Identifying challenges towards practical quantum advantage through resource estimation: the measurement roadblock in the variational quantum eigensolver

Jérôme F. Gonthier,[1] Maxwell D. Radin,[1] Corneliu Buda,[2]
Eric J. Doskocil,[2] Clena M. Abuan,[3] and Jhonathan Romero[1]

[1]*Zapata Computing, Inc., 100 Federal St., Boston, MA 02110, USA*
[2]*Applied Chemistry and Physics Centre of Expertise, BP Group Research, 150 West Warrenville Road, Naperville, IL 60563, USA*
[3]*Digital Science and Engineering, BP Innovation and Engineering, 501 Westlake Park Blvd, Houston, TX 77079, USA*

(Dated: December 9, 2020)

Recent advances in Noisy Intermediate-Scale Quantum (NISQ) devices have brought much attention to the potential of the Variational Quantum Eigensolver (VQE) and related techniques to provide practical quantum advantage in computational chemistry. However, it is not yet clear whether such algorithms, even in the absence of device error, could achieve quantum advantage for systems of practical interest and how large such an advantage might be. To address these questions, we have performed an exhaustive set of benchmarks to estimate number of qubits and number of measurements required to compute the combustion energies of small organic molecules to within chemical accuracy using VQE as well as state-of-the-art classical algorithms. We consider several key modifications to VQE, including the use of Frozen Natural Orbitals, various Hamiltonian decomposition techniques, and the application of fermionic marginal constraints. Our results indicate that although Frozen Natural Orbitals and low-rank factorizations of the Hamiltonian significantly reduce the qubit and measurement requirements, these techniques are not sufficient to achieve practical quantum computational advantage in the calculation of organic molecule combustion energies. This suggests that new approaches to estimation leveraging quantum coherence, such as Bayesian amplitude estimation [1, 2], may be required in order to achieve practical quantum advantage with near-term devices. Our work also highlights the crucial role that resource and performance assessments of quantum algorithms play in identifying quantum advantage and guiding quantum algorithm design.

I. INTRODUCTION

In the last decade, quantum computers have evolved from laboratory prototypes of a few qubits to machines with tens of qubits that are commercially available for researchers and businesses to use [3, 4]. Google recently announced the realization of the quantum supremacy milestone: their 53-qubit chip accomplished a specific task that would be extremely difficult to simulate with a classical supercomputer [5]. This task was specifically designed to be well suited to the quantum processor and challenging for classical computers, and does not solve a practical problem. The next milestone, and arguably the most pressing one [6], is finding a *practical quantum advantage* with noisy intermediate-scale quantum (NISQ) devices [7], that is, running an algorithm on a NISQ device that provides an improved solution for a commercially relevant task. This improvement can manifest in different ways, either as a reduction in the time to solution or an increase in the quality of the solution. Accomplishing this goal requires first a steady improvement in the quality of quantum computing hardware. Fortunately, we are witnessing a rapid growth in the number of qubits and fidelity of these machines as indicated by the recent trends in metrics such as quantum volume [8]: in less than a year this went from 32 in January, [1] to 64 in June,[2] 128 in September[3], and finally 4 million in October (still awaiting experimental confirmation).[4] The second requirement towards quantum advantage is identifying commercially relevant tasks for which a near-term quantum algorithm can provide a measurable improvement compared to classical alternatives.

Quantum chemistry has been identified as a likely candidate [9–11] for quantum advantage for multiple reasons. First, electronic structure calculations are used extensively in the development of many technologies, for example in the chemicals industry [12], drug development [13], and battery materials research [14]. Second, electronic structure calculations rely on the Schrödinger equation, for which a general exact solution has exponential cost on a classical computer with all known classical methods. Third, quantum computers can store exponentially scaling representations of the

[1] https://www.ibm.com/blogs/research/2020/01/quantum-volume-32/
[2] https://www.ibm.com/blogs/research/2020/01/quantum-volume-32/ and https://newsroom.ibm.com/2020-08-20-IBM-Delivers-Its-Highest-Quantum-Volume-to-Date-Expanding-the-Computational-Power-of-its-IBM-Cloud-Accessible-Quantum-Computers
[3] https://www.honeywell.com/en-us/newsroom/news/2020/09/achieving-quantum-volume-128-on-the-honeywell-quantum-computer
[4] https://ionq.com/news/october-01-2020-most-powerful-quantum-computer

wavefunction on a linear number of qubits and also provide means to implement Hamiltonian evolution efficiently, giving rise to quantum algorithms to estimate ground state energies of some molecular Hamiltonians using polynomially scaling resources.

Quantum approaches to electronic structure calculations can be divided into two categories: 1) algorithms based on the Quantum Phase Estimation subroutine and related techniques and 2) quantum heuristic algorithms [10], such as the variational quantum eigensolver (VQE) [15] and related methods based on different versions of the time-dependent variational principle [11]. Under certain assumptions, approaches in the first category can provide an advantage in computational scaling compared to exact classical algorithms, however they require a fault-tolerant implementation [9, 16–20], and therefore are not applicable in the near-term. In contrast, quantum heuristics such as VQE can be implemented on NISQ devices thanks to the flexibility in their construction, but do not provide any proven asymptotic advantage compared to classical algorithms. Demonstrating a quantum advantage in this context requires a comparison of the computational cost and performance of the quantum heuristic against the state-of-the-art classical approximations for specific problem instances. Performance metrics might include aspects such as total runtime and accuracy of the solution. In addition, this comparison must go hand in hand with an optimization of the algorithmic choices of the heuristic to maximize performance. From this perspective, a necessary step towards achieving quantum advantage in the near term is establishing protocols for evaluating quantum computational resources for specific sets of problem instances and target accuracy of the solution, a procedure we will refer to as resource and performance assessment (RPA).

While many studies have estimated resource requirements for quantum chemistry using fault-tolerant algorithms such as Quantum Phase Estimation [16–21], only a small number have assessed the resource requirements for NISQ approaches. McLean *et al.* has analyzed the asymptotic measurement requirements of VQE [22] while Kühn *et al.* numerically examined qubit requirements and required circuit depth for UCC-derived ansatzes [23]. Numerical studies have explored the VQE measurement costs of diatomic molecules and hydrogen chains/rings when applying fermionic marginal constraints and a low-rank factorization of the Hamiltonian [24, 25]. Elfving et al. estimated the number spin-orbitals required for industrially-relevant calculations and concluded that the required VQE execution time was prohibitively large, although the method used to estimate the execution time was not specified [21].

While these studies provide valuable insight into the performance of VQE and its variants, several questions about the feasibility of these techniques for systems of practical interest remain unanswered. For example, although VQE measurement requirements have been assessed for small basis sets, such analysis has not been carried out for the basis sets required to achieve a useful accuracy with respect to the infinite basis set limit. Another key question is how these measurement requirements compare to state-of-the-art classical quantum-chemistry techniques. Furthermore, previous studies have employed canonical orbitals and not considered Frozen Natural Orbitals (FNO) [26–28], which are known to significantly reduce the computational cost of classical wavefunction-based quantum chemistry methods.

To address these questions, we have performed an RPA to estimate the number of qubits, number of measurements, and total runtime required for calculating combustion energies for small organic molecules to within chemical accuracy with a single VQE energy evaluation. These estimates consider Frozen Natural Orbitals as well as measurement reduction techniques such as Hamiltonian term grouping [29–33], the application of fermionic marginal constraints [24], and low-rank factorization of the Hamiltonian [25, 34].

Our results indicate that between 120 and 260 qubits are required for chemical accuracy (error of at most 1.6 mHa [35]) for our benchmark systems. Under optimistic assumptions about the ansatz requirements and the sampling rate of the device, we show that a single energy evaluation could take several days to weeks, rendering the calculations impractical and inferior to classical methods, in particular when considering the large number of such evaluations required for the optimization loop of VQE. Our results also show that although certain grouping techniques greatly reduce the number of measurements, they are not enough to guarantee practical runtimes in the regime where quantum advantage is expected. This suggests that making VQE practical in the near-term requires the use of new approaches to measurement that leverage quantum coherence to reduce estimation runtimes, such as the recently proposed Bayesian amplitude estimation techniques [1, 2].

The rest of the paper is organized as follows: in Section II, we introduce the general structure of an RPA and proceed to describe the set of reactions we chose for our study. Our RPA method includes establishing accurate classical quantum chemistry reference values by comparison with experimental reaction energies. We then truncate the active space to establish the minimal number of qubits necessary to preserve chemical accuracy. The last step of our estimation evaluates the number of necessary measurements to reach chemical accuracy on a quantum computer, including measurement reduction techniques. This evaluation is made for increasing active space sizes to establish asymptotic scaling relationships. We present and discuss our numerical results for each step in Section III. We also introduce an empirical extrapolation formula to establish runtimes and resource requirements of more general systems than the ones specifically studied in this work. Finally, we present our conclusions and further research avenues in Section IV.

II. METHODS

In this section we describe our methodology for resource estimation, starting with an outline of the RPA concept as applied to VQE, followed by a detailed description of the methods employed for the estimation of classical and quantum computational resources.

A. Outline of the resource and performance assessment

The goal of an RPA is simple: we want to estimate as accurately as possible the resources, such as number of qubits, number of measurements, fidelity, among others, needed to achieve a given quality of solution for a specific choice of quantum algorithm and a set of problem instances. By fixing a target quality in the solution, it is possible to compare the cost with that of state-of-the-art classical approaches, establishing whether a quantum advantage is possible. The process can be divided in five stages:

1. Define a set of problem instances and the quantum algorithm to be assessed.
2. Set a target metric for performance. For example, choose a target quality of solution or time to solution.
3. Select a classical approach for comparison, ideally the state-of-the-art method for the instances of interest, and estimate the amount of classical resources required to achieve the target performance.
4. Estimate the amount of quantum computational resources required to achieve the target performance using the quantum algorithm.
5. Compare the performances and computational cost of the quantum and classical approaches. Establish whether a practical quantum advantage is attainable.

RPAs start by delimiting the problem we want to study to a set of relevant instances and choosing specific performance metrics to compare against state-of-the-art classical approaches. In this work, the problem under study is the calculation of the combustion energy for a set of small organic molecules. Our target performance metric is to compute these reaction energies to chemical accuracy. The next step is to select a classical algorithm for comparison. Here, we selected the gold standard for quantum chemistry, CCSD(T), and evaluated the cost of estimating combustion reactions to chemical accuracy. This resource estimate includes determining the basis set and number of spin-orbitals needed to achieve chemical accuracy. The performance of the classical approach provides the reference to be outperformed by the quantum algorithm, setting chemical accuracy as the target metric for quantum advantage. In principle, it is also possible to compare with available experimental data, provided reliable enthalpic corrections to the electronic energies are available. In our case, we instead ensured that the classical approach was reproducing experimental data, and then used the classical results as reference.

With the target accuracy fixed, we proceeded to estimate the number of qubits and total number of measurements needed to achieve such accuracy assuming access to a sufficiently expressive ansatz and high enough gate fidelity. By incorporating assumptions about the characteristics of the variational circuit and the quantum hardware, we established realistic runtime estimates for achieving chemical accuracy. These estimates are based on extrapolations of empirical formulas obtained from extensive numerical data generated for instances of different sizes. Crucially, our analysis takes into account the system and size dependence of different performance metrics of the algorithm. For this reason, we focus on techniques that can be scaled up to larger molecules or clusters, which is why considerations of spatial symmetry for example are not included.

To guarantee a realistic estimation of resources for a quantum algorithm, we need to consider the interplay of different aspects of the algorithm and characteristics of the problem instance. For example, in the case of VQE, the system size and the choice of active space determine the number of qubits for the calculation and thus influence the size of the variational circuit required and the number of measurements. The algorithmic choices also impact the final performance. For instance, the number of measurements in VQE is influenced by the grouping strategy used for the Hamiltonian and the way measurements are distributed among groups. We made these algorithmic choices to maximize the performance of the VQE approach.

While the most accurate RPA would require executing the algorithm, we can resort to educated approximations to estimate resources without executing quantum computations, taking advantage of our knowledge of the algorithm and the problem. The procedure might involve investigating the empirical scaling of the resources with system size in order to establish relationships that allow extrapolation to larger instances, which might not be accessible with existing quantum devices or classical simulators. Some performance metrics, such as the number of measurements in VQE, do not require simulation of the quantum circuits and only depend on properties of the problem instance. By performing estimates for problem instances of different sizes, we can establish whether there is a regime where quantum advantage is feasible, either in terms of quality of solution or time to solution. In the rest of this section we describe in details the benchmark data set chosen for our RPA as well as the methodology for estimating classical and quantum computational resources.

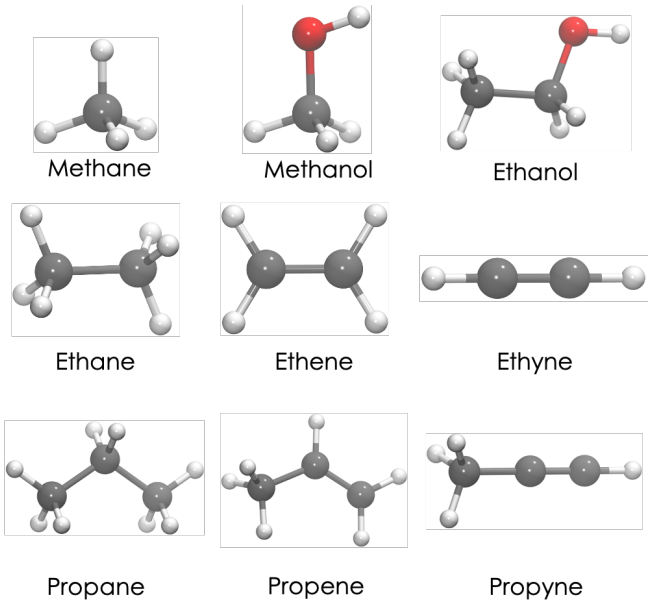

FIG. 1. Set of hydrocarbons for which we compute combustion enthalpies and corresponding quantum resources.

