



OPEN

## Reduction of the molecular hamiltonian matrix using quantum community detection

Susan M. Mniszewski<sup>1✉</sup>, Pavel A. Dub<sup>2</sup>, Sergei Tretiak<sup>3</sup>, Petr M. Anisimov<sup>4</sup>, Yu Zhang<sup>3</sup> & Christian F. A. Negre<sup>3</sup>

Quantum chemistry is interested in calculating ground and excited states of molecular systems by solving the electronic Schrödinger equation. The exact numerical solution of this equation, frequently represented as an eigenvalue problem, remains unfeasible for most molecules and requires approximate methods. In this paper we introduce the use of *Quantum Community Detection* performed using the D-Wave quantum annealer to reduce the molecular Hamiltonian matrix in Slater determinant basis without chemical knowledge. Given a molecule represented by a matrix of Slater determinants, the connectivity between Slater determinants (as off-diagonal elements) is viewed as a graph adjacency matrix for determining multiple communities based on modularity maximization. A gauge metric based on perturbation theory is used to determine the lowest energy cluster. This cluster or sub-matrix of Slater determinants is used to calculate approximate ground state and excited state energies within chemical accuracy. The details of this method are described along with demonstrating its performance across multiple molecules of interest and bond dissociation cases. These examples provide proof-of-principle results for approximate solution of the electronic structure problem using quantum computing. This approach is general and shows potential to reduce the computational complexity of post-Hartree–Fock methods as future advances in quantum hardware become available.

At this stage in quantum computing, it is useful to demonstrate the application of quantum algorithms to real-world problems even at small scale where classical solutions are available, though may be approximate. This allows for verification and validation, and helps us prepare for the larger-scale problems with unknown results on future quantum hardware. Noisy Intermediate-Scale Quantum (NISQ) technology has provided us the opportunity to explore new approaches to simulate chemistry and physics<sup>1</sup>. By taking advantage of quantum–mechanical effects, NISQ devices promise to improve solutions, provide new pathways to solutions and even solve the problems that are intractable on current classical computers. In addition, quantum formulations of complex network algorithms have gained interest in quantum information science<sup>2</sup>. The cross-disciplinary approach of combining network science and quantum computing has provided new ways to solve chemistry and physics problems that can be framed as graphs. In this work we explore reducing the Hamiltonian matrix for electronic structure calculations using *Quantum Community Detection*<sup>3,4</sup> on the D-Wave Quantum Annealer. The approach does not rely on any prior knowledge about the chemical system of interest and is based on a matrix formulation only.

Computational chemistry aims to solve the Schrödinger equation numerically for many-body quantum systems of interest such as molecules or solid-state materials. For the vast majority of applications, we are interested in solving the time-independent, non-relativistic form of this equation for electrons by treating nuclei as classical point charges (using the Born–Oppenheimer approximation). The resulting electronic Hamiltonian can be solved exactly within the space spanned by the finite one-electron orbital basis set or a numerical grid. The corresponding method is known as full configuration interaction (FCI) or the diagonalization of the Hamiltonian matrix in the basis of Slater determinants (SDs). For an  $N$ -electron molecule, the HF states  $\Psi_0$  form the lowest-energy SD. Other SDs can be described by showing their difference from the HF determinant  $|\Psi_0\rangle$ , including singly excited SDs  $|\Psi_i^a\rangle$  (which differ from  $|\Psi_0\rangle$  by exchanging an occupied with an unoccupied orbital), doubly excited SDs  $|\Psi_{ij}^{ab}\rangle$ , and up to  $N$ -tuple excited determinants  $|\Psi_{ijk\dots}^{abc\dots}\rangle$ . These SDs form a complete basis of the problem in Hilbert space accounting for the exchange and correlation energy of the multielectronic system. The diagonalization of the resulting FCI matrix with elements  $\langle\Psi_m|H|\Psi_n\rangle$ , where  $m/n$  goes over all possible SDs, provides an

<sup>1</sup>Computer, Computational and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. <sup>2</sup>Chemistry Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. <sup>3</sup>Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. <sup>4</sup>Accelerator Operations and Technology Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ✉email: smm@lanl.gov

unambiguous standard with which to compare more approximate methods. Since FCI scales factorially with the number of electrons and spin-orbitals, its practical implementation is limited to very small molecules/basis sets. For example, molecular dicarbon ( $C_2$ ) in aug-cc-pVTZ basis (with 12 electrons to be distributed on 184 spin-orbitals) is one of the largest molecules ever simulated by means of FCI using the Oak Ridge National Lab Cray-X1 supercomputer<sup>5</sup>. To overcome these limitations, various approximate solutions have been developed by either truncating the problem in a reduced user-defined active space (e.g. multiconfigurational self-consistent methods such as the complete active space self-consistent field (SCF) method (CASSCF)<sup>6–8</sup>) or in a few SDs by employing Configurational Interaction (CI) or Coupled Cluster (CC) approximations<sup>9</sup>. Examples of the truncation methods include CI + single excitations (CIS), CI + single and double excitations (CISD), CI + single, double, and triple excitations (CISDT), CI + single, double, triple, and quadruple excitations (CISDTQ), and coupled cluster single-double perturbative triple (CCSD(T)), to name a few. In their appreciable level of flavor, they scale as  $O(N^{6–10})$  and are also applicable only to few atom systems. The Hartree–Fock (HF) mean-field approximation reduces the problem to a single electron moving in an average field of others with  $O(N^4)$  complexity, and serves as an upper bound for an energy representing an uncorrelated system of electrons.

Quantum chemistry is regarded as one of the first disciplines that will benefit from quantum computing. Predicting the properties of atoms and molecules will become one of the most important applications of quantum computers<sup>10–12</sup>. Specifically, quantum computing promises to revolutionize quantum simulations of molecules and materials by bringing down the intractable exponential cost on classical computers [ $O(\alpha^N)$ ] to a polynomial scaling on quantum computers. This can be achieved by using either of the two available forms of quantum computation: gate-based quantum computing and adiabatic quantum computing. Current quantum computers are hardware-limited in both the number and quality of qubits, qubit connectivity, presence of high noise levels, and the need for full error-correction<sup>13</sup>. Noisy gate-based quantum computers are currently available at the scale of ~50 to 70 qubits. Adiabatic quantum computers are available in the form of quantum annealers, such as the D-Wave 2000Q (with ~2048 qubits) and the new D-Wave Advantage machine (with ~5000 qubits). Quantum annealing architectures require a problem to be framed as an Ising Model or quadratic unconstrained binary optimization (QUBO) and allow for slow continuous transformation of an initial Hamiltonian to a target Hamiltonian settling into a low energy state. Qubit count and limited connectivity between qubits can restrict the size of the problem that can be run directly on the hardware, necessitating the use of quantum–classical algorithms when larger. Gate-based quantum architectures require that problems be transformed into a circuit of reversible gates applied to a given set of qubits prior to execution. The use of shallow depth circuits and variational quantum–classical algorithms mitigate the effects of gate fidelity, short coherence time, and noisy qubits. The D-Wave quantum annealer exhibits an even shorter coherence time, though sufficient for solving optimization problems.

Solutions to electronic structure problems have been demonstrated for small molecules on gate-based quantum computers<sup>14–17</sup> and the D-Wave 2000Q quantum annealer<sup>18–20</sup>. Polynomial scaling has not yet been demonstrated for the electronic structure problem on a quantum annealer. However, as we show in this work, the most advanced D-Wave 2000Q quantum annealer is well suited to solve combinatorial optimization problems, which can be used to reduce the complexity of the electronic structure problem. In this work we reduce the molecular Hamiltonian on a quantum annealer and then calculate the approximate ground and/or excited state energies on a classical computer through diagonalization. The advantages of solving the graph partitioning and community detection problem on a Quantum Annealer were recently demonstrated in<sup>3,4</sup>. The smaller sub-matrices resulting from the clusters of SDs are then candidates for the best low energy solution. The one containing the lowest energy SD (i.e. the HF SD) or resulting in the lowest gauge metric (described later) is the cluster that provides the best approximation of the original system. The energy is calculated through diagonalization on a classical computer. This typically results in energy many times within chemical accuracy ( $\leq 1$  kcal/mol or  $1.6e-03$  Hartrees).

