | Bravyi–Kitaev parity$^e$ |               |
| Spin-orbital grouping                     | same spin: first all $\alpha$ then $\beta$ |               |               |
| Number of qubits$^f$                      | 4                | 6                | 14             |
| Length of the qubit $\hat{H}$             | 100              | 165              | 825            |

$^a$ A similar setup for electronic structure calculations has been considered in Ref. 56.
$^b$ The full symmetry groups for diatomics LiH and N$_2$ are $C_{\infty v}$ and $D_{2h}$, respectively, but the maximal Abelian subgroups with all-real irreducible representations were chosen instead as required by the ALDET module of the GAMESS program [57, 58].
$^c$ Basis Set Exchange library [59].
$^d$ Used to calculate one- and two-electron integrals in Eq. (2).
$^e$ Described in Ref. 60.
$^f$ After qubit reduction; see Sec. IIIA.

### III. NUMERICAL EXAMPLES

#### A. Electronic structure calculations and qubitization

All electronic-structure calculations—generation of the Hartree–Fock orbitals for selected molecular configurations and computing molecular integrals—were carried out with a locally modified version [Feb. 14, 2018 (R1)] of the GAMESS electronic structure package [57, 58]. Details are summarized in Table I. Note that we use a concept of the active space, a subset of (spin-)orbitals, for which the fermionic Hamiltonian (2) is constructed.

Qubit Hamiltonians were derived by choosing an appropriate fermion-to-qubit transformation; see Table I. In all cases the qubit Hamiltonians possessed stationary qubits [61]: namely, qubits in positions $N_{so}$ and $N_{so}/2$, ($N_{so}$ is the number of spin-orbitals, twice the number of molecular orbitals in the active space) enter Hamiltonian terms either as $\tilde{e}$ or the identity operators $\hat{e}$. $z$-projections ($\pm 1$) of these qubits are, thus, constant (stationary) and the corresponding operators can be replaced with their eigenvalues to define qubit-reduced operators. This procedure has been applied to all Hamiltonians, and the final qubit counts are listed in Table I. Eigenvalues of $\tilde{e}$ operators for stationary qubits were chosen so as to ensure that the ground states of the full and reduced Hamiltonians are the same. The qubit reduction procedure has also been applied to operators other than a Hamiltonian; see discussion in Appendix.

#### B. Different sources of entanglers: comparison of various variational spaces

The rate of convergence and the limiting value for a VQE solution depend on the composition of a unitary and the initial state in Eq. (7). The most popular choice is the UCCSD parametrization [11, 33, 38], Sec. II D. Following this approach, we use spin-adapted UCCSD cluster operators as one possible source of entanglers.

Inspired by the “hardware efficient VQE approach” of Kandala et al. [10], we also consider spaces of fixed-rank Pauli words for some low ranks, in particular, from 2 to 6 (wherever it was possible). Such spaces can be thought of as constructed by analogy with the coupled-cluster excitation hierarchy, but as was mentioned above, there is no direct relation between fermionic excitation levels and ranks of Pauli words.

We compare all variational spaces introduced above: the DIS, UCCSD and various fixed-rank spaces. Among them, the DIS is special: any operator belonging to this space satisfies by construction the first-order optimum condition (25) and guarantees energy lowering with the initial rate proportional to the magnitude of the corresponding gradient. Other spaces may contain entanglers with zero gradients. Zero-gradient entanglers are not necessarily useless: they may still lead to energy lowering at a higher, most likely at the second, order (see Ref. 3 for the discussion of second-order criteria), but they may also drive the mean-field solution toward an excited state, which is clearly undesirable. More zero-gradient entanglers are in a space, more gradient tests (25) are needed to find a suitable entangler, but even if such an entangler is found, it might be beneficial to continue screening to identify an entangler with the largest (in absolute value)
gradient. Thus, the relative fraction of operators for a given gradient value is a convenient characteristic of a variational space. Ideally, the best variational space should only contain high-gradient operators.