B. Benchmark data set

In this work, we aim to establish resource estimates that can be extrapolated to larger systems. To ensure our results are not specific to a single molecular system, we wish to apply our resource estimation procedure to a benchmark set of molecules. Ideally, this set would be of practical relevance, contain small enough molecules to allow chemically accurate computations, and correspond to well-established, accurate experimental data. For this reason, we chose to study combustion reactions for the small hydrocarbons depicted in Figure 1. For clarity, we explicitly write the general formula for the reaction's stoichiometry:

$$\mathrm{C}_x\mathrm{H}_y\mathrm{O}_z + \left(x + \frac{y}{4} - \frac{z}{2}\right)\mathrm{O}_2 \rightleftharpoons x\mathrm{CO}_2 + \frac{y}{2}\mathrm{H}_2\mathrm{O} \qquad (1)$$

Experimental enthalpies of combustion for the hydrocarbons in our benchmark set can easily be calculated from available enthalpies of formation,[36] reported in Table SI. By combining electronic ground state energies with vibrational, rotational, and translational enthalpic contributions, we can obtain simulated combustion enthalpies that can be compared to the experimental values. Most of our work focuses on getting accurate electronic energies, as harmonic vibrational corrections are obtained from the second derivatives of the electronic energies. Anharmonic effects are expected to be important for larger, flexible molecules but only play a very minor role in our benchmark reactions, as numerically verified in Section III A.

Algorithms to compute anharmonic vibrational spectra on quantum computers exist, and have been argued to be better candidates than electronic structure for early quantum advantage.[37] This assessment was based on considerations of scaling of the number of terms and their locality in the respective Hamiltonians. However, the relation between these quantities and the actual resources needed on a quantum computer is non-trivial, and notably depends on the coefficients in front of each Pauli term and of the measurement or qubit reduction techniques that can be applied. We hope that our method for resource estimation, detailed below, will be useful in future work to estimate resources for anharmonic vibrational spectra computations, which would allow to more accurately compare prospects of both problems for early quantum advantage.

Our chosen set of molecules is dominated by dynamical correlation. As pointed out by Elfving *et al.*, this means that a very large number of orbitals is needed for accurate treatment [21]. Hence, a very large number of qubits would be needed on a quantum computer to rival with quantum chemistry capabilities on classical computers. In that sense, systems dominated by non-dynamical correlations would be better candidates for demonstrations of near-term quantum advantage. However, we believe that most of our extrapolation and resource estimation results are valid for general molecular systems, whether dominated by dynamical or by non-dynamical correlations. In particular, our results regarding the scaling of the number of measurements necessary to reach chemical accuracy with the size of the system should be transferable to most cases.

C. Methodology for resource estimation

This section describes the methods used for the resource and performance assessment. All calculations were deployed using Zapata Computing's Orquestra® workflow management platform.

1. *Classical benchmarks*

The first component of the RPA consists of establishing a classical quantum chemistry approach to be used for comparison, evaluating its limitations and the classical resources needed to achieve chemical accuracy[5] for a benchmark set of molecules. The current gold standard for ground state electronic structure calculations is the Coupled-Cluster with Singles, Doubles and perturbative Triples CCSD(T) method [38], which is polynomially-scaling as N^7 where N is a proxy for

[5] Chemical accuracy is defined as 1 kcal/mol or approximately 1.6 mHa. Its use as a standard for the accuracy of chemistry modeling methods is motivated by the exponential sensitivity of properties such as equilibrium constants and reactions rates to changes in reaction energies [35].

molecular size. For closed-shell molecules, and with sufficiently large basis sets, CCSD(T) can reach chemical accuracy, i.e. an error of 1 kcal/mol compared to experimental data [35, 39]. Unfortunately, the N^7 scaling limits its application to small systems. Approximate CCSD(T) methods were developed[40–43] that take advantage of the spatial locality of electron correlation, which allowed computation of much larger systems[43, 44] by reducing the overall asymptotic scaling. However, the accuracy of the local approximations in coupled-cluster methods was recently questioned.[45, 46]

In addition, methods such as CCSD(T) are not suitable for some classes of molecular systems exhibiting non-dynamical correlation. In these systems, the mean-field solution to the electronic structure problem is qualitatively incorrect. For a more detailed discussion of dynamic versus non-dynamic correlation, we refer the reader to the existing literature.[47] Various methods have been developed and are still being perfected to treat systems with non-dynamic correlation with sub-exponential cost like DMRG,[48, 49] Quantum and Full CI Monte-Carlo [50–52], various selected CI schemes [53–55], variational reduced density matrix optimization [56, 57] etc. However, until now their application has been limited to relatively small active spaces, though much larger than what was previously possible with exponentially scaling methods. Large reactive metallocenters, of which FeMo-co is a prominent example, remain out of reach [21].

The geometries of all closed-shell molecules in our test set were optimized at the CCSD(T)[38]/cc-pVQZ[58] level with the density fitting approximation[59] in Psi4.[60] For O_2, the same level of theory was used with a ROHF reference and without the density fitting approximation for technical reasons. Harmonic vibrational frequencies were computed at the same level of theory as the geometry optimization for all molecules.

To assess the performance of some commonly used electronic structure methods, we computed electronic reaction energies with Hartree-Fock, MP2 [61], and the density functionals B3LYP [62, 63] and ωB97M-V [64] in the very large aug-cc-pV5Z [65] basis set. For all closed-shell molecules the RHF formalism was used whereas the UHF formalism was employed for O_2. The default Psi4 (version 1.2.1) options were used for other aspects of the computations. These calculations correlated all electrons.

Convergence of energies to within chemical accuracy of experimental data was achieved by computing high-level electronic structure corrections. This data also allowed us to investigate the importance of various high-order contributions, and provided reference numbers for further numerical experiments on the size of the active space. For the smallest molecules in our benchmark set, our final electronic energy E_{final} is computed as:

$$E_{\text{final}} = E^{\text{AV5Z}}_{\text{CCSD(T)}} + \Delta E^{\text{VTZ}}_{\text{CCSDT}} + \Delta E^{\text{VTZ}}_{\text{CCSDT(2)}_\text{Q}} + \Delta E^{\text{VDZ}}_{\text{CCSDTQ}} + \Delta E^{\text{ACVTZ}}_{\text{core}} \quad (2)$$

where the subscript identifies the contribution and the superscript the basis set used. In this paper, we abbreviate the cc-pVXZ, aug-cc-pVXZ and aug-cc-pCVXZ[66] basis sets as VXZ, AVXZ and ACVXZ respectively, where X = D, T, Q or 5. $E^{\text{AV5Z}}_{\text{CCSD(T)}}$ is the density-fitted CCSD(T)/AV5Z electronic energy with core orbitals frozen, except for O_2 where the density fitting approximation was not used. In addition, we also computed the CCSD(T) energy in the AVDZ, AVTZ and AVQZ basis sets to verify the convergence of the corresponding reaction energies. The AVQZ and AV5Z results were used to extrapolate the CCSD(T) correlation energy to the basis set limit using a cubic extrapolation formula.[39, 67, 68]

Higher-order effects are estimated through incremental contributions. $\Delta E^{\text{VTZ}}_{\text{CCSDT}}$ estimates the effect of a full treatment of triples amplitudes[69] relative to CCSD(T) perturbative treatment:

$$\Delta E^{\text{VTZ}}_{\text{CCSDT}} = E^{\text{VTZ}}_{\text{CCSDT}} - E^{\text{VTZ}}_{\text{CCSD(T)}} \quad (3)$$

The additional effect of quadruple excitations is taken into account perturbatively by the $\Delta E^{\text{VTZ}}_{\text{CCSDT(2)}_\text{Q}}$ correction:

$$\Delta E^{\text{VTZ}}_{\text{CCSDT(2)}_\text{Q}} = E^{\text{VTZ}}_{\text{CCSDT(2)}_\text{Q}} - E^{\text{VTZ}}_{\text{CCSDT}} \quad (4)$$

where all energies are computed without the density fitting approximation in NWChem.[70] In the case of propane, the CCSDT(2)$_\text{Q}$[71–73] computations proved to be too expensive and the correction is computed with the VDZ basis set instead of VTZ. Note that the corrections (3) and (4) could be computed simultaneously, but we separate them here to gain insight into their individual contributions.

The difference between the perturbative CCSDT(2)$_\text{Q}$ treatment and the full, iterative CCSDTQ[74, 75] treatment of the quadruple amplitudes was estimated for all but the largest systems in our benchmark (namely ethane, ethanol, propane and propene) as:

$$\Delta E^{\text{VDZ}}_{\text{CCSDTQ}} = E^{\text{VDZ}}_{\text{CCSDTQ}} - E^{\text{VDZ}}_{\text{CCSDT(2)}_\text{Q}} \quad (5)$$

with all energies computed in NWChem without the density fitting approximation.

Correlation effects for core electrons should be estimated by including core polarization functions in the basis set. Our incremental estimation for core correlation effects is computed as:

$$\Delta E^{\text{ACVTZ}}_{\text{core}} = E^{\text{ACVTZ}}_{\text{CCSD(T),all}} - E^{\text{ACVTZ}}_{\text{CCSD(T),fc}} \quad (6)$$

where the subscript "all" indicates that all electrons are correlated whereas "fc" indicates that electrons in core

orbitals were frozen. Computations were done with the density fitting approximation in Psi4.

To obtain reaction enthalpies ΔE_{tot}, it is necessary to add translational, rotational and vibrational enthalpic contributions at finite temperature to E_{final}. The total enthalpic correction ΔH_{final} was computed as:

$$\begin{aligned}\Delta H_{\text{final}} &= H^{\text{harm}}_{\text{CCSD(T)}} + \Delta H^{\text{VDPT2}}_{\text{B3LYP}} \\ &= H^{\text{trans}}_{\text{CCSD(T)}} + H^{\text{rot}}_{\text{CCSD(T)}} + H^{\text{vib, harm}}_{\text{CCSD(T)}} + \Delta H^{\text{VDPT2}}_{\text{B3LYP}}\end{aligned} \tag{7}$$

where $H^{\text{trans}}_{\text{CCSD(T)}}$, $H^{\text{rot}}_{\text{CCSD(T)}}$ and $H^{\text{vib, harm}}_{\text{CCSD(T)}}$ are harmonic translational, rotational and vibrational enthalpies computed at the density-fitted CCSD(T)/cc-pVQZ level and without density fitting for O_2. Their sum is the total harmonic enthalpy $H^{\text{harm}}_{\text{CCSD(T)}}$. $\Delta H^{\text{VDPT2}}_{\text{B3LYP}}$ corrects the enthalpy for anharmonic vibrational contributions and is computed as:

$$\Delta H^{\text{VDPT2}}_{\text{B3LYP}} = H^{\text{vib, VDPT2}}_{\text{B3LYP}} - H^{\text{vib, harm}}_{\text{B3LYP}} \tag{8}$$

where $H^{\text{vib, harm}}_{\text{B3LYP}}$ is the harmonic vibrational enthalpy and $H^{\text{vib, VDPT2}}_{\text{B3LYP}}$ is the anharmonic vibrational enthalpy from VDPT2[76] calculations in the Quartic Force Field (QFF)[77] approximation as obtained from anharmonic ZPVE and frequencies computed by GAMESS. All these contributions were computed at the B3LYP/def2-TZVPPD[78, 79] level on geometries optimized with the same method, using a Lebedev grid with 99 radial and 590 angular points. All enthalpic contributions are computed at 298.15 K.

2. *Number of qubits*

To estimate the minimal number of qubits that can be used while recovering CCSD(T)/AV5Z results with sufficient accuracy, we explore different truncations of the virtual active space. Although we are using CCSD(T)/AV5Z as a reference for practical purposes, we are assuming that higher-order correlations would be recovered with a similar accuracy. In all cases, we include a correction for the missing correlation energy at the MP2 level, and all computations use the density fitting approximation. Hence, the total CCSD(T) energy $E^{\text{tot}}_{\text{CCSD(T)}}$ is computed as:

$$E^{\text{tot}}_{\text{CCSD(T)}} = E^{\text{trunc}}_{\text{CCSD(T)}} + \Delta E_{\text{MP2}} \tag{9}$$

where $E^{\text{trunc}}_{\text{CCSD(T)}}$ is the CCSD(T) energy with virtual orbitals truncated and

$$\Delta E_{\text{MP2}} = E^{\text{full}}_{\text{MP2}} - E^{\text{trunc}}_{\text{MP2}} \tag{10}$$

where $E^{\text{full}}_{\text{MP2}}$ is the MP2 energy in the full space while $E^{\text{trunc}}_{\text{MP2}}$ is the MP2 energy with virtual orbitals truncated.