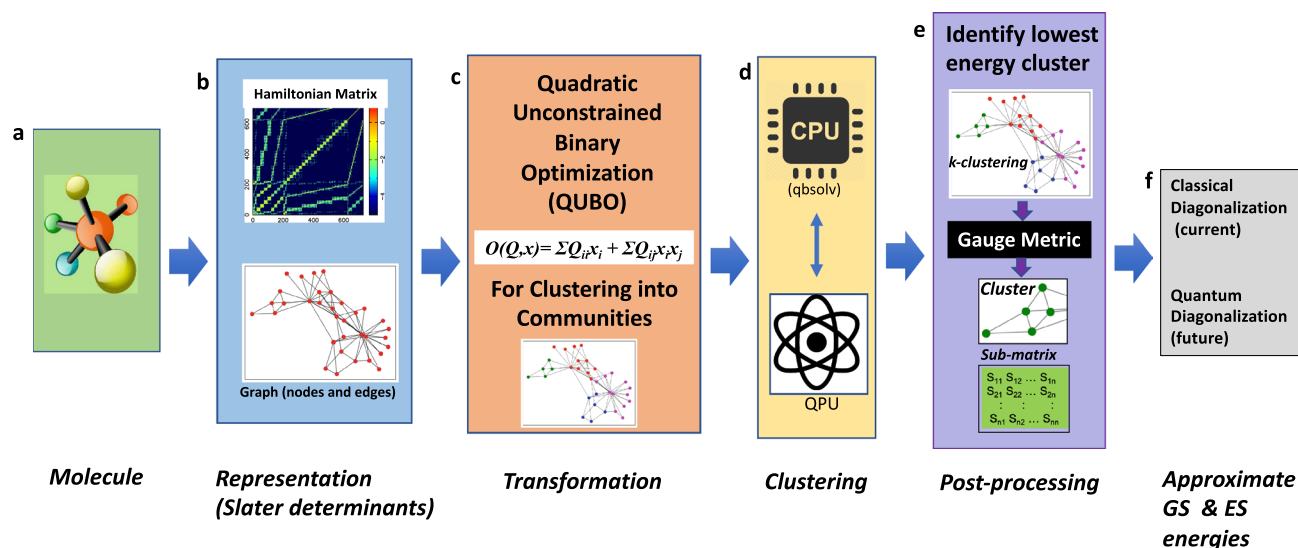
A quantum annealer, such as the D-Wave 2000Q<sup>21,22</sup>, is able to solve graph problems as combinatorial optimization formulated as an Ising model or QUBO problem. Quantum annealing uses the quantum-mechanical effects of tunneling, superposition and entanglement<sup>23</sup> to minimize and sample from energy-based models. Evidence that these effects play a useful role in the processing have been discussed in<sup>24–26</sup>. The following objective function in Ising Model form is minimized.

$$O(\mathbf{h}, J, s) = \sum_i h_i s_i + \sum_{i < j} J_{ij} s_i s_j \quad (1)$$

where  $s_i \in \{-1, +1\}$  encodes the binary results;  $h_i$  and  $J_{ij}$  represent the weights on the qubits and strengths on the couplers between qubits of the problem Hamiltonian. The D-Wave quantum computer is composed of qubits with sparse connectivity as a fixed sparse graph, known as a *Chimera* graph. Qubits are in a “superposition” state (both a “– 1” and a “+ 1” simultaneously) during the annealing process. Once the annealing is done, each qubit settles into an Ising spin value  $\in \{-1, +1\}$ , resulting in a low-energy ground state. Due to sparse connectivity, a logical variable maps to a chain of qubits<sup>26</sup> in the D-Wave quantum processing unit (QPU).

Quantum annealers are useful for tackling NP-hard complex problems including optimization, machine learning and sampling. Maximization can also be solved by using the negative of Eq. (1) as the objective function. The QUBO formulation where variables  $x_i$  take values of either 0 or 1 is an alternative representation. The QUBO and Ising Model are related by the following linear transformation:  $s = 2x - 1$ .

*Quantum Community Detection* is Community Detection<sup>27,28</sup> formulated as a QUBO problem and is run on the D-Wave quantum annealer. The community detection algorithm is an unsupervised machine learning technique that allows for the discovery of network substructure as tightly knit clusters or communities. The quantum version of this algorithm as a QUBO problem running on a quantum annealer uses superposition to explore the design space and settles into a low-energy solution. The molecule Hamiltonian matrix  $W$  is treated as an adjacency matrix  $A$  for a graph with weighted edges.



**Figure 1.** Conceptual workflow for the reduction process. (a) Cartesian co-ordinates for atoms, atomic numbers and a chosen basis set define a molecule. (b) An electronic Hamiltonian matrix is constructed in the basis of Slater Determinants (SDs) of spin-orbitals by using a classical quantum chemistry code. This matrix is further represented as a graph adjacency matrix where each SD is a node. (c) The graph is transformed to a Quadratic Unconstrained Binary Optimization (QUBO) form of *Community Detection*. (d) The QUBO is run using quantum-classical *qbsolv* + D-Wave quantum annealer. (e) The resulting clusters/communities as sub-matrices are candidates for the best low energy solution. The gauge metric is used to identify the lowest energy cluster. (f) Approximate ground state and excited state energies are calculated using diagonalization.

$$A_{ij} = 0, i = j$$

$$A_{ij} = |W_{ij}|, i \neq j$$

The modularity matrix  $B$  is constructed from the adjacency matrix  $A$  and degree of each node (number of edges connected to a node),  $g_i$ , where,  $2m = \sum_i g_i$ .

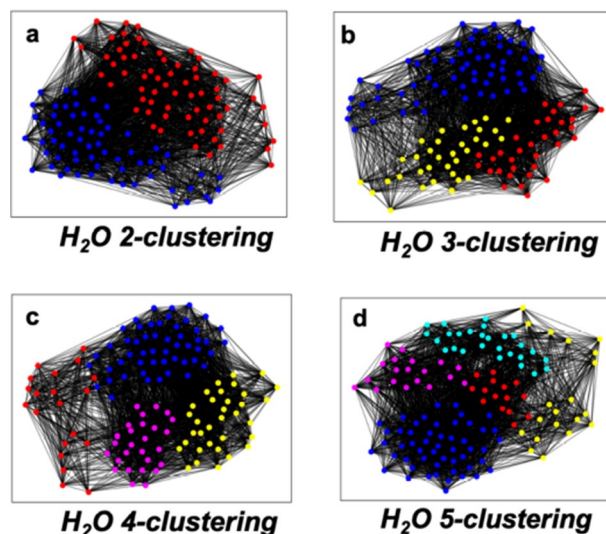
$$B_{ij} = A_{ij} - \frac{g_i g_j}{2m} = A_{ij} - \frac{g_i g_j}{\sum_i g_i}$$

The objective function for the optimal modularity  $Q$  as a QUBO is maximized as shown in Eq. (2). The elements of  $x$  are the solution and are binary, where  $x_i$  is 0 or 1. Details of the mathematical formulation for multiple communities have been developed previously and are available<sup>4</sup>. When solving for multiple communities, a *one hot encoding* or super-node concept is used with unary encoding. The objective function is shown as the following.

$$Q(x) = \max(x^T B x) \quad (2)$$

The D-Wave 2000Q is limited in the number of variables that can be embedded and solved directly on the hardware. Therefore, the largest fully connected graph that can be embedded is 65. A hybrid quantum-classical approach is required to address larger numbers of variables as nodes (here representing SDs)<sup>13</sup>. We use the D-Wave developed *qbsolv* classical solver<sup>29</sup> to orchestrate the QUBO solution process between the central processing unit (CPU) and QPU. Global minimization is performed by *qbsolv*, making multiple calls to the D-Wave to solve sub-QUBOs. This is followed by local minimization using tabu search. The D-Wave Ocean application programming interface (API) or the command line is used to initiate this process. A low-energy solution is returned as a bitstring of zeros and ones that requires translation based on the problem representation.