Composition of different variational spaces for LiH, H$_2$O, and N$_2$ is plotted in Fig. 1. All spaces are computed with the initial qubit Hamiltonian and the lowest-energy QMF solution, which for the chosen set of MOs coincides with the Hartree–Fock one [47]. All plots demonstrate a common trend: larger-qubit problems have more diverse set of entanglers, but with an exception of the DIS, zero-gradient entanglers become vastly dominating. The UCCSD space is quite competitive for the LiH molecule, but its quality quickly deteriorates with increasing the size of the problem: for the N$_2$ molecule it has more than 80% of zero-gradient solutions and, thus, extensive screening is necessary to find the highest-gradient ones. It also evident that the fixed-rank variational spaces are especially poor choice: the lowest possible (i.e. rank-2) space does not only contain a very high fraction of zero-gradient generators, but also systematically misses the highest-gradient entanglers completely—note the absence of the corresponding bars in Fig. 1.

Apart from relative composition of a variational space, the total size matters: if only few entanglers are to be included into the iQCC Ansatz, large-scale screening might not be justified. This disfavours the DIS since its size grows exponentially compared to polynomial growth for UCCSD or any of the fixed-rank spaces. Table II reports actual sizes of all variational spaces.

Indeed, the size of the DIS increases faster (with the qubit count) than the size of other spaces. However, the use of the DIS might still be preferable for the following reasons. First, building the DIS by the algorithm of Sec. II G is very efficient: even for the N$_2$ molecule it is still 3 times faster than the screening of the UCCSD space by Eq. (24). Secondly, as one aims for finding entanglers with non-zero gradients, it is not necessary to build the whole DIS. Given the fact the DIS maximizes the probability of finding the top-gradient generators, a search in a small subset of the DIS could be sufficient. Finally, as we show below, only the entanglers drawn from the DIS provide systematic convergence to the FCI limit if a single generator ($N_{\text{ent}} = 1$, see Sec. III C) is used in the iQCC procedure.

### C. Convergence of the iQCC method

The ultimate goal of the iQCC method is meaningful electronic structure calculations using current NISQ devices. The iQCC procedure allows for flexible partition of computational efforts between quantum and classical devices. Screening, dressing, compression, and energy minimization at each iteration are all done by a classical computer, but if more entanglers are encoded by a quantum circuit, fewer iterations are needed to achieve the same accuracy. The iQCC method also trades the complexity of encoding a quantum state for the number of measurements: the shorter gate sequences (smaller $N_{\text{ent}}$) result in more iterations and more intermediate Hamiltonians that need to be measured [62].

The minimal iQCC form is a single multi-qubit entangler [$N_{\text{ent}} = 1$ in Eq. (22)] augmented with $\sim n$ single-qubit gates which represent an independent-qubit reference state parametrized by Bloch (Euler) angles. Although a single Pauli word (entangler) is converted into a sequence of multi-qubit entanglement gates and this de-
composition is quantum architecture-dependent (e.g. different for superconducting-qubit and ion-trap machines), in general, a single-entangler form is among the simplest possible. We discuss a single-entangler variant of the iQCC procedure first and then consider its generalization with $N_{\text{ent}} > 1$. Since our numerical implementation is not fully optimized, we mostly focus on the results for the LiH and H$_2$O molecules. Additional details that are relevant to a larger-scale problem (the N$_2$ molecule), such as efficiency and impact of the compression procedure are reported in Sec. III D

Single-entangler case, $N_{\text{ent}} = 1$

Convergence of the iQCC procedure for different variational spaces is shown in Fig. 2. The iQCC procedure with the DIS as a source of entanglers demonstrates almost a linear (geometric) rate of convergence toward the FCI limit. All other variational spaces show saturation before the exact energy is reached; notable exceptions are the full-rank spaces that contain all possible (for a given qubit size of a problem) Pauli words of rank greater than 2. However, these spaces contain the DIS as a subspace, and hence, this result is hardly surprising. Overall, the fixed-rank spaces perform poorly and recover only a small fraction of correlation energy; rank-2 space is especially bad—its limiting value is almost indistinguishable from the reference QMF energy.

The behavior of the iQCC/UCCSD curve for the LiH molecule is somewhat surprising. A space of single and double fermionic excitations is complete for two active electrons, and one expects the energy to converge to the exact value, whereas the actual curve demonstrates saturation similar to that demonstrated by the fixed-rank counterparts. We return to this problem in the following Section.