The simplest choice for orbital truncation is to eliminate canonical virtual orbitals with the highest energy. However, a better choice is known in the quantum chemistry literature as the Frozen Natural Orbitals (FNO)[26–28] method. In FNO, pseudo-natural virtual orbitals are obtained by diagonalizing the virtual-virtual block of the MP2 density matrix. The virtual space is then truncated by ranking the pseudo-natural virtual orbitals by their occupation number, which is usually interpreted as an indication of their importance in the correlation energy. We used two different methods to select how many virtual orbitals to keep in the active space. In the first, we simply fix the total number of available qubits, and select the number of virtuals that fit on these qubits together with the active occupied orbitals. In the second, we used the FNO threshold as implemented in Psi4 for the FNO-CCSD(T) [59, 80] method: virtual orbitals with occupations lower than half the threshold for both the α and β spins are excluded from the active space while all others are kept. We can then count the number of virtual orbitals in the active space to deduce the corresponding number of qubits.

We also note that there exists other methods to reduce the number of qubits necessary to encode a problem, most of which rely on using existing symmetries [81, 82]. We did not investigate these methods as only a small number of qubits can be eliminated without adding significant complexity to the Hamiltonian, a trade-off beyond the scope of the present study. We also chose not to restrict our analysis to include a specific spin symmetry. When extrapolating our results to systems that can benefit from symmetry, the number of necessary qubits should be divided by the appropriate factor.

3. *Measurement analysis*

a. General considerations To estimate expectation values of operators in the Variational Quantum Eigensolve (VQE) algorithm, it is necessary to perform many measurements and average their results. In the case of quantum chemistry simulations, the total number of necessary measurements M can be estimated as follows:

$$M = \frac{K}{\epsilon^2}, \tag{11}$$

where ϵ is the desired precision on the estimation and K is a proportionality constant that depends on the Hamiltonian, the state being measured and the measurement strategy employed for the estimation [22, 83], as described below. Note that in quantum chemistry, an accuracy of of 1 mHa with respect to the exact ground state in the infinite basis set limit is typically desired, which means that the uncertainty due to sampling error ϵ must be less than this amount.

The variance of the Hamiltonian estimator, which determines K, depends on details of the measurement process and the qubit encoding. Here, we consider the

widely used Jordan-Wigner encoding of the fermionic Hamiltonian to qubits. The qubit Hamiltonian takes the form:

$$\hat{H} = \sum_i h_i \hat{P}_i, \tag{12}$$

where $\hat{P}_i$ is a product of Pauli operators acting on one or more qubits and h_i is the associated coefficient. While the simplest approach would be to measure each $\hat{P}_i$ independently, grouping and other Hamiltonian decomposition techniques can reduce the number of measurements required [29–33]. Applying the Lagrangian approach of Rubin *et al.* [24] shows that when measurements are optimally allocated to groups, the proportionality constant between the number of required measurements and the inverse square precision is given by

$$K = \left(\sum_C \sqrt{\sum_{\alpha,\beta \in C} h_\alpha h_\beta \mathrm{Covar}\left(\hat{P}_\alpha, \hat{P}_\beta\right)} \right)^2 \tag{13}$$

where C represents a group and α and β label terms in a group. This measurement allocation scheme assumes that the covariances between all operators $\hat{P}_\alpha$ and $\hat{P}_\beta$ are already known, including the variances $\mathrm{Var}\left(\hat{P}_\alpha\right) = \mathrm{Covar}\left(\hat{P}_\alpha, \hat{P}_\alpha\right)$. Eq. 13 represents a lower bound on the value of K that could be achieved in practice because, in general, one does not know the values of these covariances and must estimate them.

Because K depends on the variances and covariances of the operators $\hat{P}_i$, K depends on the specific Ansatz used and the values of the Ansatz parameters. To avoid these complications, and allow estimation of K for up to 80 qubits, we employ some simplifying approximations. For the variances, we consider two approximations. In the first, we assume variances to be 1, the upper bound. In the second, we estimate variances from CISD density matrices computed with Psi4 [84]. The variances can be computed from the expectation value $\left\langle \hat{P} \right\rangle$ of each operator since $\hat{P}^2 = 1$:

$$\mathrm{Var}\left(\hat{P}\right) = \left\langle \hat{P}^2 \right\rangle - \left\langle \hat{P} \right\rangle^2 = 1 - \left\langle \hat{P} \right\rangle^2 \tag{14}$$

We assume all of the covariances between different terms to be zero. This does not correspond to a worst or best case scenario, but approximates the effect of a random distribution of covariances within bounds given by the Cauchy-Schwartz inequality:

$$\left| \sqrt{\mathrm{Var}\left(\hat{P}_\alpha\right) \mathrm{Var}\left(\hat{P}_\beta\right)} \right| \geq \mathrm{Covar}\left(\hat{P}_\alpha, \hat{P}_\beta\right) \tag{15}$$

In practice, we observed that this approximation resulted in K being overestimated by a factor of $\sim$2 relative to estimates using covariances obtained from circuit simulations of optimized Ansatzes. Note that when the upper bound is used for variances and covariances are set to zero, the estimated value of K is determined by the coefficients of the Hamiltonian and does not depend on the Ansatz or its parameters.

b. Hamiltonian decomposition methods In the conventional approach to VQE, the Hamiltonian is decomposed into Pauli terms. It is possible to measure two operators $\hat{P}_i$ and $\hat{P}_j$ at the same time if they commute. However, general commutation implies that multi-qubit measurements may be necessary to obtain information about both operators [29]. Additional circuit operations are necessary to obtain these multi-qubit measurements, hence we restrict ourselves to grouping methods relying only on single-qubit measurements. The simplest one groups operators which are Qubit-Wise Commuting (QWC) [29] instead of the more general commutation relation. Qubit-Wise Commutativity implies that for both $\hat{P}_i$ and $\hat{P}_j$, the Pauli operators acting on the same qubit individually commute. Information about the expectation values of a set of mutually QWC operators can be obtained with a single measurement of all qubits, each in the appropriate single-qubit Pauli basis. Finding the optimal grouping of QWC operators is equivalent to solving the Minimum Clique Cover graph problem and is NP-hard in the general case [29]. Here we use a heuristic greedy algorithm that goes through all operators and adds each one to the first group with which it is qubit-wise commuting [85, 86]. In addition, we sort the list of operators according to their coefficients h_i, so operators with the largest coefficients are grouped first.

We also consider the orbital frames approach to Hamiltonian decomposition [25, 34, 86]. The Hamiltonian terms that only contain Z operators are measured in the usual way, while an eigenvalue decomposition is used to obtain a low-rank factorization of the remaining two-body terms. The expectation values of this low-rank factorization and the remaining one-body terms can be obtained by applying a linear-depth basis rotation circuit after the ansatz.

We also assessed the performance of a method based on grouping mutually anticommuting Pauli terms [32, 33]. Our results show that anticommuting grouping is less performant than QWC grouping (see Figure S10), therefore these results are not included in the main text.

A precise assessment of the number of measurements required and its scaling with increasing system size for both QWC and basis rotation grouping is fundamental in predicting the runtime of energy estimation on NISQ devices. For this purpose, we computed K for all the molecules in our benchmark set with the exception of O_2 for technical reasons. For each molecule, K was computed for different active space sizes, where we always used an integer number of qubits per active electron to facilitate extrapolation. This was done for up to 80 qubits, and the exponents and prefactors obtained by a power fit through our data were used to extrapolate the number of measurements necessary for the 100 to 200 qubits region. In addition, we performed these estimations for both the upper bound approximation to the variances and variances computed from CISD. Finally, for each case we also computed the Hamiltonian coefficients based on canonical orbitals (as is usual in most

VQE publications) and based on FNOs (consistent with our active space size estimations), relying on the AVDZ basis set in all cases. This results in a total of six K estimations for each molecule and active space, giving us unprecedented insight into the relative performance of the different variants examined.

c. Variance reduction Grouping the Hamiltonian terms is not the only possibility to reduce the total number of measurements needed. The Hamiltonian can also be transformed so that its overall variance is reduced. Here, we explored the Reduced Density Matrix Constraints (RDMC) method proposed by Rubin *et al.*[24] In brief, this method adds operators to the Hamiltonian that sum to zero and optimizes their coefficients to reduce the total variance. We implemented this method directly in the qubit picture, which was suggested by Rubin *et al.* to have better performance than the original implementation in the fermionic picture. We present a comparison of both implementations in Figure S9 for the interested reader. We apply RDMC to a small set of molecules, with up to 20 qubits included in the active space, and examine the reduction obtained in K for the case of no grouping, for QWC grouping and for the basis rotation grouping. The Hamiltonians examined were computed with FNOs based on the AVTZ basis set and variances were estimated from CISD density matrices.

In conclusion, our resource estimation method includes a benchmark of classical methods, which can then be used as a reference to estimate the number of qubits needed to reach chemical accuracy in the general case. We establish empirical scaling relations for the number of measurements using state-of-the-art grouping and measurement reduction techniques, various approximations for the variances involved and two different molecular orbital bases. These scaling relations and their prefactors allow us to estimate the number of measurements needed to reach chemical accuracy when the qubit active space for the molecules in our benchmark reaches 100 to 200 qubits. These relations are also useful as a general guide for the scaling of QWC and basis rotation methods, and for the performance of RDMC.

III. RESULTS AND DISCUSSION

A. Benchmarking classical chemistry methods

The main purpose of this section is to establish whether classical quantum chemistry methods can reach chemical accuracy for combustion reaction of small, closed-shell hydrocarbons, and to quantify how much effort is necessary to reach chemical accuracy. The results of our assessment of classical resources will establish a reference for the next step of our resource evaluation which is concerned with the number of qubits required for chemical accuracy. We begin our evaluation by considering the importance of harmonic enthalpic effects and the performance of common, relatively inexpensive quantum chemistry methods, before moving on to the contributions of various high-level corrections.

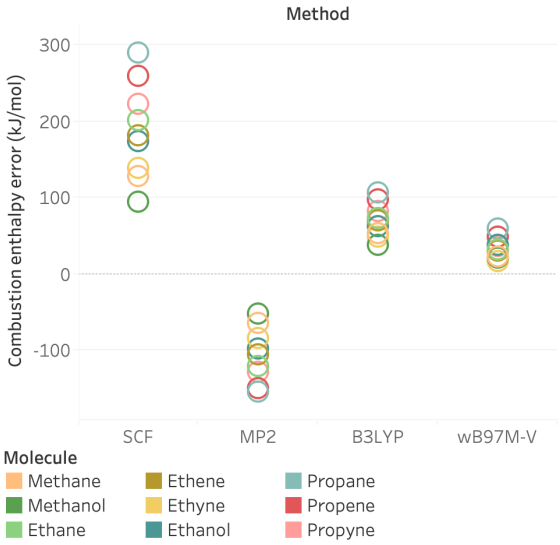

FIG. 2. Combustion enthalpy errors in kJ/mol using Hartree-Fock, B3LYP, ωB97M-V and MP2 in the AV5Z basis set. $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ contributions at the df-CCSD(T)/cc-pVQZ level are included.

The harmonic enthalpy contributions to the combustion energies $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ are reported in Table I. Contributions vary between -1.39 kJ/mol for CH_4O and 21.14 kJ/mol for C_2H_2. Since chemical accuracy is generally defined as a maximum error of 4.2 kJ/mol with respect to experimental data, the contribution of $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ cannot be neglected for our set of combustion energies.

In Figure 2, electronic combustion energies from Hartree-Fock, B3LYP, ωB97M-V and MP2 are added to $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ and the resulting error with respect to experiment is plotted. The very large AV5Z basis set used ensures all self-consistent field calculations are converged. In the case of MP2, the differences between AVQZ and AV5Z is about 10 kJ/mol (see Figure S1), much smaller than the plotted errors. As expected, the error for Hartree-Fock is the largest and ranges between 100 and 300 kJ/mol because of the neglect of dynamical correlation effects. MP2 improves over these numbers, but overshoots the correct value, a behavior which was previously reported [87]. MP2 errors approximately range between -50 and -150 kJ/mol, well outside the region of chemical accuracy. The popular B3LYP density functional has slightly lower errors of up to 100 kJ/mol. ωB97M-V is a recent functional that provides significant improvements, but still cannot reach chemical accuracy with errors ranging from 17 to 59 kJ/mol. In conclusion, the performance of common quantum chemistry methods that can routinely be applied to larger systems is insufficient to reach chemical accuracy.

To improve our results, we need to turn to the more

Molecule	CH_4	CH_4O	C_2H_6	C_2H_4	C_2H_2	C_2H_6O	C_3H_8	C_3H_6	C_3H_4
CCSD(T)/AV5Z	−809	−671.5	−1434.2	−1336.7	−1279.2	−1273.9	−2046.7	−1939	−1872.8
$\Delta H^{\text{harm}}_{\text{CCSD(T)}}$	8.44	−1.39	7.02	13.79	21.14	−0.43	7.65	14.42	20.64
CCSD(T)/AV5Z + $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$	−800.6	−672.9	−1427.2	−1322.9	−1258.1	−1274.3	−2039.0	−1924.6	−1852.2
Experiment	−802.5	−676.1	−1428.4	−1323.0	−1256.2	−1277.6	−2043.9	−1925.9	−1849

TABLE I. CCSD(T)/AV5Z combustion energies, harmonic enthalpic contributions $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ to combustion energies, and experimental values[36] in kJ/mol. All values at 298.15 K and 1 atm, see text for methods.