The conceptual workflow for this process is shown in Fig. 1. Starting from the Cartesian coordinates for atoms, atomic numbers and a chosen basis set defining a molecule (Fig. 1a), an electronic Hamiltonian matrix is constructed in the basis of SDs of spin-orbitals by using one of the classical quantum chemistry codes such as Psi4<sup>30</sup>. This matrix is represented as a graph (Fig. 1b) and is transformed into a QUBO form of *Community Detection* (Fig. 1c) to be solved on the D-Wave Quantum Annealer (Fig. 1d). Due to the matrix/graph size, a quantum-classical approach using *qbsolv* and the D-Wave is required. The resulting clusters or communities are then candidates for the best low-energy solution. The one containing the HF SD or identified from the gauge metric in the post-processing step (Fig. 1e) is the lowest energy cluster. The approximate ground state and excited state energies are calculated by diagonalization on a classical computer (Fig. 1f).



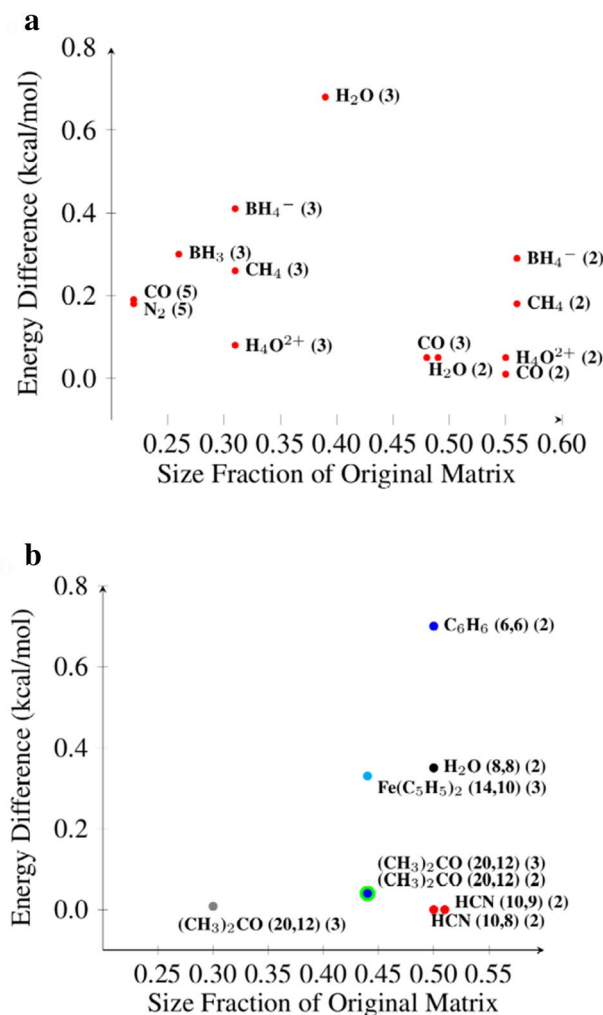
**Figure 2.** Showing H<sub>2</sub>O 2-, 3-, 4-, and 5-clustering from *Quantum Community Detection* on the D-Wave Quantum Annealer. For all the  $k$ -clustering, the cluster or community of nodes with the lowest energy is shown in blue. (a) The 2-clustering produced communities of size 68 (red) and 65 (blue). (b) The 3-clustering is shown with communities of size 34 (red), 65 (blue), and 34 (yellow). The 2- and 3-clustering share the same low energy cluster of 65. (c) The 4-clustering produced communities of size 20 (red), 52 (blue), 34 (yellow), and 27 (magenta). (d) The 5-clustering resulted in communities of size 15 (red), 52 (blue), 20 (yellow), 19 (magenta), and 27 (cyan). The 4- and 5-clustering share the same low energy cluster of 52. The clustering pictures were generated using NetworkX 2.5<sup>31</sup> (<https://networkx.org>) and Matplotlib 3.3.3<sup>32</sup> (<https://matplotlib.org>).

## Results

**The reduction process.** The reduction approach proceeds as follows. A molecular Hamiltonian  $\hat{H}$  matrix in the basis of SDs serves as input. The  $k$ -clustering sets of the Hamiltonian matrix are explored starting with  $k = 2$ . *Quantum Community Detection* performed using the D-Wave quantum annealer determines the  $k$  communities or clusters. Each of the  $k$  clusters translates into a sub-matrix of SDs. The one containing the HF SD (or identified by the gauge metric) is the cluster of interest out of the  $k$  clusters and yields the lowest energy,  $E_{CL}$ , when diagonalized classically. The results obtained by this procedure fall between the HF ( $E_{HF}$ ) and FCI energy of the full matrix ( $E_{FCI}$ ), with a tendency to be closer to  $E_{FCI}$ . The accuracy of the result can be measured as the difference with respect to  $E_{FCI}$ , (energy difference =  $E_{CL} - E_{FCI}$ ). This can be repeated for 3-, 4-, 5-clustering and more as needed. Each  $k$ -clustering is independent of the others. One does not need to go through the  $k$ -clustering in order, but it is natural to do so. Based on knowledge about a molecule, one may even choose to start from a large  $k$  and work backwards. The cluster with the lowest energy difference over all  $k$ -clustering sets is the best result from an accuracy perspective. This sub-matrix is the reduced matrix, constituting an approximate representation of the full Hamiltonian. The goal of this procedure is to discover a reduced matrix with an energy difference within chemical accuracy. The focus can be on accuracy or size. One should expect the larger the size, the higher the accuracy. Multiple clusters with energy differences within chemical accuracy may be found. Any of these is sufficient for ground state energy calculations. We note here that the reduction method is not limited to the energy and could potentially be extended to compute the atomic forces, for geometry optimization or molecular dynamics.

**Illustrating the method with H<sub>2</sub>O.** The Hamiltonian FCI matrix of the H<sub>2</sub>O molecule in the sto-3g minimal basis set is a  $133 \times 133$  sparse Hermitian (i.e., real symmetric) matrix after imposing the unitary groups U(1) and SU(2) (spin and particle conservation) and  $C_{2V}$  point group symmetries. The sparse matrix can be represented as a weighted graph of 133 nodes with 3032 edges. Each off-diagonal matrix element is considered an edge weight. Running *Quantum Community Detection* for 2-clustering produces the communities shown in Fig. 2a. The 65 nodes of the blue community represent the SD indices of the reduced matrix at this level with the lowest energy. The energy difference for this cluster is 0.05 kcal/mol for the lowest eigenvalue (ground state energy), which is achieved by almost a 2-fold decrease in matrix size. The 3-clustering shown in Fig. 2b produces two other communities and the blue community with the same ground state energy. A further reduction occurs with 4-clustering as shown in Fig. 2c. Classical diagonalization of this blue community of 52 SDs produces a ground state energy value with an acceptable energy difference of 0.68 kcal/mol. No further reduction is seen with 5-clustering, as shown in Fig. 2d. A significant 49% and 39% of the original molecule Hamiltonian matrix is retained with the 65 and 52 SD communities.

An excited state of a molecule is any electronic quantum state of the system that has a higher energy than the ground state. Table 1 also shows the energy differences for the H<sub>2</sub>O ground state and next five excited states for the reduced matrices based on the communities of size 65 (CL65) and 52 (CL52). Refer to the Supplementary Information for additional data. Chemically accurate energies for five excited states are shown for CL65 and



**Figure 3.** Comparison of energy difference and size reduction for different molecule clusters within chemical accuracy. **(a)** Molecule clusters using FCI and the sto-3g minimal basis set are shown. Those in the lower right corner from a  $k$ -clustering with  $k = 2$ , shown as “(2)”, are the best results from an accuracy perspective. The others have smaller size and less accuracy. **(b)** Molecule clusters using CASSCF and different basis sets are shown. These are all from a 2- or 3-clustering. In this case the accuracy is independent of size. The basis set used for each molecule is noted by color: sto-3g is red, cc-PVQZ is blue, def2-svp is gray, def2-TZVP is cyan, cc-PVDZ is black, and 6-31G\* is green.

only the first three for CL52. This indicates that the lower accuracy CL52 does not contain all the SDs necessary for higher excited state calculations.