Multiple-entanglers case, $N_{\text{ent}} > 1$

When several entanglers are to be included into the QCC Ansatz (22), a problem of degenerate solutions appears. Namely, as we discussed in Sec. II G, there are many entanglers $T_J$ that are characterized by identical gradients in Eq (24); for example, all “outer” solutions are like that. It is highly probable then that the first $N_{\text{ent}}$ will have identical gradients. Including them into Eq. (22) may lead to a bias in sampling of the variational space and limit search directions to a particular sector of solutions. To avoid such a bias and diversify representation of the variational space we group entanglers with identical (or numerically close, e.g. to $10^{-5}$ or any other convenient threshold) gradients and randomly draw $N_{\text{ent}}$ of them from different groups, implementing some sort of stochastic strategy. This increases uncertainty in our results, as new different runs lead to a different selection of entanglers and, thus, slightly different convergence patterns. Nevertheless, we repeat our calculations several times to guarantee that the results and conclusions made about various aspects of iQCC do not depend on the exact sequence of the entanglers employed in the iQCC procedure.

If multiple entanglers are included into the QCC form, faster convergence is expected. As evident from Fig. 3, this is indeed the case. Only 5 iterations are needed for $N_{\text{ent}} = 6$ to bring the energy closer than $1 \times 10^{-8} E_h$ to the FCI limit. Similar trends were observed for the H$_2$O molecule (not shown).

Sometimes grouping found less than $N_{\text{ent}}$ distinct gradient groups. In this case $N_{\text{ent}}$ is dynamically adjusted to match the number of groups; these numbers are also shown in Fig. 3. We do not expect this problem to occur for larger-qubit problems since the number of gradients groups will be much larger than $N_{\text{ent}}$.

Simultaneous use of several entanglers in the QCC form may lead to a different set solutions as compared to that obtained when the same number of entanglers is used sequentially. For example, we found (see Fig. 4) that using $N_{\text{ent}} = 6$ the iQCC procedure with rank-2 variational

![Graph of LiH and H$_2$O](https://via.placeholder.com/150)

**FIG. 2.** Convergence of a single-entangler ($N_{\text{ent}} = 1$) iQCC procedure for various choices of variational spaces. Full-rank, i.e. those that contain all possible (for a given qubit size of a problem) Pauli words of rank greater than 2, spaces are labelled with (*).
Existence of alternative solutions for $N_{\text{ent}} > 1$ explains the UCCSD/LiH quandary mentioned in the previous Section: it is necessary to include more than one entangler in the QCC form to reach the FCI energy; missing operators come out as commutators of those existing in the UCCSD space. However, zero-gradient entanglers might also be needed to close the algebra, so that the question whether completeness of the UCCSD space is compatible with the gradient condition (25) remains open.

As we established in Sec. II H, the dressing procedure must increase the size of Hamiltonians at each iteration by a factor of $3/2$ “on average”. We verified this expectation on the iQCC calculations for the $N_2$ molecule with $N_{\text{ent}} = 1$ and the UCCSD variational space. The ratio of the size of the Hamiltonian after dressing, $\hat{H}^{(n)}$, to the size of the Hamiltonian at previous iteration, $\hat{H}^{(n-1)}$, is plotted in Fig. 5. Indeed, sizes of Hamiltonians after the dressing procedure steadily increase at each iteration, but the relative growth never exceeds 1.5. Nevertheless, even with this milder growth, intermediate Hamiltonians quickly become intractable and compression seems to be imperative. We applied the compression procedure described in Sec. II H to intermediate Hamiltonians in the $N_2$ problem with the compression threshold $\epsilon = 5 \text{ mE}_h$, which is slightly larger than the chemical accuracy. However, since this estimate is conservative, we expected to reach the accuracy close to chemical for the ground state notwithstanding. In Fig. 6 we plot sizes of intermediate Hamiltonians at each iteration after the compression. The compression procedure remarkably changes the rate of growth of intermediate Hamiltonians from expected exponential to presumably polynomial allowing much larger calculations to be performed.

IV. CONCLUSIONS

We have presented and tested a novel technique for quantum chemistry calculations on universal quantum computers dedicated specifically for use with NISQ devices—the iterative qubit coupled cluster (iQCC) method. It is an iterative version of our previous technique, the QCC method [3], which belongs to the VQE class of algorithms.