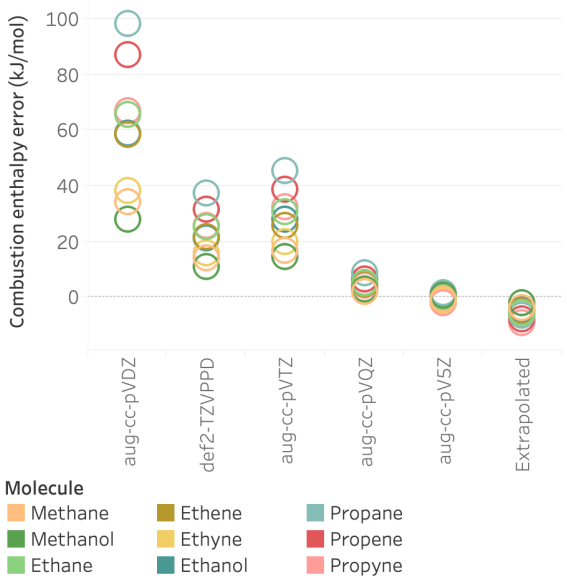

FIG. 3. CCSD(T) combustion enthalpy errors in kJ/mol in various basis sets and with AVQZ/AV5Z extrapolation. $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ contributions at the df-CCSD(T)/cc-pVQZ level are included.

computationally demanding gold standard of quantum chemistry: CCSD(T), which scales as N^7. In Figure 3, we see that convergence of the error for this method as a function of the basis set angular momentum is quite slow. At the AVTZ level, errors are almost as large as for ωB97M-V, whereas chemical accuracy is reached for all reactions at the AV5Z level. We also plotted results with the def2-TZVPPD basis set, which performs slightly better than AVTZ in spite of being slightly smaller. In general, we expect the def2-TZVPPD basis to be better suited for DFT calculations than the similar AVTZ basis, and this is indeed what we observe in Figure S1. Even at the AV5Z level, the agreement of the sum of CCSD(T) and $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ contributions with experimental values is somewhat fortuitous, and the energies are not completely converged yet. Indeed, a cubic extrapolation of the AVQZ and AV5Z correlation energies increases the error from experimental values, while providing results that should be closer to the complete basis set limit. The final errors vary between −2 and −10 kJ/mol.

Going beyond CCSD(T) is possible, albeit at significant cost, by computing high-level corrections in smaller basis sets. Indeed, the difference between CCSD(T) and high-level methods is expected to converge faster with the size of the basis set than the respective total energies.[88] We examine these corrections by increasingly high order treatment of amplitudes, and our results are gathered in Table II.

The ΔCCSDT correction represents the difference between the perturbative treatment of the triples amplitudes in CCSD(T) and the full treatment of these amplitudes which scales as N^8. Here, it amounts to at most 2 kJ/mol in the VTZ basis, confirming the accuracy of the CCSD(T) treatment. Next, we introduce perturbatively the effect of quadruples amplitudes through the $\Delta\text{CCSDT(2)}_\text{Q}$ correction that represents the difference between CCSDT and the CCSDT(2)_Q perturbative method, scaling as N^9. This correction is significant, and contributes up to 5.74 kJ/mol to combustion energies. Omitting the largest molecules, we confirm the accuracy of CCSDT(2)_Q by computing the full, N^{10} scaling, CCSDTQ energy at the VDZ level, resulting in a correction of at most 0.46 kJ/mol.

The overall effect of high-order excitations is the sum of ΔCCSDT, $\Delta\text{CCSDT(2)}_\text{Q}$ and ΔCCSDTQ and ranges between 2.5 and 7.7 kJ/mol. This effect is almost entirely canceled out by the $\Delta E^{\text{ACVTZ}}_{\text{core}}$ core correlation contributions as observed in Table II. The ACVTZ basis set used is likely to be sufficient for converging core correlation effects, as supported by comparison with results obtained in the ACVDZ basis in Figure S2. By contrast, the same contribution shows very large variations when computed with the AVTZ and AVQZ bases which lack core polarization functions, in spite of their higher angular momentum. This highlights the importance of using core polarization functions when computing core correlation effects.

Finally, anharmonic contributions to combustion enthalpies are negligible for the molecules considered and range between −0.77 kJ/mol and +0.18 kJ/mol. This is expected since vibrational anharmonicity plays a minor role in small rigid molecules, but would become significant in large and flexible systems with soft vibrational modes. Combining all corrections, we obtain an RMSE of 2.4 kJ/mol if the main contributions are computed with CCSD(T)/AV5Z, but of 5.1 kJ/mol if the AVQZ/AV5Z extrapolation of correlation energies is used (see Table SII). By comparison, experimental enthalpies of formation have uncertainties of about 1 kJ/mol or lower for our test set. The lower perfor-

Molecule	CH_4	CH_4O	C_2H_6	C_2H_4	C_2H_2	C_2H_6O	C_3H_8	C_3H_6	C_3H_4
ΔCCSDT (VTZ)	0.98	0.70	0.99	1.49	0.80	1.23	2.00	1.92	1.24
$\Delta CCSDT(2)_Q$ (VTZ)	2.63	1.70	4.39	3.86	3.55	3.49	5.32 (VDZ)	5.74	5.39
ΔCCSDTQ (VDZ)	0.21	0.10	N/A	0.31	0.37	N/A	N/A	N/A	0.46
$\Delta E_{\text{core}}^{ACVTZ}$	−3.15	−2.64	−5.41	−4.86	−3.81	−4.73	−7.54	−6.82	−5.60
Δanharmonic	+0.18	−0.34	−0.11	−0.16	−0.32	−0.77	−0.13	−0.25	−0.10
ΔE_{tot} error	2.75	2.72	1.06	0.74	−1.31	2.52	4.55	1.89	−1.81

TABLE II. High-level, core correlations, and anharmonic enthalpy corrections to combustion enthalpies in kJ/mol. All corrections computed with the indicated basis set, except for $\Delta CCSDT(2)_Q$ in the case of propane for which the VDZ basis was used. See Section II for details. ΔE_{tot} is the sum of all corrections combined with CCSD(T)/AV5Z + $\Delta H_{\text{CCSD(T)}}^{\text{harm}}$ results, and the error is computed relative to experimental data.

mance of the extrapolated energies is indicative that some contributions may not be fully converged, or that the AVQZ energies are not reliable enough for extrapolation. In any case, the CCSD(T)/AV5Z result is used in the following section.

In conclusion, the combination of harmonic enthalpies and CCSD(T)/AV5Z electronic energies provides combustion enthalpies within chemical accuracy. These accurate results stem from a cancellation of errors between high-level correlation effects and core correlation effects. This means that reliably reaching chemical accuracy for large systems is challenging even when only dynamical correlation is present. In cases where the observed error compensation does not occur, even CCSD(T) at the CBS limit might not be sufficient. Provided Ansätze on quantum computers can take into account high-order excitations at a sufficiently low polynomial cost, they could provide a better path to chemical accuracy. However, a significant number of qubits would be needed, as we demonstrate in the next section.

B. Number of qubits

In this section we explore how truncation of the active space impacts the combustion energy errors. We take as reference the CCSD(T)/AV5Z electronic combustion energies unless indicated otherwise. We then compute the CCSD(T) combustion energy in various basis sets with truncated virtual spaces, and compare the energy obtained to our CCSD(T)/AV5Z reference computed with all spin-orbitals. In all cases the ΔE_{MP2} correction is included to compensate part of the truncation error. We assume that the errors observed would be similar when truncating the virtual space for an ideal Ansatz that can effectively yield the FCI energy on a quantum computer. In our first experiment, we truncate the virtual space by keeping a fixed number of spin-orbitals: 40, 72 and 128.

As was previously reported in the literature [27], canonical virtual orbitals are not an optimal basis for virtual space truncation. Indeed, we observe very large errors in that case even with 128 qubits and the AV5Z basis sets, which are the largest active space and basis sets explored, respectively (see Figure S4). The smallest

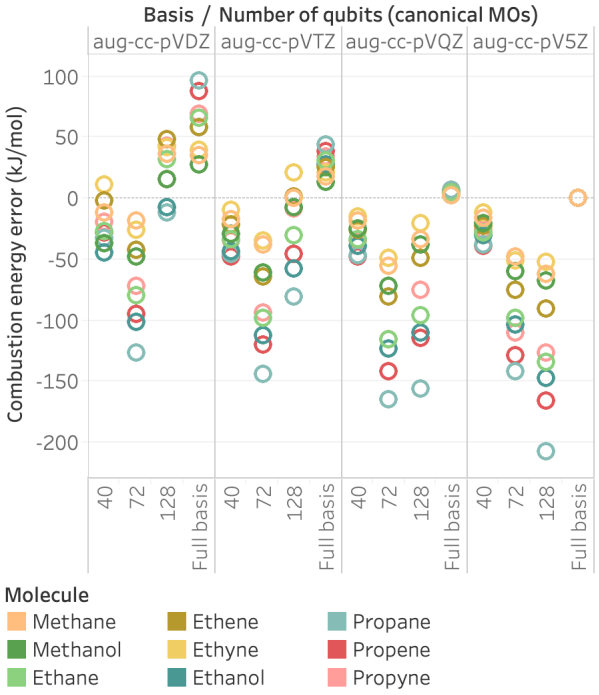

FIG. 4. Error relative to CCSD(T)/AV5Z for the FNO method with a fixed number of qubits and frozen core orbitals. ΔE_{MP2} correction is included in the results.

errors range between −50 and −200 kJ/mol. Moreover, the errors do not converge smoothly as the active space size is increased from 40 to 128 qubits.

A better truncation basis for correlated calculations is provided by Frozen Natural Orbitals (FNOs). In Figure 4, we plot the errors obtained for different basis sets and active space sizes. We first notice that combustion energy errors visibly converge towards the full basis set limit for each basis examined when going from 40 to 128 qubits. However, even in the largest active space chemical accuracy cannot be reached and the final errors range from 3 to 13 kJ/mol for AV5Z.

To facilitate the exploration of active space sizes, our second experiment switches to the FNO threshold as a criterion to select active virtual orbitals. Focusing on AVQZ and AV5Z, we present results for thresholds of

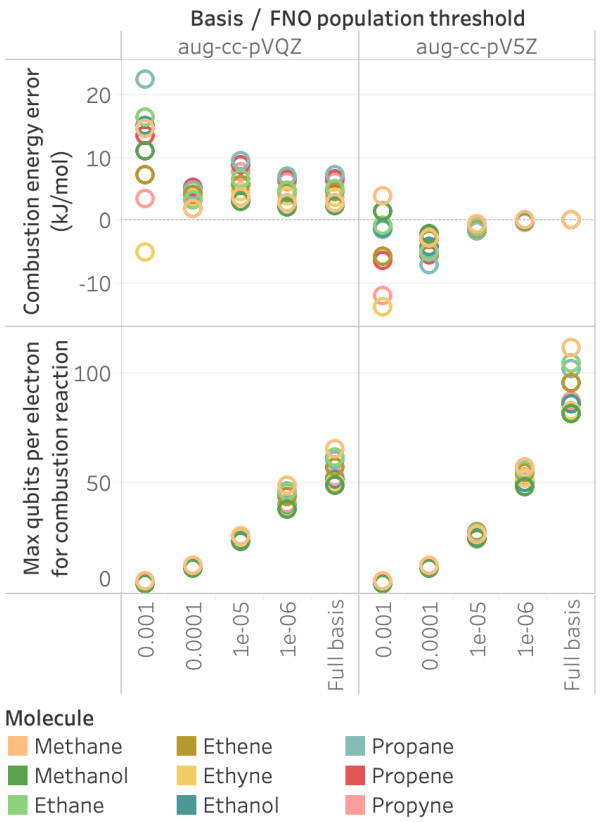

FIG. 5. (Top) Error relative to CCSD(T)/AV5Z for the FNO method using the FNO threshold for truncation and frozen core orbitals. (Bottom) Largest number of qubits per active electrons that would be needed to compute the combustion energy for each molecule in the given active space. ΔE_{MP2} correction is included in the results.

10^{-3}, 10^{-4}, 10^{-5} and 10^{-6}. The upper part of Figure 5 shows that the combustion energy error is indistinguishable from the full basis value at a threshold of 10^{-6}. A threshold of 10^{-5} yields a maximum deviation of -1.8 kJ/mol from the full basis result, whereas a threshold of 10^{-4} results in a maximum error of -7.1 kJ/mol, larger than chemical accuracy.

To connect the FNO threshold to the size of the active space in a transferable way, we plot the maximum number of qubits per active electron for each combustion reaction in the lower part of Figure 5. This number is obtained by dividing the number of active FNO spin-orbitals by the number of active electrons for each molecule in a combustion reaction, and then selecting the largest result. A threshold of 10^{-4} corresponds to about 13 qubits per electron, which is the number we will use to estimate the size of the active space necessary to reach chemical accuracy. This is an optimistic estimate: the errors we observe are slightly larger than chemical accuracy relative to the full basis limit for AV5Z, but this could be compensated for by including orbital optimization [89], or by using some of the qubit reduction techniques [81, 82] mentioned above.

To conclude, our estimation for the number of qubits N_{q} necessary to obtain accurate dynamical correlation energies is at least

$$N_{\mathrm{q}} \approx 13 N_{\mathrm{el}} \tag{16}$$

where N_{el} is the number of active electrons in the system.