**Molecules using FCI and the sto-3g minimal basis set.** Quantum chemists are predominantly interested in ground state energies. Molecular systems used to create FCI matrices in the sto-3g minimal basis set for further evaluation of the performance of *Quantum Community Detection* are shown in Fig. 3a. Refer to the Supplementary Information for additional data. H<sub>2</sub>O, CO, CH<sub>4</sub>, BH<sub>4</sub><sup>-</sup>, and H<sub>4</sub>O<sup>2+</sup> show a similar pattern for the  $k$ -clustering that gave the best low energy result. The 2-clustering produced a relevant cluster that gave good initial results (within chemical accuracy, i.e. < 1 kcal/mol), followed by a 3- or 4-clustering cluster that reduced the matrix size further at slightly lower accuracy. BH<sub>3</sub> and N<sub>2</sub> show a different pattern. A  $k$ -clustering with  $k > 2$  produced a cluster with greater reduction in size and still within chemical accuracy. Figure 3a shows the same molecules in comparison based on the size fraction of the original vs. the energy difference relative to FCI. Only the reduced molecule clusters resulting in energies within chemical accuracy are shown. Those from 2-clustering produced the most accurate results with ~50% reduction in size. Those from clusterings where  $k > 2$  produced clusters of smaller size, though still within chemical accuracy.

**Molecules using the CASSCF method.** The complete active space self-consistent field method (CASSCF) is a multi-configurational SCF method<sup>6–8</sup>. The active space consists of the electrons and orbitals that are necessary for a reliable description of the electronic structure of a molecule. CASSCF can be thought of as a

Excited state	$E_{CL65} - E_{FCI}$ (kcal/mol)	$E_{CL52} - E_{FCI}$ (kcal/mol)
0	0.05	0.68
1	0.06	0.19
2	0.09	0.49
3	0.02	0.51
4	0.04	33.15
5	0.13	17.07

**Table 1.** Approximate ground and excited states energies for H<sub>2</sub>O in the sto-3g basis.

truncation of FCI, which is based on a full active space within which all orbitals are considered. The total number of orbitals depends directly on the number and type of atoms in the system. Given a number of electrons (e) distributed across a number of orbitals (o), a Hamiltonian consisting of a linear combination of SDs or configuration state functions (CSFs) is obtained. Careful design of the active space requires chemical intuition based on the molecule of interest<sup>33,34</sup>. This method is useful when HF and DFT are not adequate (frequently the case of static electronic correlations), and FCI is computationally intractable. *Quantum Community Detection* could potentially be used for further optimal truncation of the initially chosen active space.

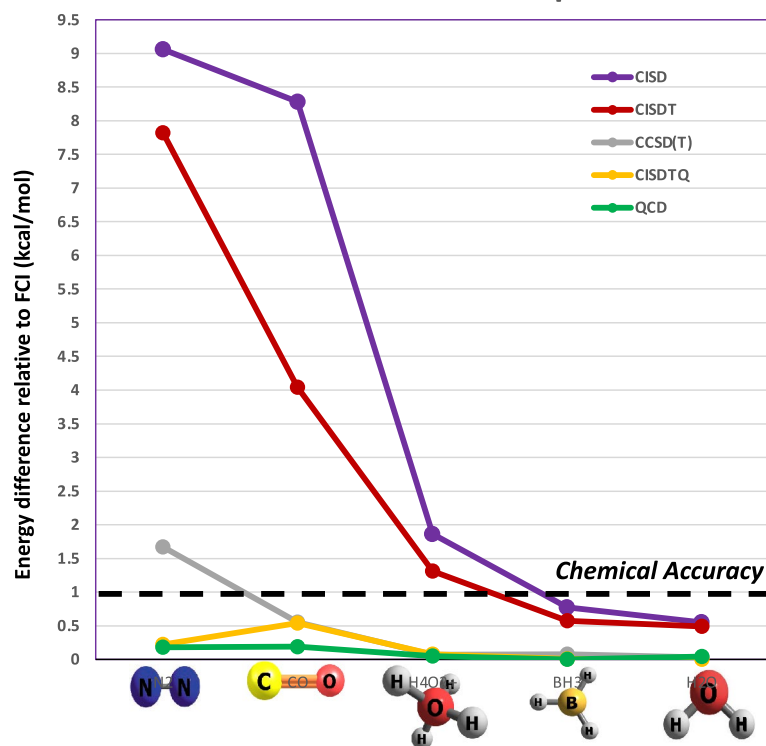
Molecules used to create the CASSCF matrices in different basis sets are shown in Fig. 3b. For HCN (hydrogen cyanide) and (CH<sub>3</sub>)<sub>2</sub>CO (acetone), the choice of active space was standard, based on the highest occupied molecular orbital as a reference point. For C<sub>6</sub>H<sub>6</sub> (benzene), Fe(C<sub>5</sub>H<sub>5</sub>)<sub>2</sub> (ferrocene) and C<sub>8</sub>H<sub>10</sub>N<sub>4</sub>O<sub>2</sub> (caffeine), the choice of active space was chemically inspired (see Supplementary Information). Acetone with different basis sets and HCN with different active spaces show the best accuracy. Clusters with less accuracy are seen for C<sub>6</sub>H<sub>6</sub>, Fe(C<sub>5</sub>H<sub>5</sub>)<sub>2</sub>, and H<sub>2</sub>O. The sizes, energies, and energy differences along with additional data are shown in the Supplementary Information. For most of these cases, chemical accuracy was achieved with the 2- or 3-*clustering* sets, as shown in Fig. 3b. Additionally, two molecules, HCN (10,9) using the 6-31G\* basis set and caffeine using the sto-3g basis set resulted in low energy clusters greater than chemical accuracy (see Supplementary Information). In the case of caffeine, expert chemical knowledge was used in choosing the original active space of 12 electrons across 9 orbitals and left no room for improvement. The gauge metric indicated that this could not be reduced further while retaining chemical accuracy (see Supplementary Information). This was further confirmed by *Quantum Community Detection* for 2-*clustering* which led to an energy difference of 24.36 kcal/mol.

**Comparison with other methods.** In Fig. 4, we compare *Quantum Community Detection* with commonly used current methods, CISD, CISDT, CCSD(T), and CISDTQ, all relative to FCI. The hierarchy of approximations consists of HF ⇒ CISD ⇒ CISDT ⇒ CCSD(T) ≈ CISDTQ ⇒ FCI. The inclusion of higher excitations leads to a larger expansion and produces an answer which is typically closer to the exact solution. The higher configuration interaction methods are more accurate, but at the expense of an increased problem size which scales steeply ( $N^6$  to  $N^{10}$ ). Figure 4 shows the energy differences for the five methods across the five molecules (N<sub>2</sub>, CO, H<sub>4</sub>O<sup>2+</sup>, BH<sub>3</sub>, and H<sub>2</sub>O) analyzed here. In this case, CISD, CISDT, and CCSD(T) produced results greater than chemical accuracy for one or more molecules. The size produced by *Quantum Community Detection* compared to CISDTQ and CCSD(T) is smaller or similar (see Supplementary Information). In general, for this small set of molecules, we can see that *Quantum Community Detection* produces results that are competitive with current methods, in line with CISDTQ and the gold standard method, CCSD(T).

**H<sub>3</sub>O<sup>+</sup> bond dissociation.** A bond dissociation energy experiment was performed for H<sub>3</sub>O<sup>+</sup> for O–H bond distances of 0.39–3.39 Å by enforcing C<sub>2v</sub> molecular symmetry. We simulated a constrained potential energy surface of an oxonium ion (H<sub>3</sub>O<sup>+</sup>) in the sto-3g basis (FCI method), as shown in Fig. 5a. Specifically, the constrained parameter was chosen as an O–H bond. Starting from the optimized geometry of H<sub>3</sub>O<sup>+</sup>, the latter was stretched/shortened by using a step of 0.2 Å in a constrained scan optimization (at HF level). The overall process can be viewed as a protonation of H<sub>2</sub>O. The optimized geometries were used to generate FCI matrices of size 3136. The *Quantum Community Detection* method resulted in reduced matrices of size 1576 for all geometries (see Fig. 5c), a 50% reduction. Their diagonalization provided approximate ground state energies within  $1.57 \times 10^{-5}$  kcal/mol relative to the original FCI matrices as seen in Fig. 5b. Refer to the Supplementary Information for data on all the clusterings performed.