The iQCC method allows for flexible partition of com-
As few as a single Pauli word can be used as a generator of wave functions. Symmetry breaking can be used as a shortcut to highly-correlated solutions for some complicated problems. However, not every operator can be reduced, as its symmetry may not be compatible with the symmetry of the Hamiltonian. In particular, the UCC Ansatz, whose cluster operators are written in a spin-orbital basis, preserves the electron-number but not the spin symmetry—some of $\hat{T}$ operators do not commute with $S^2$. Such operators couple different spin sub-blocks of the Hamiltonian and their qubit expressions have operators other than $\hat{\sigma}$ at the position of the stationary qubits. To find reducible combinations, we derived spin-adapted cluster amplitudes following a general scheme given by Eqs. 13.7.2 and 13.7.2 of Ref. 14. In particular, we solved equations

\begin{align}
[S_{\pm}, \hat{T}_k] &= 0, \\
[S_z, \hat{T}_k] &= 0, \quad k = 1 \text{ (singles), } 2 \text{ (doubles), } \ldots
\end{align}

for the qubit expressions of $\hat{T}_k$, $\hat{S}_z$, and $\hat{S}_{\pm}$ (spin raising/lowering) operators. The solutions are spin-free cluster amplitudes that have stationary qubits at the same positions as the Hamiltonian, and thus, can be reduced by the same procedure.

Qubit reduction, however, alters the size of the UCC variational space as some of the cluster amplitudes do not contribute to the UCC form anymore. The reader must be cautious comparing our results (like those listed in Table II) with other implementations of the UCC Ansatz reported elsewhere.

Overall, the iQCC procedure seems to be the most economical VQE variant presented in the literature to date; other iterative approaches require orders of magnitude more quantum resources for the same problems. We hope, therefore, that the iQCC method will become the method of choice for quantum chemistry calculations on NISQ devices.
ACKNOWLEDGMENTS

Authors thank Prof. A.F. Izmaylov for multiple useful discussions.
aspects in chemistry and physics, George Fisher Baker non-resident lectureship in chemistry at Cornell University (Wiley, 1988).

[47] I. G. Ryabinkin, S. N. Genin, and A. F. Izmaylov, J. Chem. Phys. 149, 214105 (2018), arXiv:1806.00514 [physics.chem-ph].

[48] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed., Springer Series in Operations Research and Financial Engineering (Springer New York, 2006).

[49] J. R. McClean, R. Babbush, P. J. Love, and A. Aspuru-Guzik, J. Chem. Phys. 149, 214105 (2018), arXiv:1806.00514 [physics.chem-ph].

[50] In fact, there are two degenerate mean-field solutions, |↓↑⟩ and |↑↓⟩. To break degeneracy and prevent their mixing, which may result in values of Bloch angles θ_0,1 ≠ (0, π) artificially inflating the DIS, one can introduce a small external perturbation of the form ε_0 with ε → 0.

[51] M. Nielsen and I. Chuang, Quantum Computation and Quantum Information: 10th Anniversary Edition (Cambridge University Press, 2010).

[52] An operator is said to be involutory if it satisfies \hat{σ}^2 = 1.

[53] H. Weyl, Math. Ann. 71, 441 (1912).

[54] R. Bhatia, Matrix Analysis, Graduate Texts in Mathematics (Springer-Verlag New York, Inc, 1996).

[55] To align our consideration with the Weyl’s theorem, one should use the negate of the real Hamiltonian, −\hat{H}.

[56] M. L. Abrams and D. C. Sherrill, Chem. Phys. Lett. 404, 284 (2005).

[57] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. Montgomery, J. Comput. Chem. 14, 1347 (1993).

[58] M. S. Gordon and M. W. Schmidt, in Theory and Applications of Computational Chemistry: The first forty years, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005) pp. 1167–1189.

[59] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. Windus, J. Chem. Inf. Model. 47, 1045 (2007).

[60] M. A. Nielsen, School of Physical Sciences The University of Queensland (2005).

[61] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, ArXiv e-prints (2017), arXiv:1701.08213 [quant-ph].

[62] A. F. Izmaylov, T.-C. Yen, and I. G. Ryabinkin, Chem. Sci. 10, 3746 (2019).

[63] Z. Wang, Y. Li, and J. Lu, J. Chem. Theory Comput. 15, 3558 (2019), pMID: 31042383.

[64] G. E. Scuseria, C. A. Jimnez-Hoyos, T. M. Henderson, K. Samanta, and J. K. Ellis, J. Chem. Phys. 135, 124108 (2011).

[65] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, arXiv e-prints (2018), arXiv:1812.11173 [quant-ph].