C. Measurement estimation

In this section, our aim is to estimate the number of measurements needed for a single energy estimation step in the VQE procedure. We consider measurement reduction techniques based on qubit-wise commutativity of Pauli terms [29] and orbital basis rotation [25, 34], realistic variance estimation, and an efficient orbital basis so that our final estimates reflect conditions close to a large experiment. We aim to obtain extrapolation formulas for the number of measurements for each molecule in our benchmark set. This will allow us to extrapolate the number of necessary measurements for the large qubit active spaces needed for chemical accuracy (see Section III B). We also provide empirical scaling relations for two grouping methods.

1. *Performance of Hamiltonian decomposition methods*

We evaluated the Hamiltonian variance K for QWC grouping and basis rotation approach with two different bases for the Hamiltonian (canonical orbitals or FNOs) and two different estimates for the variances (upper bounds or CISD), giving a total of 4 different settings for each grouping method. We ran computations for all molecules in the set depicted on Figure 1, and also included H_2O and CO_2 that are necessary for computing combustion energies. Due to technical limitations in our code at the time of computation, the open-shell O_2 was omitted. We expect scaling results would be similar to those obtained for CO_2. For each molecule, we computed different active spaces with an integer number of qubits per active electron up to a total of 80 qubits. This represents the most extensive investigation of the number of measurements in VQE to our knowledge. We fit our results to a power law for each grouping method:

$$K = a(N_{\mathrm{q}})^b \tag{17}$$

where N_q is the number of qubits, a and b are fitted parameters. The obtained scaling exponents b are reported next to the corresponding curves on Figure 6.

The number of terms in the quantum chemistry Hamiltonian scales as N^4, where N is the number of qubits. However, the QWC grouping method with optimal measurement allocation approximately scales between N^5 and N^6. The optimal measurement allocation tends to attribute very little to no measurements

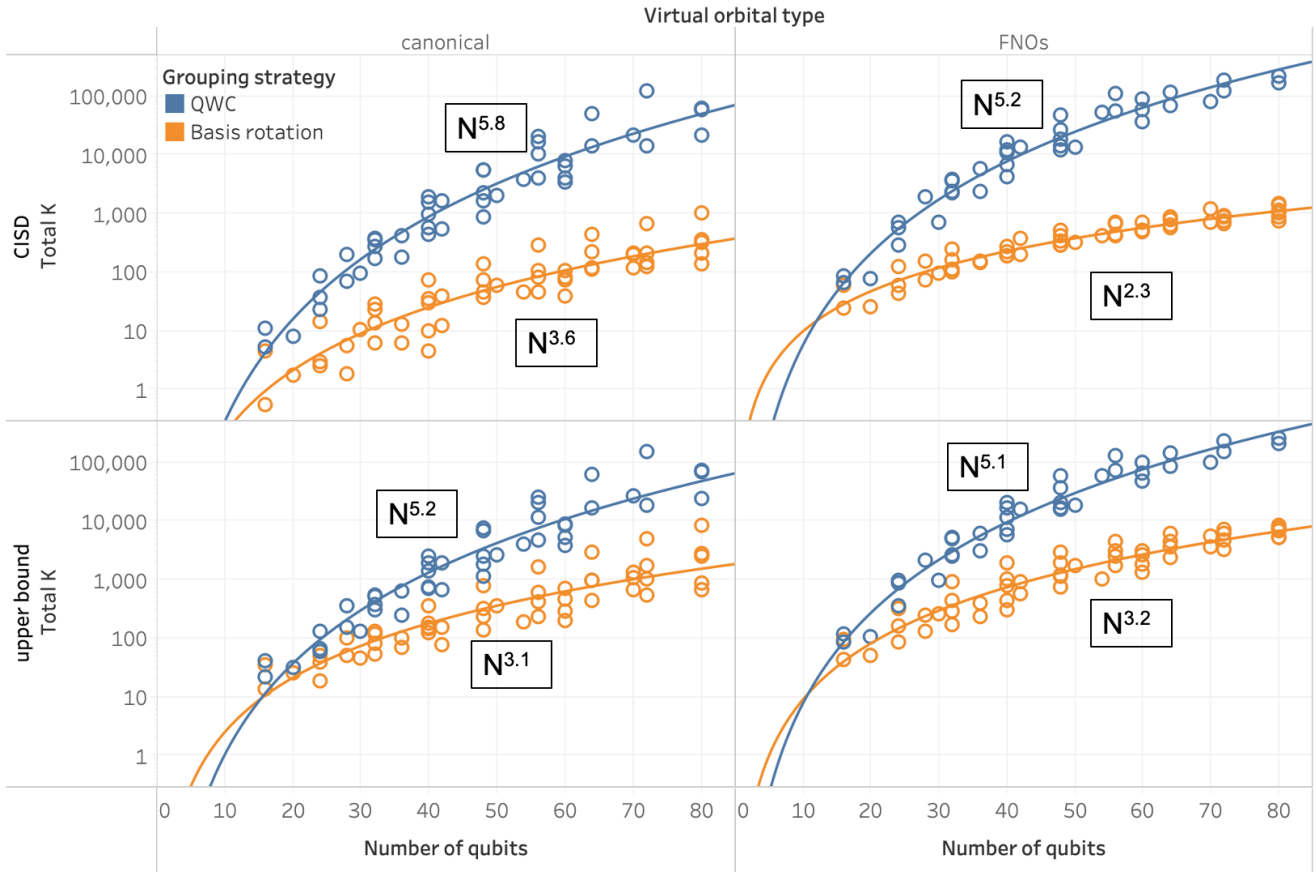

FIG. 6. Values of K computed for molecules in our benchmark set using QWC grouping (blue) and basis rotation grouping (orange). The top row approximates variances with CISD density matrices and the bottom row sets variances to their upper bounds. Covariances are set to zero in both cases. The left column represents the Hamiltonians in the canonical orbital basis and the right column in the FNO basis. A power law is fit through the data for each grouping method and the obtained exponent is reported next to the curve.

to terms with very small Hamiltonian coefficients that can safely be neglected. Thus, the observed scaling for QWC grouping only constitutes a modest improvement over the estimated upper bound of N^6 for scaling without grouping.[22]

Basis rotation grouping offers significantly better scaling, as hinted by the data presented by Huggins *et al.* for up to 32 qubits.[25] We observe that the scaling varies between $N^{2.3}$ and $N^{3.6}$, a very significant improvement compared to QWC grouping results. In addition, the effect of this improved scaling is already beneficial at low number of qubits, so that QWC grouping never appears advantageous in our computed data. The power law fits indicate that there is a crossover point at which QWC grouping could be preferred, but it only happens for very low number of qubits. Such an example appears in the next section, in Figure 7 for 12 qubits. Basis rotation grouping practically always yields a lower number of measurements, however it necessitates the addition of a basis rotation circuit before measurements are performed. Although this circuit has a very shallow depth,[90] in situations where fidelity needs to be maximized it might become too costly.

To facilitate the comparison of K computed with upper bound and CISD variances, we plotted again the data in Figure 6 so as to highlight the difference between the two variance estimation methods in Figure S5. As expected, this clearly shows that CISD variances always yield a lower number of measurements, albeit by only 20 to 30% when combined with QWC grouping. With basis rotation grouping, the benefit is significantly larger and reaches a factor 5 to 10. This shows that variance approximation is an important aspect to consider when estimating measurements: errors of an order of magnitude can occur when using upper bounds.

The effect of changing the orbital basis of the Hamiltonian from canonical orbitals to FNOs is visualized on Figure S6, which contains the same data as Figure 6 but highlighting the difference of interest in color. In section II C 2, we showed that FNOs yield significantly more correlation energy than canonical orbitals for the same number of qubits, which allows chemically accurate results in smaller active spaces. However, this increased accuracy comes at a price since the value of K is system-

atically higher for FNOs, by a factor of up to 10 in some cases. This is slightly compensated by a lower scaling exponent (compare left and right column of Figure 6), that reflects the fact that K saturates faster for FNOs. Indeed, when all virtual orbitals are included, the canonical and FNO spaces are the same and they must have the same K.

To obtain extrapolations of the value of K for each molecule, we must choose one of the eight variants investigated. For variance estimation, the CISD approximation is closest to what would be experimentally observed. In spite of the increased number of measurements, we believe the FNO basis is more advantageous since it yields more compact active spaces. Finally, we consider that the circuit fidelity is high enough to afford the orbital rotation circuit from basis rotation grouping. Hence, we fit Equation 17 for each molecule using K computed with basis rotation grouping, FNO Hamiltonians and CISD variances. We report our results in Table III, and we also plot the fits in Figure S7. The fit to all molecular data presented in Figure 6 yielded an exponent of 2.3, while for individual molecules b varies between 1.8 and 2.7. Most prefactors a have the same order of magnitude, except for H_2O where the prefactor is 5 to 10 times larger than for other molecules. However, H_2O also has the lowest exponent.

The fits presented above represent the scaling of K when increasing the number of qubits for a fixed molecule, which is convenient to extrapolate K to the very large active spaces needed for chemical accuracy. This "virtual scaling" is not the same as the "size scaling", where both the number of active electrons and the number of qubits increase. To investigate size scaling, we fitted Equation 17 through our data for increasing numbers of active electrons while keeping the number of qubits per active electrons fixed. We only have enough data to obtain meaningful fits for up to 5 qubits per electrons. Beyond that, the extrapolation gives incoherent results where larger active spaces would need fewer measurements, whereas for 2, 3, 4 and 5 qubits per electrons the obtained scaling is consistent. Overall, size scalings are slightly more favorable than virtual scalings (see Figure S8). The QWC grouping method scales around N^4 to $N^{5.5}$ in most cases, whereas the basis rotation method scales between N^2 and $N^{2.5}$. Thus, this data suggests that the basis rotation method provides a considerable asymptotic improvement in the number of measurements compared to QWC and related approaches.

2. *Variance reduction*

We now turn to a method that transforms the Hamiltonian to reduce the number of required measurements: the application of fermionic marginal constraints introduced by Rubin *et al.* [24], that we will abbreviate as RDMC for Reduced Density Matrix Constraints. Since

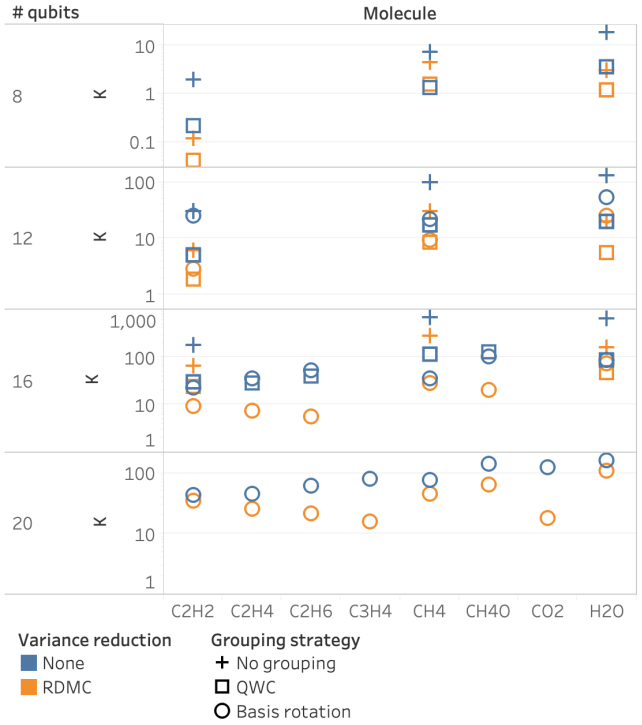

FIG. 7. Values of K computed for various molecules with no grouping (crosses), QWC grouping (squares) and basis rotation grouping (circles), both with (orange) and without (blue) RDMC. 8 qubit data is freezing 6 electrons for CH_4 and H_2O, others only freeze core electrons. Hamiltonians were represented with FNOs in the AVTZ basis set and variances estimated from CISD.

the application of RDMC takes significant classical resources using our prototype implementation, we restrict our study to a few molecules and active spaces. Our goal is to obtain an empirical estimation of the improvement in K that RDMC yields. Our results are presented in Figure 7, where we compare the performance of various grouping methods combined with or without RDMC. In all cases, RDMC yields reductions in the values of K. The reduction factor obtained is about 3 to 5 when no grouping methods is used. In the case of QWC grouping, the reduction provided by RDMC decreases a bit to a factor of 2 to 3. Basis rotation grouping usually yields the lowest K and has the best scaling with molecular size or number of qubits. Even in this case, RDMC is able to yield an additional improvement to K, of approximately a factor of 2. We note that the observed performance of RDMC seems to vary significantly among tested cases, and a factor of 2 is a somewhat conservative estimate. In general, the smaller reduction factors are obtained for larger number of qubits.

For low number of qubits, there are some irregularities in the patterns usually observed. For example, in active spaces of 12 qubits, the QWC grouping method generally performs better than the basis rotation grouping. This also happens for H_2O in 16 qubits with RDMC.

Molecule	H_2O	CO_2	CH_4	CH_4O	C_2H_6	C_2H_4	C_2H_2	C_2H_6O	C_3H_8	C_3H_6	C_3H_4
b	1.8	2.4	2.2	2.2	2.5	2.6	2.5	2.4	2.4	2.6	2.7
$a \cdot 10^2$	45	4.4	5.8	9.0	1.9	1.3	1.6	3.8	2.5	1.6	1.0

TABLE III. Results of fitting data for each molecule with $K = a(N_q)^b$ (Equation 17) when using basis rotation grouping, CISD variances and FNOs. a is multiplied by 100 in the table for clarity.