**N<sub>2</sub> bond dissociation.** A bond dissociation energy experiment was performed for N<sub>2</sub> for distances of 0.5–2.4 Å. The ground state for each N<sub>2</sub> matrix was optimized using the restricted Hartree–Fock (RHF) method and the sto-3g basis set. The Hamiltonian matrices were generated using Psi4<sup>30</sup> along with the rHF and rFCI energies. Low energy communities were discovered for each bond distance using our *Quantum Community Detection* approach producing energies well within chemical accuracy as seen in Fig. 5d–f and the Supplementary Information. Interestingly, the  $E_{rCL} - E_{rFCI}$  values decrease with increasing bond distance. The reduced sub-matrices are all similar in size ranging from 363 to 397 SDs (see Fig. 5f) as compared to the original 1824 SDs, up to a 4-fold reduction in size. It is a well-known fact that RHF cannot qualitatively describe the electronic structure in a molecule undergoing homolytic bond breaking. N<sub>2</sub> has been widely used as a test molecule for many multi-reference approaches<sup>35</sup>. However, the bond-dissociation curve is quantitatively captured by the

## Molecule Method Comparison



**Figure 4.** Comparison of *Quantum Community Detection* with current methods. Results for CISD, CISDT, CCSD(T), CISDTQ, and our *Quantum Community Detection* (QCD) relative to FCI are shown for five molecules, N<sub>2</sub>, CO, H<sub>4</sub>O<sup>2+</sup>, BH<sub>3</sub>, and H<sub>2</sub>O. CISD (violet), CISDT (red) and CCSD(T) (gray) produced energy differences greater than chemical accuracy for one or more molecules. For these five molecules we can see that *Quantum Community Detection* produced results that are competitive with current methods, similar to CISDTQ and the gold standard, CCSD(T).

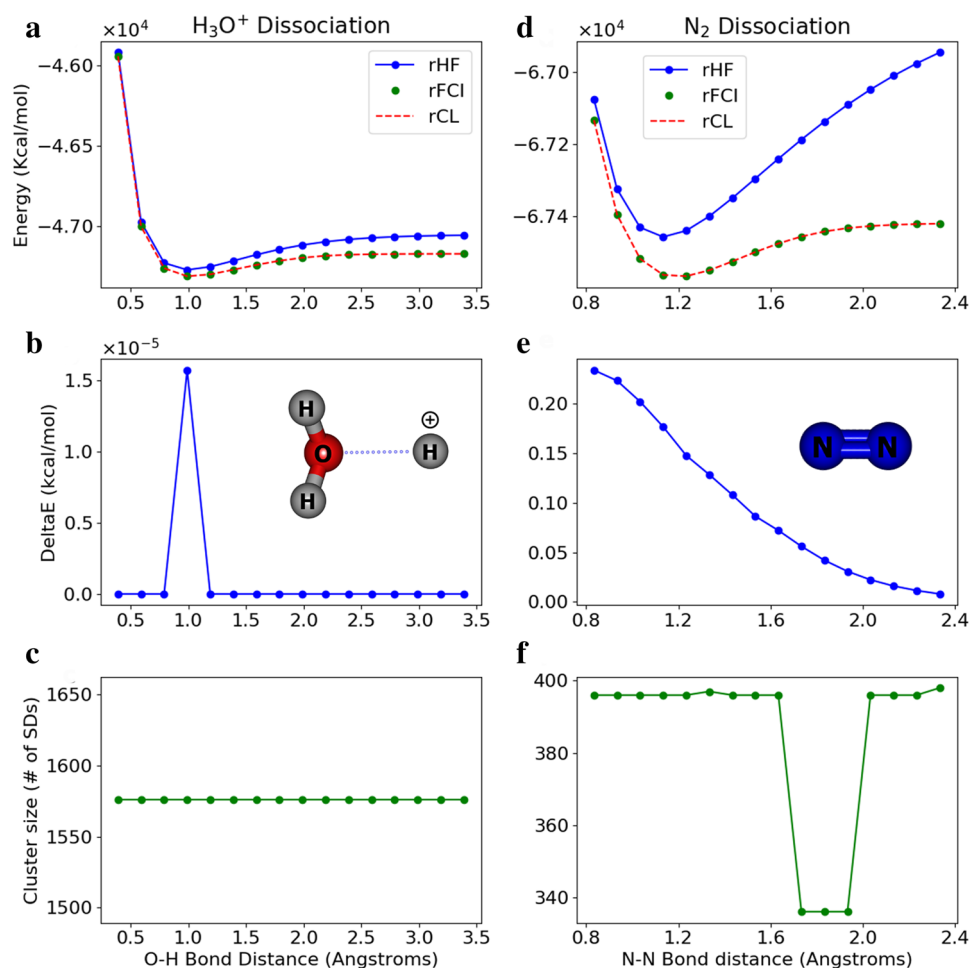
*Quantum Community Detection* method shown in Fig. 5d, which indicates that the method successfully selects the reduced sub-matrices that include all the important SDs to recover the correct electronic structure along the dissociation curve.

**Cluster analysis based on energy and connectivity.** The best low energy cluster for a molecule from *Quantum Community Detection* depends on the energy of the individual SDs in a cluster and the connectivity between them. Community detection determines the clusters by relying on the off-diagonal values of the Hamiltonian matrix as edge weights between SDs as the connectivity contribution.

We observe that the energy difference depends on the cluster, and it is not necessarily a monotonic function of the number of clusters within a *k-clustering*. Because of this, there is an optimal *k* after which, if *k* is further increased, important elements from the Hamiltonian matrix are lost, and the energy difference is further increased.

A metric to “gauge” the energy difference for the corresponding sub-matrix of each cluster of a particular *k-clustering* can be calculated involving SD energies (diagonal elements of the Hamiltonian matrix) and the weights between SDs (off-diagonal elements of the Hamiltonian matrix). Notably, the former quantity is related to the energies of spin-orbitals composing a given SD, which historically served as a criterion for selecting active spaces for truncated methods. Therefore, an analysis of the diagonal elements of the Hamiltonian may be useful to rationalize the *Quantum Community Detection* truncation. The cluster with the lowest metric will also result in the lowest energy and smallest energy difference for this set. This reduces the diagonalization calculations down to one for each *k-clustering*. The energy difference can be gauged with the approximations typically used in quantum chemistry, such as perturbation theory or energy re-normalization<sup>36</sup>.

Following this same idea, we correct the energy by considering the interaction of multiple states. In Fig. 6a we can see a network representation of an SD (node) interacting with many others within a network representation and within an energy representation. In Fig. 6b, the energy representation, we show how the energy of node 0 (diagonal element  $\epsilon_0$ ) gets re-normalized (corrected) to  $\tilde{\epsilon}_0$  due to the interaction with another node via the coupling elements or weights  $V_{0j}$ . The gauge metric  $f$  calculated across all SDs in a cluster is shown in Eq. (3). A value of  $\delta$  is added to the denominator to avoid a division by 0. This parameter is typically known as “phase shift” or “denominator shift” within the context of many-body methods<sup>37,38</sup>. The value of  $\delta$  will need to be determined for every particular problem. For the cases analyzed in this manuscript, a value of  $\delta = 1.0$  not only shows monotonicity with the energy difference, it is also linear. Refer to the Supplementary Information for a detailed derivation.



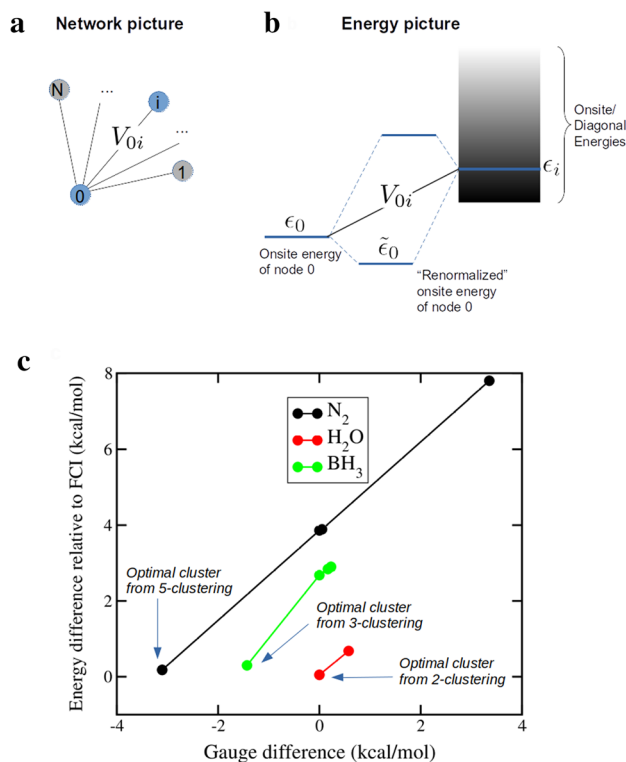
**Figure 5.** Bond dissociation for  $\text{H}_3\text{O}^+$  and  $\text{N}_2$ . (a) Showing  $\text{H}_3\text{O}^+$  bond dissociation using the reduced sub-matrices discovered using *Quantum Community Detection*. The lowest energy bond distance is 0.99 Å. The FCI and cluster (CL) energies across all distances are nearly the same. (b) Most of the energy differences are within  $10^{-10}$  kcal/mol across all O-H bond distances, though the lowest energy point is within  $10^{-5}$  kcal/mol. (c) The reduced sub-matrix size for all bond distances is 1576 (from low energy clusters). This is a 50% reduction from the original matrices of size 3136 SDs. (d) Showing potential energy curve for  $\text{N}_2$  bond dissociation. (e)  $\text{N}_2$  energy differences decrease with larger bond distance. (f) Reduced sub-matrix sizes for  $\text{N}_2$  vary from 363 to 397 SDs out of 1824 across available bond distances.

$$f = \min_i \left( \epsilon_{ii} - \sum_k \frac{V_{ik}^2}{|\epsilon_{ii} - \epsilon_{kk}| + \delta} \right) \quad (3)$$

The metric  $f$  can be used to determine the relevant cluster of a  $k$ -clustering without having to compute  $k$  diagonalizations. Although we currently have a partial understanding of why there is no clear decreasing monotonicity between  $k$  and the accuracy of the selected cluster; the explanation of the differences in accuracy between clusters are fully contained in the factors of this gauge metric  $f$ . This metric is derived and analyzed further in the Supplementary Information.

Looking at the pair-wise differences in the metric  $f$  values of the best clusters across multiple  $k$ -clusterings, we see that the lowest difference determines the cluster with the best accuracy for a given molecule. This relationship between the metric  $f$  differences and the energy differences is monotonic and linear as shown in Fig. 6c for  $\text{H}_2\text{O}$ ,  $\text{N}_2$ , and  $\text{BH}_3$ . The most accurate cluster for  $\text{H}_2\text{O}$  was discovered from the 2-clustering with a small energy difference of 0.05 kcal/mol. This remained the same for 3-clustering resulting in a metric  $f$  difference of 0.0. For  $\text{N}_2$ , the smallest energy difference of 0.18 kcal/mol and a low metric  $f$  difference between the 4- and 5-clustering confirm that the most accurate cluster comes from the 5-clustering. Similarly, for  $\text{BH}_3$ , the smallest energy difference of 0.30 kcal/mol and lowest metric  $f$  difference between the 3- and 2-clustering indicate that the most accurate cluster is from the 3-clustering. These results are shown in the Supplementary Information. The most accurate cluster includes low energy SDs (the lower the better), which are close in energy, with high weights between the SDs (as nodes). The gauge thus provides a quick test if the inclusion of low energy SDs is sufficient





**Figure 6.** Gauge metric description. **(a)** Network representation of a cluster. **(b)** Scheme of how the energy of node 0 (diagonal element  $\epsilon_0$ ) gets re-normalized (corrected) to  $\tilde{\epsilon}$  due to the interaction with the other nodes (SDs) via the coupling elements or weights  $V_{0i}$ . **(c)** Gauge and energy differences are shown for  $H_2O$ ,  $BH_3$ , and  $N_2$ . The lowest gauge metric differences and corresponding energy differences denote the most accurate cluster. Pair-wise differences in the gauge metrics and the corresponding energy differences have a linear relationship as shown. The most accurate cluster across all  $k$ -clusterings is noted.

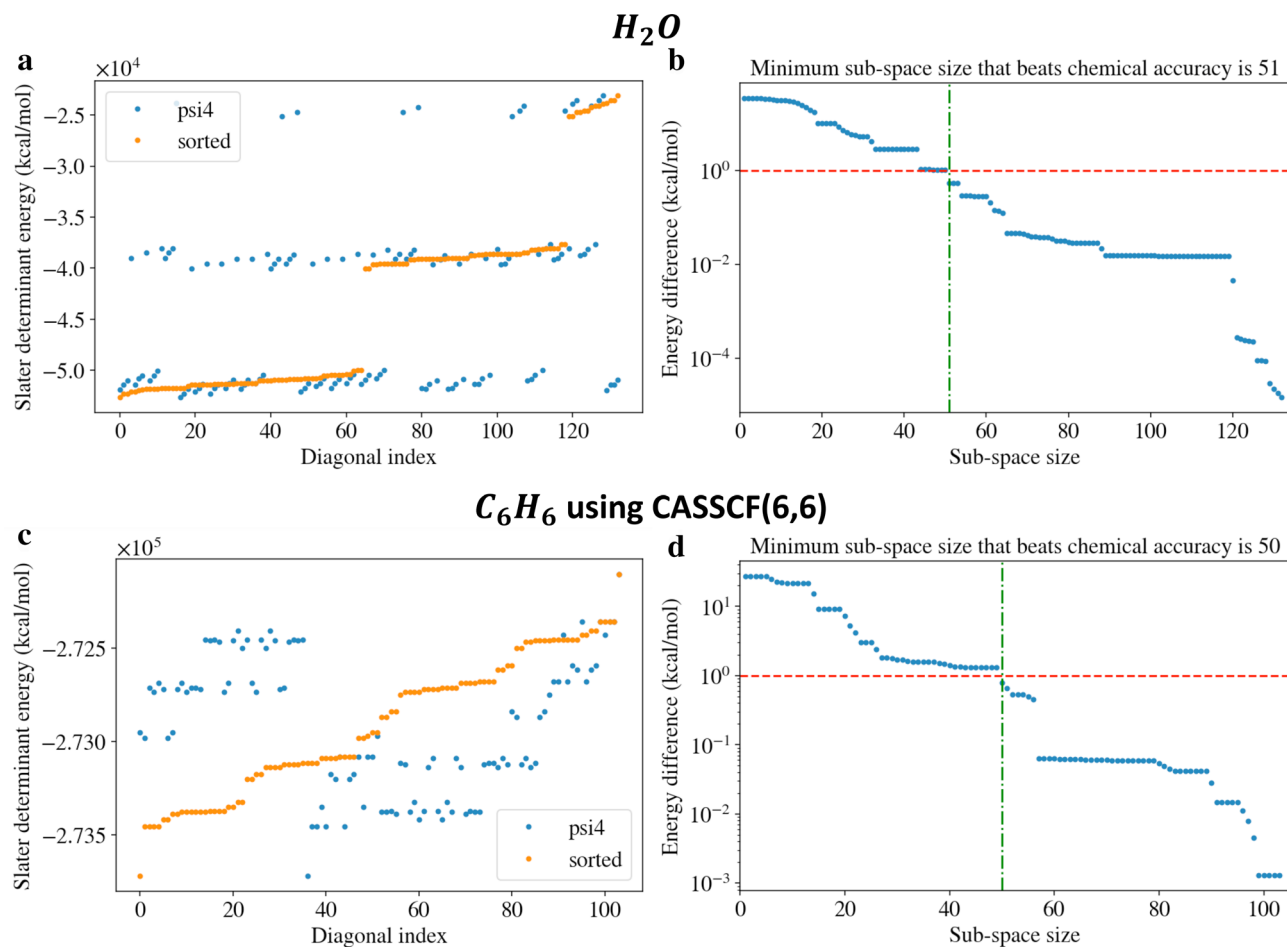
for a given  $k$ . The gauge can also be used to indicate whether a low energy cluster will result in chemical accuracy or not (see Supplementary Information).