At these low number of qubits, very few virtual orbitals are included for each active electron, less than one for 12 qubits. This makes it difficult to extrapolate the behavior of the methods examined to large qubit numbers, and highlights the importance of running systematic benchmarks on large enough systems.

As highlighted in section II C 3, our RDMC implementation performs the Hamiltonian transformation in the qubit picture, as was suggested in the original work.[24] In Figure S9, we compare our results to the original implementation in the fermionic picture, and confirm that the qubit picture implementation systematically yields equivalent or better results. In general, RDMC shows a reduction in measurement count in all cases tested and therefore it could provide practical improvements for the implementation of VQE in the near-term. However, a more extensive analysis of the classical computational cost of this technique and the magnitude of its improvement when scaled to larger systems would improve the current assessment.

D. Overall qubit and runtime requirements

In this section, we summarize and gather the previous results to obtain estimates for the number of qubits, number of measurements and runtimes required to reach chemically accurate results for the set of investigated combustion reactions. The number of qubits N_q is estimated simply from Equation 16 and the number of valence electrons in each molecule. The number of measurements is computed as:

$$M = \frac{K}{2\epsilon^2} \tag{18}$$

where K is extrapolated for each molecule separately from Equation 17 with a and b taken from Table III. The extrapolation takes into account basis rotation grouping, approximated variances from CISD and assumes the Hamiltonian is expressed in the FNO basis. The extra factor of $1/2$ in Equation 18 approximately accounts for the additional measurement reduction provided by RDMC on top of the basis rotation grouping. We fix $\epsilon = 0.5$ mHa instead of the usual chemical accuracy of 1.6 mHa. Indeed, we allow 1.1 mHa for additional errors arising from truncation of the active virtual space and from device noise effects. Note that reducing the effect of device noise to below chemical accuracy in general is still a subject of research, and the low error we are assuming can only be achieved on the smallest circuits with the best devices currently.

To convert the number of measurements to actual runtimes, several additional assumptions are necessary. The first and perhaps most speculative regards the Ansatz. Although the UCCSD Ansatz generally yields good results in spite of deficiencies for strong correlation [91], the corresponding quantum circuit is extremely deep and not appropriate for NISQ devices. Alternatives have been designed [92, 93] however we will assume here that we can use a shallower, hardware-efficient Ansatz. Such Ansatz makes use of parametrized entangling gates that are taken to be hardware native or easily compiled to hardware native gates without significant overhead. We are assuming a linear connectivity of the qubit array, in which case a single layer of a hardware-efficient Ansatz is defined as the circuit of depth 2 that entangles every neighboring pair of qubits. We further assume that the number of layers needed to reach the ground state energy scales linearly with the number of qubits, and for the purposes of our estimation, we choose the prefactor in the scaling to be 2. It is likely that this depth constitutes a lower bound for the Ansatz depth that would be necessary in practice. Since our extrapolation for K assumes the basis rotation grouping, we also need to add the depth of the circuit for basis rotations, which is $N_\mathrm{q} - 3$ on a linear array of qubits if α and β spins can be transformed independently [90]. The final depth of the circuit would then be $5N_\mathrm{q} - 3$ in terms of two-qubit gates. Our final assumption is that runtime is dominated by execution times of two-qubit gates, which is assumed to be 100 ns, a value on the faster side of current superconducting gate times (see Table 1 in the review by Kjaergaard *et al.*[94]). The final formula we use to obtain runtimes t in seconds from the values of M and N_q reads:

$$t = 10^{-7} M(5N_\mathrm{q} - 3) \tag{19}$$

We report the results of our runtime estimates in Table IV. We also plot our estimated runtimes from the computed K values of Figure 6 and their extrapolation on Figure S11. The picture painted by these runtimes is very pessimistic for VQE. The shortest runtime for energy estimation, for CH_4, is 1.9 days. This is in spite of using rather optimistic estimates for the Ansatz depth, the number of qubits needed and neglecting the time for qubit reset, cloud latency times or measurement overheads for error mitigation. Moreover, we highlight again that this is the time necessary for a single energy evaluation. Running the full VQE algorithm involves

Molecule	H_2O	CO_2	CH_4	CH_4O	C_2H_6	C_2H_4	C_2H_2	C_2H_6O	C_3H_8	C_3H_6	C_3H_4
N_{el}	8	16	8	14	14	12	10	20	20	18	16
N_q	104	208	104	182	182	156	130	260	260	234	208
$K \cdot 10^{-3}$	1.9	16	1.6	8.4	8.5	6.6	3.1	24	16	23	18
$M \cdot 10^{-9}$	3.9	32	3.2	17	17	13	6.2	48	31	46	36
t (days)	2.3	39	1.9	18	18	12	4.6	71	47	62	44

TABLE IV. Estimated runtimes t in days for a single energy evaluation using the number of measurements M from extrapolated values of K (Equation 17 and Table III), with $\epsilon = 0.5$ mHa and the effect of RDM constraints included by a factor of 1/2 (see Equation 18). The number of qubits N_q is computed from the number of active electrons N_{el} and our empirical estimations of active space size (Equation 16).

optimizing the circuit parameters, which requires at least a few dozen to hundreds of iterations even with excellent optimizers. Hence, the total VQE runtime would be about a month for the smallest molecules in our test set. Larger molecules like ethanol already have a runtime of 71 days for a single energy evaluation.

These runtimes originate essentially in the considerable number of measurements necessary to obtain chemically accurate energies for molecules. Even on devices where the error rate would be small enough to warrant reliable VQE execution, the runtime to solution would be prohibitive for molecules in our test set. Parallelization of measurements over several quantum devices is a potential solution, provided all of these quantum devices are sufficiently similar, and the distribution of measurements designed to achieve chemical accuracy. However, parallelization could only bring a constant factor improvement and will not change the scaling of the runtimes with molecular size. In the case of systems dominated by non-dynamical correlation, a smaller active space might be sufficient to demonstrate quantum advantage over classical computing power. A recent paper[21] proposes the chromium dimer with a (24, 24) active space as a potential candidate. At 48 qubits, our extrapolation on Figure S11 indicates a runtime of a few hours, which could allow for a full VQE optimization with considerable effort. However, Hamiltonian coefficients for heavier, strongly correlated atoms like Cr might be larger, which would result in larger values of K. Moreover, even if such a computation becomes possible, the transition to practically relevant advantage could require active spaces beyond 100 qubits.[21]

Focusing on the scaling b and omitting the prefactor a, our results for the basis rotation grouping technique suggest that VQE has the potential to scale better with system size than methods such as Coupled Cluster. To transform this difference in scaling into an actual practical advantage, research should focus on two directions: 1) developing linear scaling Ansätze that provide sufficient accuracy on NISQ devices and 2) improving the measurement techniques, in particular to reduce the dependency of the number of measurements on the required precision. Regarding the first direction, having sufficiently accurate Ansätze for VQE with a circuit depth scaling only linearly implies an empirical runtime scaling of N^3 to N^4, which would be very competitive with the scaling of approaches such as CCSD(T). A number of Ansätze with linear scaling have been proposed [90, 95], but more studies should be devoted to investigating their representational power for chemical systems of interest, and the impact of noise on their accuracy. Along this line, the development and benchmarking of error mitigation techniques is crucial towards achieving sufficient accuracy on NISQ devices. Regarding the second direction, methods that can reduce the dependency of the number of measurements with respect to the required accuracy should be prioritized to make VQE competitive. One such method has been recently proposed by Wang *et al.*[1] and Koh *et al.*[2] which trade circuit fidelity for a reduction in the number of measurements.

IV. CONCLUSIONS

The Variational Quantum Eigensolver (VQE) is a heuristic algorithm, which does not have yet a demonstrated quantum speed-up over classical algorithms for quantum chemistry. Hence, it is of utmost importance to adequately benchmark VQE to evaluate its performance and prospects for quantum advantage. One significant step has recently been made in this direction [21] by identifying what molecules are the most likely candidates for quantum advantage, and in particular for practically relevant quantum advantage.

Here, we outlined a general procedure to assess quantum advantage with a quantum heuristic by carrying out a resource and performance assessment (RPA). We performed a specific RPA for computing a set of combustion energies with VQE, but our general method is also applicable to other variational algorithms. First, it is essential to assess the performance of state-of-the-art classical algorithms to check whether they can solve the problem at hand and estimate the compute resources required. Then, the number of qubits necessary to obtain a solution that is accurate enough should be established. Finally, a rigorous estimation of the number of measurements needed to evaluate expectation values with sufficient accuracy is performed. Measurement requirements are crucial to obtain approximate runtimes,

which are ultimately decisive for the practicality of the quantum algorithm.

Our classical benchmarks show that CCSD(T)/AV5Z complemented with harmonic enthalpic corrections at the CCSD(T)/VQZ level is sufficient to reproduce experimental combustion enthalpies to within chemical accuracy. Our high-level corrections reveal that the accuracy obtained originates from error cancellation between high-level correlation contributions (up to CCSDTQ) and core correlation effects. Anharmonic contributions to enthalpy are negligible for our systems. CCSD(T)/AV5Z is taken as our reference energy to estimate the minimal size of the active space that still yields chemical accuracy. Using the well-known FNO method for virtual space truncation, we observe that at least 13 qubits per active electrons must be included to obtain dynamical correlation energies within chemical accuracy.

Our RPA results further show that the number of measurements necessary for QWC grouping scales as N^5 to N^6, whereas the basis rotation grouping only needs about N^2 measurements, at the cost of a small addition to the overall circuit depth. The application of RDMC on the Hamiltonian in addition to grouping warrants another reduction in the number of measurements by a factor of 2. Unfortunately, the ϵ^{-2} precision dependence of measurement requirements introduces a very large multiplicative factor. With optimistic assumptions regarding the total circuit depth and the execution time of quantum circuits, estimating a single energy for molecules in our test set to chemical accuracy would take between a few days and a couple of months. Combined with the necessity for a large number of energy evaluations to optimize VQE parameters, this indicates that VQE with operator averaging is not currently practical even for molecules with only a few heavy atoms.

There are several possible ways to resolve this issue. One is to work on better Hamiltonian decomposition methods, and hopefully achieve reduction in the prefactor or the scaling of the number of measurements needed as a function of the system size. Another would be to work on improving Hamiltonian transformations to reduce the Hamiltonian variance further. Some of these directions have been explored by the authors without significant success. However, a more concrete improvement tackles the ϵ^{-2} dependence of the number of measurements. Recently, the use of Bayesian techniques combined with Engineered Likelihood Functions [1, 2] offered a way to exploit better device fidelity to reduce the number of measurements, bridging VQE and Quantum Phase Estimation in a practical way. Engineered Likelihood Functions may then be combined with grouping and variance reduction techniques to further reduce measurement requirements and runtime. Any proposed solution to the measurement bottleneck for the application of VQE should be benchmarked on various molecules and active spaces to assess its robustness and scaling with system size.

ACKNOWLEDGEMENT

The authors would like to acknowledge support from BP. The authors also acknowledge insightful scientific discussions and suggestions from Alex Kunitsa, Peter Johnson, Christopher Brown, and Peter Love, and support from the team of scientists and engineers at Zapata Computing.

[1] G. Wang, D. E. Koh, P. D. Johnson, and Y. Cao, Bayesian inference with engineered likelihood functions for robust amplitude estimation, arXiv:2006.09350v2.

[2] D. E. Koh, G. Wang, P. D. Johnson, and Y. Cao, A framework for engineering quantum likelihood functions for expectation estimation, arXiv:2006.09349v1.

[3] D. Castelvecchi, Ibm's quantum cloud computer goes commercial, Nature News **543**, 159 (2017).

[4] M. Mohseni, P. Read, H. Neven, S. Boixo, V. Denchev, R. Babbush, A. Fowler, V. Smelyanskiy, and J. Martinis, Commercialize quantum technologies in five years, Nature **543**, 171 (2017).

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).

[6] Feasibility and timeframes of quantum computing, in *Quantum computing: progress and prospects*, edited by E. Grumbling and M. Horowitz (National Academies Press, Washington, D.C., 2019) Chap. 7, pp. 156–189.

[7] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[8] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, Validating quantum computers using randomized model circuits, Phys. Rev. A **100** (2019).

[9] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005).

[10] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, *et al.*, Quantum chemistry in the age of quantum computing, Chem. Rev. **119**, 10856 (2019).

[11] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[12] P. Deglmann, A. Schäfer, and C. Lennartz, Application of quantum calculations in the chemical industry-An overview, International Journal of Quantum Chemistry **115**, 107 (2015).

[13] A. Heifetz, *Quantum Mechanics in Drug Discovery* (Springer Science+Business Media, LLC, 2020).

[14] A. Van der Ven, Z. Deng, S. Banerjee, and S. P. Ong, Rechargeable alkali-ion battery materials: Theory and computation, Chemical Reviews **120**, 6977 (2020).

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

[16] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, Proceedings of the National Academy of Sciences **114**, 7555 (2017).

[17] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, arXiv:2007.14460v1.

[18] D. Wecker, B. Bauer, B. K. Clark, M. B. Hastings, and M. Troyer, Gate-count estimates for performing quantum chemistry on small quantum computers, Phys. Rev. A **90**, 022305 (2014).

[19] I. D. Kivlichan, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, H. Neven, and R. Babbush, Improved Fault-Tolerant Quantum Simulation of Condensed-Phase Correlated Electrons via Trotterization, Quantum **4**, 296 (2020).

[20] J. Lee, D. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction (2020), arXiv:2011.03494 [quant-ph].