Could better clusters be formed from the lowest energy SDs only? Arranging the SDs in sorted order and making “brute force” diagonalization calculations for clusters of increasing size, showed that this approach works for some cases. In Fig. 7, we show the results for  $H_2O$  and  $C_6H_6$  from a sorted energy perspective. Sorting the  $H_2O$  SDs from low to high energy clearly shows three energy levels (see Fig. 7a). The minimum number of SDs required to be within chemical accuracy (just less than 1.0 kcal/mol) is 51 (see Fig. 7b). In comparison, *Quantum Community Detection* results in 65 SDs for high accuracy and 52 SDs for lower accuracy. Considering  $C_6H_6$ , the sorted SDs are somewhat separated in energy as seen in Fig. 7c. The minimum number of SDs to be within chemical accuracy is 50, just below 1.0 kcal/mol as seen in Fig. 7d. *Quantum Community Detection* results in 52 SDs, with an energy delta of 0.70 kcal/mol and 28 SDs for an energy delta of 0.72 kcal/mol. Within a  $k$ -clustering, the majority of the SDs included in the lowest energy cluster are of low energy. However, a more complex interplay between different SDs could further lower the energy as can be deduced from the factors of the metric  $f$ .

## Discussion

The solution of the Schrödinger equation lies at the heart of computational chemistry and is usually accomplished by numerical diagonalization of the Hamiltonian matrix to obtain energies of the ground and excited states. Over decades, exponential hardness of the original problem was eased in practice by various many-body approximations allowing for reduction of the Hamiltonian size. Availability of quantum computing hardware provides new and unconventional means to solve traditionally difficult numerical problems. While the majority of quantum efforts have been devoted to turning exponential-to-polynomial scaling of numerical expenses, automatic reduction of the Hamiltonian size on the quantum hardware remains an important algorithmic development with practical implications. Here we have introduced a new approach for reducing the molecular Hamiltonian matrix for electronic structure calculations using *Quantum Community Detection* running on the D-Wave quantum annealer. The algorithm automatically samples the Hamiltonian space and identifies clusters allowing for accurate approximation of the lowest eigenvalue of the Hamiltonian through diagonalization that corresponds to the ground state energy of the molecule.

We have applied our *Quantum Community Detection* method to a variety of different molecular systems of different sizes and chemical properties. For several cases, the resulting clusters produced ground state energies within chemical accuracy ( $\leq 1$  kcal/mol) providing size reductions of 50% or more, and serve as good



**Figure 7.** The sorted energy approach is demonstrated for H<sub>2</sub>O and C<sub>6</sub>H<sub>6</sub>. **(a)** The SDs for H<sub>2</sub>O in sorted order are shown displaying distinct energy levels. **(b)** The minimum number of SDs required to be within chemical accuracy, 51, is shown. **(c)** The sorted SDs for C<sub>6</sub>H<sub>6</sub> are shown, somewhat separated in energy. **(d)** The minimum number of SDs within chemical accuracy, 50, is shown.

approximations to the original larger problems (i.e. calculations at full configuration interaction level). We note here that the approach is not variational, however, and does not guarantee a reduction within chemical accuracy. Though the gauge metric can be used to indicate whether reduction is possible and if the reduced sub-matrix will be within chemical accuracy.

Beyond exploring chemical diversity, we have also demonstrated application to two chemical reaction/conformational dynamics examples. Bond dissociation of H<sub>3</sub>O<sup>+</sup> achieved a ~ 50% reduction in matrix size while maintaining chemical accuracy. While N<sub>2</sub> bond dissociation required only ~ 25% of the original matrix size, also within chemical accuracy.

We further have shown connections to the traditional approaches of Hamiltonian reduction via selecting molecular orbitals that provide low-energy Slater Determinants (SD). Our analysis shows that both the energies of the SDs and the connectivity between them are important in determining the best approximate representation of a molecule within chemical accuracy for electronic structure calculations. The high accuracy clusters overall include mostly low energy SDs (the lower the better), which are close in energy. However, they may also include high energy SDs as well and the connectivity or weights between the SDs (as nodes) is typically high. The *Quantum Community Detection* approach was generally able to reduce the size of the molecular Hamiltonian without chemical knowledge or intuition. For the case study of five molecules, the algorithm resulted in an efficient reduction of Hamiltonian space toward reaching chemical accuracy, which is competitive with the current traditional approaches achieving reduction by limiting configuration interaction expansion (i.e. CISD, CISDT, CCSD(T), and CISDTQ levels).

Similar to selected configuration interaction (sCI)<sup>39,40</sup>, the method presented in this paper is able to prune the space of all directly interacting states in the Hamiltonian matrix as a graph to account for the most important ones that are representative of the original. The initial FCI matrix still needs to be constructed and the resulting reduced sub-matrix still needs to be diagonalized. In our proof-of-principle paper, they were required to gain an understanding of the results and provide ground truth comparisons. Working with larger molecular systems with potentially larger FCI matrices we will not have this luxury and will need to alleviate this by closer integration with the FCI matrix construction producing a reduced sub-matrix as part of the process. Techniques such

as sCI, incremental FCI construction<sup>41</sup>, and Monte Carlo CI methods<sup>42</sup> that operate on batches may be helpful here. Diagonalization can be avoided by using perturbation-based methods. The *Quantum Community Detection* reduction method could be used as part of sCI or as a post-processing step to further reduce the matrix. It could also be used as a pre-conditioner combined with other methods such as density matrix renormalization group (DMRG)<sup>43</sup>, where a reduced Hamiltonian sub-matrix could improve the process. These ideas will be explored in future studies.

*Quantum Community Detection* formulated for the D-Wave quantum annealer is a promising algorithmic development making use of emerging quantum hardware toward solving important quantum chemical problems. This study reports proof-of-principle concept, while practical utility of this algorithm requires further investigation. This motivates further studies targeting exploration of deeper insights into the relevance of quantum communities and energetics of SDs and molecular orbitals. Analysis of the molecular orbitals contributing to the communities will allow for uncertainty quantification and construction of algorithms gradually expanding an active space as well as reduction.

## Methods

**Molecular Hamiltonian matrix preparation.** All molecular geometries were first optimized at the restricted HF level (gas) with Gaussian 09 (E01) code<sup>44</sup>. The FCI and CASSCF matrices were generated using the in-house modified Psi4 code<sup>30</sup> with the optimized Cartesian coordinates of atoms as input data. The matrices were generated imposing the unitary groups U(1) and SU(2) (spin and particle conservation) and point group symmetries and contain nonzero matrix elements only. If optimized by Gaussian 09, the molecule belongs to a non-Abelian point group, Psi4 automatically lowers the symmetry to one of its subgroups. The energy threshold of  $10^{-8} E_h$  for diagonal and of  $10^{-10} E_h$  for non-diagonal elements was set up. The nuclear repulsion (NR) term was added manually for clustered matrices.