[21] V. E. Elfving, B. W. Broer, M. Webber, J. Gavartin, M. D. Halls, K. P. Lorton, and A. Bochevarov, How will quantum computers provide an industrially relevant computational advantage in quantum chemistry?, arXiv:2009.12472v1.

[22] J. R. McClean, R. Babbush, P. J. Love, and A. Aspuru-Guzik, Exploiting locality in quantum computation for quantum chemistry, J. Phys. Chem. Lett. **5**, 4368 (2014).

[23] M. Kühn, S. Zanker, P. Deglmann, M. Marthaler, and H. Weiß, Accuracy and resource estimations for quantum chemistry on a near-term quantum computer, J. Chem. Theory Comput. **15**, 4764 (2019).

[24] N. C. Rubin, R. Babbush, and J. McClean, Application of fermionic marginal constraints to hybrid quantum algorithms, New J. Phys. **20**, 053020 (2018), arXiv:1801.03524 [quant-ph].

[25] W. J. Huggins, J. McClean, N. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers (2019), arXiv:1907.13117.

[26] T. L. Barr and E. R. Davidson, Nature of the configuration-interaction method in ab initio calculations i ne ground state, Phys. Rev. A **1**, 644 (1970).

[27] C. Sosa, J. Geertsen, G. W. Trucks, R. J. Bartlett, and J. A. Franz, Selection of the reduced virtual space for correlated calculations. an application to the energy and dipole moment of {H2O}, Chem. Phys. Lett. **159**, 148 (1989).

[28] A. G. Taube and R. J. Bartlett, Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory, Collect. Czech. Chem. Commun. **70**, 837 (2005).

[29] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, arXiv:1907.03358v4.

[30] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi, and F. T. Chong, Minimizing State Preparations in Variational Quantum Eigensolver by Partitioning into Commuting Families (2019), arXiv:1907.13623.

[31] P. Gokhale and F. T. Chong, $O(N^3)$ Measurement Cost for Variational Quantum Eigensolver on Molecular Hamiltonians (2019), arXiv:1908.11857.

[32] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi, Unitary partitioning approach to the measurement problem in the Variational Quantum Eigensolver method (2019), arXiv:1907.09040.

[33] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake, and P. J. Love, Measurement reduction in variational quantum algorithms, Phys. Rev. A **101**, 062322 (2020).

[34] M. D. Radin and P. Johnson, Measurement reduction via orbital frames decompositions on quantum computers, *World patent WO 2020/146794*, 2020.

[35] K. A. Peterson, D. Feller, and D. A. Dixon, Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges, Theoretical Chemistry Accounts **131**, 1079 (2012).

[36] W. M. Haynes, ed., *CRC Handbook of Chemistry and Physics [Online], 96th ed* (CRC Press) http://hbcponline.com/ (accessed 2016).

[37] N. P. D. Sawaya, F. Paesani, and D. P. Tabor, Near- and long-term quantum algorithmic approaches for vibrational spectroscopy, arXiv:2009.05066v1.

[38] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, A fifth-order perturbation comparison of electron correlation theories, Chem. Phys. Lett. **157**, 479 (1989).

[39] J. Rezac and P. Hobza, Benchmark calculations of interaction energies in noncovalent complexes and their applications, Chem. Rev. **116**, 5038 (2016).

[40] C. Riplinger and F. Neese, An efficient and near linear scaling pair natural orbital based local coupled cluster method, J. Chem. Phys. **138**, 034106 (2013).

[41] J. J. Eriksen, P. Baudin, P. Ettenhuber, K. Kristensen, T. Kjaergaard, and P. Jorgensen, Linear-scaling coupled cluster with perturbative triple excitations: The divide-expand-consolidate ccsd(t) model, J. Chem. Theory Comput. **11**, 2984 (2015).

[42] J. Yang, G. K.-L. Chan, F. R. Manby, M. Schutz, and H.-J. Werner, The orbital-specific-virtual local coupled cluster singles and doubles method, J. Chem. Phys. **136**, 144105 (2012).

[43] P. R. Nagy and M. Kállay, Approaching the basis set limit of ccsd(t) energies for large molecules with local natural orbital coupled-cluster methods, J. Chem. Theory Comput. **15**, 5275 (2019).

[44] O. Caldararu, M. A. Olsson, C. Riplinger, F. Neese, and U. Ryde, Binding free energies in the sampl5 octa-acid host–guest challenge calculated with dft-d3 and ccsd(t), Journal of Computer-Aided Molecular Design **31**, 87 (2017).

[45] Y. S. Al-Hamdani, P. R. Nagy, D. Barton, M. Kállay, J. G. Brandenburg, and A. Tkatchenko, Interactions between large molecules: Puzzle for reference quantum-mechanical methods, arXiv:2009.08927v1.

[46] M. Bot, V. Gorbachev, A. Tsybizova, and P. Chen, Bond dissociation energies in the gas phase for large molecular ions by threshold collision-induced dissociation experiments: Stretching the limits, The Journal of Physical Chemistry A **124**, 8692 (2020).

[47] D. K. W. Mok, R. Neumann, and N. C. Handy, Dynamical and nondynamical correlation, The Journal of Physical Chemistry **100**, 6225 (1996).

[48] G. K.-L. Chan and M. Head-Gordon, Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group, J. Chem. Phys. **116**, 4462 (2002).

[49] S. R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. **69**, 2863 (1992).

[50] N. S. Blunt, S. D. Smart, J. A. F. Kersten, J. S. Spencer, G. H. Booth, and A. Alavi, Semi-stochastic full configuration interaction quantum monte carlo: Developments and application, J. Chem. Phys. **142**, 184107 (2015).

[51] A. Lüchow, Quantum monte carlo methods, WIREs Computational Molecular Science **1**, 388 (2011).

[52] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum monte carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[53] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, Heat-bath configuration interaction: An efficient selected configuration interaction algorithm inspired by heat-bath sampling, J. Chem. Theory Comput. **12**, 3674 (2016).

[54] N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, A deterministic alternative to the full configuration interaction quantum monte carlo method, J. Chem. Phys. **145**, 044112 (2016).

[55] P.-F. Loos, Y. Damour, and A. Scemama, The performance of cipsi on the ground state electronic energy of benzene, The Journal of Chemical Physics **153**, 176101 (2020).

[56] G. Gidofalvi and D. A. Mazziotti, Spin and symmetry adaptation of the variational two-electron reduced-density-matrix method, Phys. Rev. A **72**, 052505 (2005).

[57] J. Fosso-Tande, T.-S. Nguyen, G. Gidofalvi, and A. E. DePrince, Large-scale variational two-electron reduced-density-matrix-driven complete active space self-consistent field methods, Journal of Chemical Theory and Computation **12**, 2260 (2016).

[58] T. H. Dunning, Jr., Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen, J. Chem. Phys. **90**, 1007 (1989).

[59] A. E. DePrince and C. D. Sherrill, Accuracy and efficiency of coupled-cluster theory using density fitting/cholesky decomposition, frozen natural orbitals, and a t1-transformed hamiltonian, J. Chem. Theory Comput. **9**, 2687 (2013).

[60] R. M. Parrish, L. A. Burns, A. E. D. I. D. G. A. Smith, A. C. Simmonett, E. G. Hohenstein, U. Bozkaya, R. D. R. A. Yu. Sokolov, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. S. III, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability, J. Chem. Theory Comput. **13**, 3185–3197 (2017).

[61] C. Moller and M. S. Plesset, Note on an approximation treatment for many-electron systems, Phys. Rev. **46**, 618 (1934).

[62] A. D. Becke, Density-functional thermochemistry. iii. the role of exact exchange, J. Chem. Phys. **98**, 5648 (1993).

[63] C. Lee, W. Yang, and R. G. Parr, Development of the colle-salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B **37**, 785 (1988).

[64] N. Mardirossian and M. Head-Gordon, wb97m-v: A combinatorially optimized, range-separated hybrid, meta-gga density functional with vv10 nonlocal correlation, J. Chem. Phys. **144**, 214110 (2016).

[65] R. A. Kendall, T. H. Dunning, and R. J. Harrison, Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions, J. Chem. Phys. **96**, 6796 (1992).

[66] D. E. Woon and T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. v. core-valence basis sets for boron through neon, The Journal of Chemical Physics **103**, 4572 (1995).

[67] L. A. Burns, M. S. Marshall, and C. D. Sherrill, Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (scs), explicitly correlated (f12), and specialized wavefunction approaches, J. Chem. Phys. **141**, 234111 (2014).

[68] P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, Benchmark database of accurate (mp2 and ccsd(t) complete basis set limit) interaction energies of small model complexes, dna base pairs, and amino acid pairs., Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[69] J. Noga and R. J. Bartlett, The full ccsdt model for molecular electronic structure, J. Chem. Phys. **86**, 7041 (1987).

[70] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. van Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, Nwchem: a comprehensive and scalable open-source solution for large scale molecular simulations, Comput. Phys. Commun. **181**, 1477 (2010).

[71] S. Hirata, M. Nooijen, I. Grabowski, and R. J. Bartlett, Perturbative corrections to coupled-cluster and equation-of-motion coupled-cluster energies: A determinantal analysis, J. Chem. Phys. **114**, 3919 (2001).

[72] S. Hirata, M. Nooijen, I. Grabowski, and R. J. Bartlett, Erratum: "perturbative corrections to coupled-cluster and equation-of-motion coupled-cluster energies: A determinantal analysis" [j. chem. phys. 114, 3919 (2001)], The Journal of Chemical Physics **115**, 3967 (2001).

[73] S. Hirata, P.-D. Fan, A. A. Auer, M. Nooijen, and P. Piecuch, Combined coupled-cluster and many-body perturbation theories, J. Chem. Phys. **121**, 12197 (2004).

[74] N. Oliphant and L. Adamowicz, Coupled-cluster method truncated at quadruples, J. Chem. Phys. **95**, 6645 (1991).

[75] S. A. Kucharski and R. J. Bartlett, The coupled-cluster single, double, triple, and quadruple excitation method, J. Chem. Phys. **97**, 4282 (1992).

[76] N. Matsunaga, G. M. Chaban, and R. B. Gerber, Degenerate perturbation theory corrections for the vibrational self-consistent field approximation: Method and applications, J. Chem. Phys. **117**, 3541 (2002).

[77] K. Yagi, K. Hirao, T. Taketsugu, M. W. Schmidt, and M. S. Gordon, Ab initio vibrational state calculations with a quartic force field: Applications to h2co, c2h4, ch3oh, ch3cch, and c6h6, The Journal of Chemical Physics **121**, 1383 (2004).

[78] F. Weigend and R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[79] D. Rappoport and F. Furche, Property-optimized gaussian basis sets for molecular response calculations, J. Chem. Phys. **133**, 134105 (2010).

[80] A. E. DePrince and C. D. Sherrill, Accurate noncovalent interaction energies using truncated basis sets based on frozen natural orbitals, J. Chem. Theory Comput. **9**, 293 (2012).

[81] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic hamiltonians

(2017), arXiv:1701.08213 [quant-ph].
[82] M. Steudtner and S. Wehner, Fermion-to-qubit mappings with varying resource requirements for quantum simulation, New J. Phys. **20**, 063010 (2018).
[83] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).
[84] C. David Sherrill and H. F. Schaefer, The configuration interaction method: Advances in highly correlated approaches (Academic Press, 1999) pp. 143 – 269.
[85] R. S. Smith, M. J. Curtis, and W. J. Zeng, A practical quantum instruction set architecture (2016), arXiv:1608.03355 [quant-ph].
[86] J. R. McClean, K. J. Sung, I. D. Kivlichan, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlicek, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. Rubin, N. P. D. Sawaya, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, Openfermion: The electronic structure package for quantum computers, arXiv:1710.07629v5.
[87] S. Grimme, n-alkane isodesmic reaction energy errors in density functional theory are due to electron correlation effects, Org. Lett. **12**, 4670 (2010).
[88] A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, Heat: High accuracy extrapolated ab initio thermochemistry, J. Chem. Phys. **121**, 11599 (2004).
[89] W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan, and Y.-y. Ohnishi, Orbital optimized unitary coupled cluster theory for quantum computer, Phys. Rev. Research **2**, 033421 (2020).
[90] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, Phys. Rev. Lett. **120**, 110501 (2018).
[91] B. Cooper and P. J. Knowles, Benchmark studies of variational, unitary and extended coupled cluster methods, J. Chem. Phys. **133**, 234102 (2010).
[92] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, Journal of Chemical Theory and Computation **15**, 311 (2019).
[93] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nature Communications **10**, 3007 (2019).
[94] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, arXiv:1905.13641v3.
[95] P.-L. Dallaire-Demers and N. Killoran, Quantum generative adversarial networks, Phys. Rev. A **98**, 012324 (2018).

V. SUPPLEMENTARY INFORMATION

A. Additional classical benchmark data

In this section, we present additional data relevant to our classical benchmarks. We present the experimental enthalpies of formation that were used to compute the reference combustion enthalpies in Table SI. In Figure S1, we plot the basis set error with respect to AV5Z as a function of the number of orbitals for different methods. This shows that the def2-TZVPPD basis set (in orange) is slightly more compact than the AVTZ basis set (in teal), and yields slightly larger errors for SCF and MP2. However, for the DFT methods B3LYP and ωB97M-V, def2-TZVPPD is both more compact and significantly closer to AV5Z results than AVTZ.