**D-Wave 2000Q setup.** The D-Wave 2000Q Ising resource at Los Alamos National Laboratory was used for this project. The solver used was *DW\_2000Q\_LANL* with 2032 active qubits (out of 2048) and 5924 active couplers (out of 6016). Default parameter settings were used for all runs (with an annealing time of 20 microseconds). All runs were quantum-classical using D-Wave's *qbsolv* with a sub-qubo size of 65 for the D-Wave quantum annealer.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 17 December 2020; Accepted: 4 February 2021

Published online: 18 February 2021

## References

- Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- Biamonte, J., Faccin, M. & Domenico, M. D. Complex networks from classical to quantum. *Commun. Phys.* **20**, 1–10 (2019).
- Ushijima-Mwesigwa, H., Negre, C. F. A. & Mniszewski, S. M. Graph partitioning using quantum annealing on the D-Wave system. In *Proceedings of the Second International Workshop on Post Moores Era Supercomputing (PMES'17)*, 22–29 (ACM New York, NY, USA, 2017).
- Negre, C. F. A., Ushijima-Mwesigwa, H. & Mniszewski, S. M. Detecting multiple communities using quantum annealing on the D-Wave system. *PLoS One* **15**, e0227538 (2020).
- Gan, Z. & Harrison, R. J. In calibrating quantum chemistry: A multi-teraflop, parallel-vector, full-configuration interaction program for the cray-x1. In *Proceedings of the ACM/IEEE DC 2005 Conference*, 22–22 (ACM, 2005).
- Roos, B. O., Taylor, P. R. & Siegbahn, P. E. M. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **48**, 157–173 (1980).
- Siegbahn, P. E. M., Heiberg, A., Roos, B. O. & Levy, B. Comparison of the super-CI and the Newton–Raphson scheme in the complete active space SCF method. *Phys. Scripta* **21**, 323–327 (1980).
- Siegbahn, P. E. M., Almlöf, A., Heiberg, A. & Roos, B. O. The complete active space SCF (CASSCF) method in a Newton–Raphson formulation with application to the HNO molecule. *J. Chem. Phys.* **74**, 2384–2396 (1981).
- Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, vol 1st edn. (Dover Publications, Inc., Mineola, 1996).
- Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
- McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **192**, 015003 (2020).
- Cao, Y. *et al.* Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856–10915 (2019).
- Shaydulin, R. *et al.* A hybrid approach for solving optimization problems on small quantum computers. *IEEE Comput.* **52**, 18–26 (2019).
- Kandala, A. *et al.* Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
- Nam, Y. *et al.* Ground-state energy estimation of the water molecule on a trapped ion quantum computer. arXiv:19 1–10 (2019) (arXiv preprint).
- O'Malley, P. J. J. *et al.* Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016).
- Aspuru-Guzik, A., Anthony D. Dutoi, A. D., Love, P. J. & Head-Gordon, M. Simulated quantum computation of molecular energies. *Science* **309**, 1704–1707 (2005).
- Genin, S. N., Ryabinkin, I. G. & Izmaylov, A. F. Quantum chemistry on quantum annealers. arXiv:1901.04715 (arXiv preprint) 1–10 (2019).
- Xia, R., Bian, T. & Kais, S. Electronic structure calculations and the Ising hamiltonian. *J. Phys. Chem.* **122**, 3384–3395 (2018).
- Streif, M., Neukart, F. & Leib, M. Solving quantum chemistry problems with a D-Wave quantum annealer. In *Proceedings of the International Workshop on Quantum Technology and Optimization Problems*, 111–122 (Springer, Cham, 2019).

21. D-Wave. The D-Wave 2000Q system. <https://www.dwavesys.com/d-wave-two-system>.
22. D-Wave. The D-Wave quantum computer technology overview. [https://www.dwavesys.com/sites/default/files/D-Wave%202000Q%20Tech%20Collateral\\_0117E.pdf](https://www.dwavesys.com/sites/default/files/D-Wave%202000Q%20Tech%20Collateral_0117E.pdf) (2017).
23. McGeoch, C. C. Adiabatic quantum computation and quantum annealing: Theory and practice. *Synth. Lect. Quantum Comput.* **5**, 1–93 (2014).
24. Albash, T., Hen, I., Spedalieri, F. M. & Lidar, D. A. Reexamination of the evidence for entanglement in the d-wave process. *Phys. Rev. A* **92**, 062328 (2015).
25. Boixo, S. *et al.* Computational multiqubit tunnelling in programmable quantum annealers. *Nat. Commun.* **7**, 20 (2016).
26. Lanting, T. *et al.* Entanglement in a quantum annealing processor. *Phys. Rev. X* **4**, 021041 (2014).
27. Newman, M. E. J. Modularity and community structure in networks. *PNAS* **103**, 8577–8582 (2006).
28. Newman, M. E. J. Analysis of weighted networks. *Phys. Rev. E* **70**, 056131 (2004).
29. Booth, M., Reinhardt, S. P. & Roy, A. Partitioning optimization problems for hybrid classical/quantum execution. *D-Wave Tech. Rep. Ser.* **14**, 1–9 (2017).
30. Parrish, R. M. *et al.* Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *J. Chem. Theory Comput.* **13**, 3185–3197 (2017).
31. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, SciPy 2008, 11–16 (ACM, 2008).
32. Hunter, J. D. Matplotlib: A 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95. <https://doi.org/10.1109/MCSE.2007.55> (2007).
33. Stein, C. J. & Reiher, M. Automated selection of active orbital spaces. *J. Chem. Theory Comput.* **12**, 1760–1771. <https://doi.org/10.1021/acs.jctc.6b00156> (2016).
34. Keller, S., Boguslawski, K., Janowski, T., Reiher, M. & Pulay, P. Selection of active spaces for multiconfigurational wavefunctions. *J. Chem. Phys.* **142**, 244104. <https://doi.org/10.1063/1.4922352> (2015).
35. Lyakh, D. I., MusiaÅ, M., Lotrich, V. F. & Bartlett, R. J. Multireference nature of chemistry: The coupled-cluster view. *Chem. Rev.* **112**, 182–243. <https://doi.org/10.1021/cr2001417> (2012).
36. Pastawski, H. M. & Medina, E. Tight Binding methods in quantum transport through molecules and small devices: From the coherent to the decoherent description. [arXiv:cond-mat/0103219](https://arxiv.org/abs/cond-mat/0103219) (arXiv preprint) 1–23 (2001).
37. Kelly, H. P. *Applications of Many-Body Diagram Techniques in Atomic Physics* Vol. 4, 129–190 (Wiley, New York, 1969). <https://doi.org/10.1002/9780470143599.ch4>.
38. Bartlett, R. J. & Silver, D. M. Pair-correlation energies in sodium hydride with many-body perturbation theory. *Phys. Rev. A* **10**, 1927–1931. <https://doi.org/10.1103/PhysRevA.10.1927> (1974).
39. Garniron, Y. *et al.* Quantum package 2.0: An open-source determinant-driven suite of programs. *J. Chem. Theory Comput.* **15**, 3591–3609 (2019).
40. Abraham, V. & Mayhall, N. J. Selected configuration interaction in a basis of cluster state tensor products. *J. Chem. Theory Comput.* **16**, 6098–6113 (2020).
41. Zimmerman, P. M. Incremental full configuration interaction. *J. Chem. Phys.* **146**, 104102 (2017).
42. Prentice, A. W., Coea, J. P. & Patersonb, M. J. A systematic construction of configuration interaction wavefunctions in the complete ci space. *J. Chem. Phys.* **151**, 164112 (2019).
43. Wouters, S. & Van Neck, D. The density matrix renormalization group for ab initio quantum chemistry. *Eur. Phys. J. D* **68**, 272 (2014).
44. Frisch, M. J. *et al.* Gaussian 09 revision e.01 (2009).

## Acknowledgements

Research presented in this article was supported by the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory (LANL) under project number 20200056DR. This research was also supported by the U.S. Department of Energy (DOE) National Nuclear Security Administration (NNSA) Advanced Simulation and Computing (ASC) program at LANL. We acknowledge the ASC program at LANL for use of their Ising 2000Q quantum computing resource. LANL is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218NCA000001). Assigned: Los Alamos Unclassified Report LA-UR-20-26971.

## Author contributions

S.M.M. conceived and guided the work, P.A.D. and S.M.M. designed the experiments, P.A.D. and Y.Z. generated the molecular Hamiltonians, S.M.M. ran the calculations, C.F.A.N. provided the gauge analysis, P.A.D., S.T., P.M.A., Y.Z., and S.M.M. provided other analysis. S.M.M. wrote the initial draft of the manuscript, all authors contributed to the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83561-x>.

**Correspondence** and requests for materials should be addressed to S.M.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021