In Figure S2, we plot the effect of core correlation on combustion energies computed with different basis sets. The ACVDZ and ACVTZ results only differ by 1 or 2 kJ/mol. In spite of using higher angular momentum functions, the results in the AVTZ and AVQZ basis sets that lack core polarization functions vary as much as 6 kJ/mol, and are very different from the ACVTZ results. This highlights the well-known importance of core polarization functions when computing core correlation effects.

In Figure S3, we plot the purely anharmonic contributions to the enthalpy and free energy of combustion at different temperatures. For the molecules in our test set, this contribution can be neglected at 298.15 K since it amounts to at most -0.8 kJ/mol. The temperature dependence of the enthalpy correction is pretty low and is still negligible at 800 K. Because of the temperature-dependent entropic contribution, the free energy correction reaches a value of -4 kJ/mol at 800 K, at which point it cannot be neglected for chemical accuracy. We also point out that purely anharmonic corrections are likely to become more important in larger molecules with softer vibrational modes.

B. Additional data for active space selection

In this section we present additional data regarding our active space selection procedure, which ultimately determines the number of qubits necessary for chemical accuracy. Figure S4 shows that canonical orbitals provide a poor basis for truncation of the active space. Often, the errors do not converge monotonically towards the full basis set limit when increasing the size of the active space from 40 to 128 qubits. Moreover, the 128 qubits results is still significantly different from the full basis result, especially in the larger basis sets.

C. Additional data for measurement scaling

In this section we present some additional plots regarding our measurement scaling results. Figure S5 and S6 contain the same data as Figure 6, but plotted to highlight differences between variance estimation methods and choice of orbital basis, respectively. CISD variances are expected to be closer to experimental variances than upper bounds, and generally yield lower values of K, especially in the case of basis rotation grouping. FNOs

Molecule	H_2O	CO_2	CH_4	CH_4O	C_2H_6	C_2H_4	C_2H_2	C_2H_6O	C_3H_8	C_3H_6	C_3H_4
ΔH_f (kJ/mol)	−241.8	−393.5	−74.6	−201	−84	+52.4	+227.4	−234.8	−103.8	+20	+184.9

TABLE SI. Experimental enthalpies of formation[36] in the gas phase at 298.15 K and 1 atm for computing combustion enthalpies of molecules in our benchmark set.

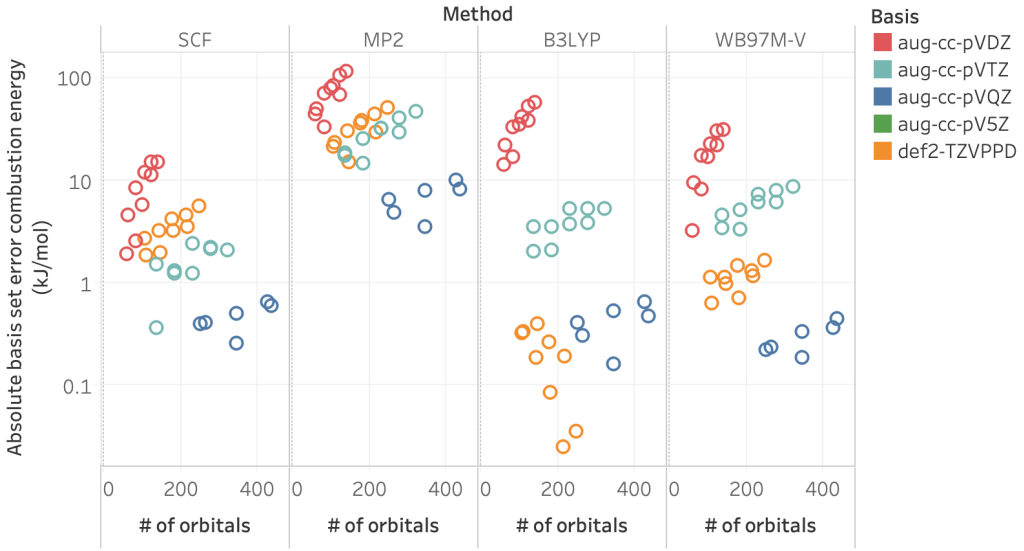

FIG. S1. Log scale error in combustion energies as a function of the number of orbitals for different methods and basis sets. For each method, the error is computed relative to results in the AV5Z basis set.

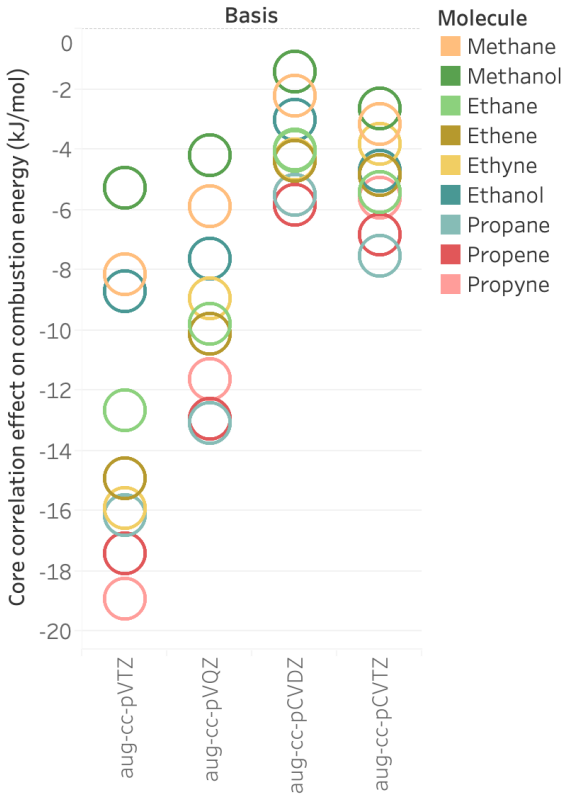

FIG. S2. Core correlation effects on combustion energies computed with different basis sets.

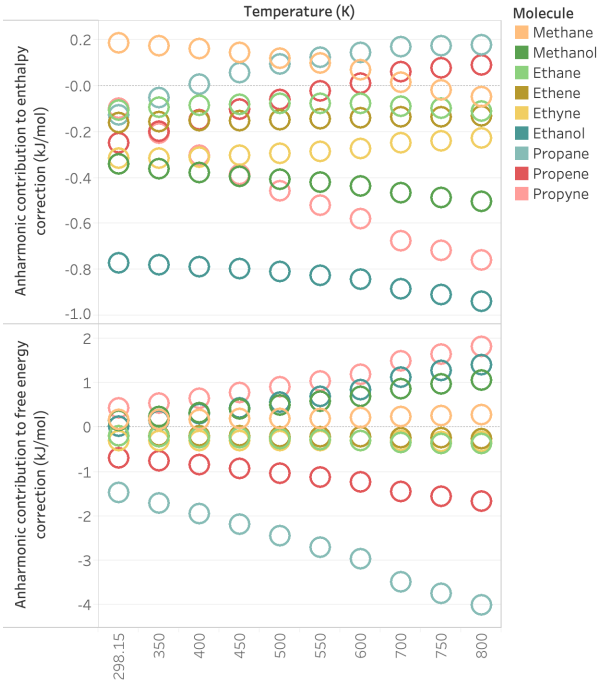

FIG. S3. Anharmonic contributions to enthalpy corrections (top) and to free energy correction (bottom) to the combustion energies. These are the purely anharmonic contributions that would be added on top of the harmonic contributions.

Molecule	CH_4	CH_4O	C_2H_6	C_2H_4	C_2H_2	C_2H_6O	C_3H_8	C_3H_6	C_3H_4
CCSD(T)/Q5 + $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$	-806.1	−678	−1434.1	−1328.1	−1262.4	−1281	−2048.2	−1932.9	−1858.1
$\Delta E_{\text{tot, extrap}}$ error	−2.75	−2.38	−5.84	−4.46	−5.61	−4.18	−4.65	−6.41	−7.71

TABLE SII. CCSD(T)/Q5 extrapolated combustion energies combined with harmonic enthalpic contributions $\Delta H^{\text{harm}}_{\text{CCSD(T)}}$ at 298.15 K and 1 atm. $\Delta E_{tot,extrap}$ includes the sum of high-order, core correlation and anharmonic contributions, similarly to ΔE_{tot} in the main text, but using the Q5 extrapolated CCSD(T) energies instead of CCSD(T)/AV5Z.

always yield larger values of K than canonical orbitals, however they are more efficient at recovering correlation energy and allow us to use smaller active spaces and thus less qubits.

In Figure S7, we plot the curves obtained by fitting the power law in Equation 17 to the K values obtained for each individual molecule using basis rotation grouping, CISD variances and FNOs. All curves are similar, which is reflected in the similar values of the coefficients a and b reported in Table III.

In Figure S8, we plot the values of K obtained for a fixed number of qubits per electrons, while increasing the number of active electrons. This represents the scaling of K for different variants of grouping methods, variance estimation and orbital bases as a function of the overall size of the system, where both the number of electrons and the number of qubits increase. Beyond 5 qubits per electrons, our data is insufficient for reliable extrapolation and is not represented here. Overall, the scaling coefficients are slightly more favorable than in the case where only the number of qubits is increased and the number of electrons is kept fixed.

Figure S9 show the difference between the implementation of the RDM constraints method in the fermionic or in the qubit picture. The original work[24] employed the fermionic picture implementation, but suggested that the qubit picture might yield better results. Indeed, we observe that our qubit implementation (in orange) performs at least as well or better than our fermionic implementation (in blue). During our experiments, we also observed that the final performance depends on the chosen optimizer.

Figure S10 compares values of K computed with no grouping (red), QWC grouping (blue) and anticommuting grouping (orange). For all molecules, the core orbitals were frozen and the Hamiltonian expressed in the canonical orbital basis contained 20, 30 or 40 qubits. The upper bound for the variances was used and covariances were set to zero. The anticommuting groups were determined through the same greedy algorithm than QWC groups, but checking for anticommutativity instead of commutativity. In all cases, anticommuting grouping offers an improvement over no grouping, while being significantly outperformed by QWC group-

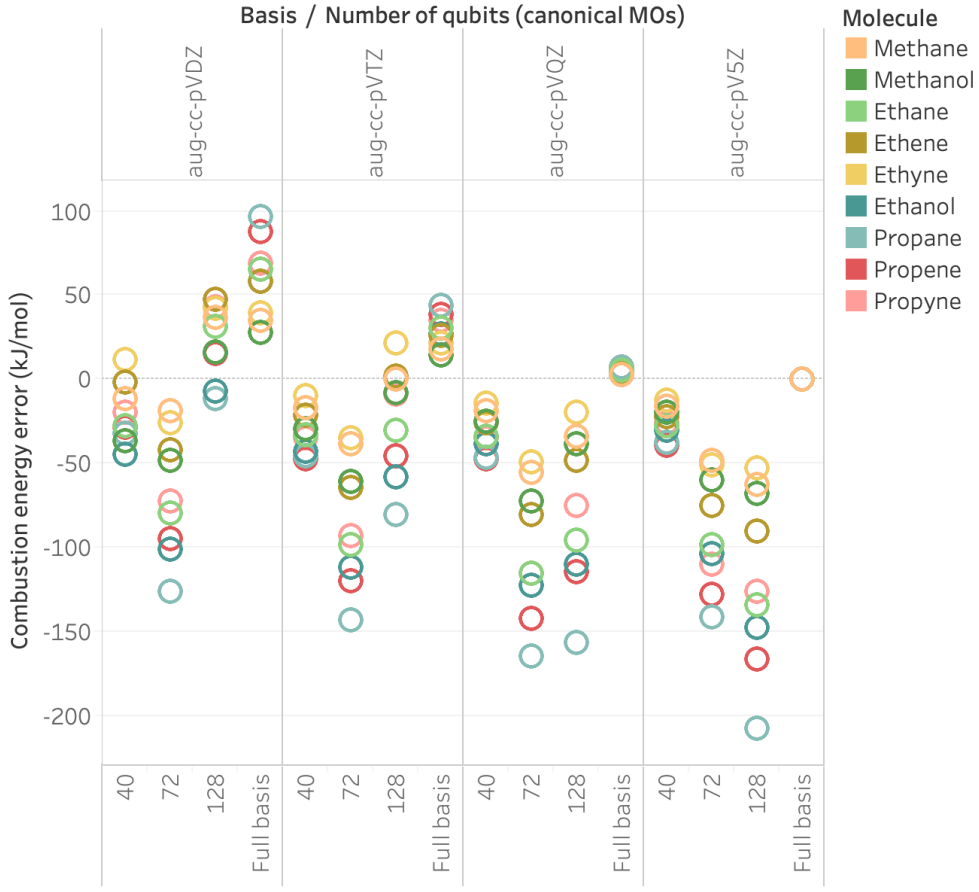

FIG. S4. For each basis set, we plot the error in combustion energy relative to the CCSD(T)/AV5Z calculation including all orbitals. In each case, we plot the error when keeping 40, 72, 128 or all qubits in the active space, using a canonical orbital basis.

ing.

D. Additional data for runtimes

In Figure S11, we plot our runtime extrapolations for each molecule, using the same estimation method than for Table IV but varying the number of qubits. We observe that at least a few days of measurement times are needed at the lower end of the gray region which indicates the range of number of qubits needed for chemical accuracy in the case of combustion reactions. Assuming a similar scaling, this plot also allows us to estimate a runtime of at least a few hours for the 48 qubit active space corresponding to 24 electrons in 24 orbitals for the chromium dimer.

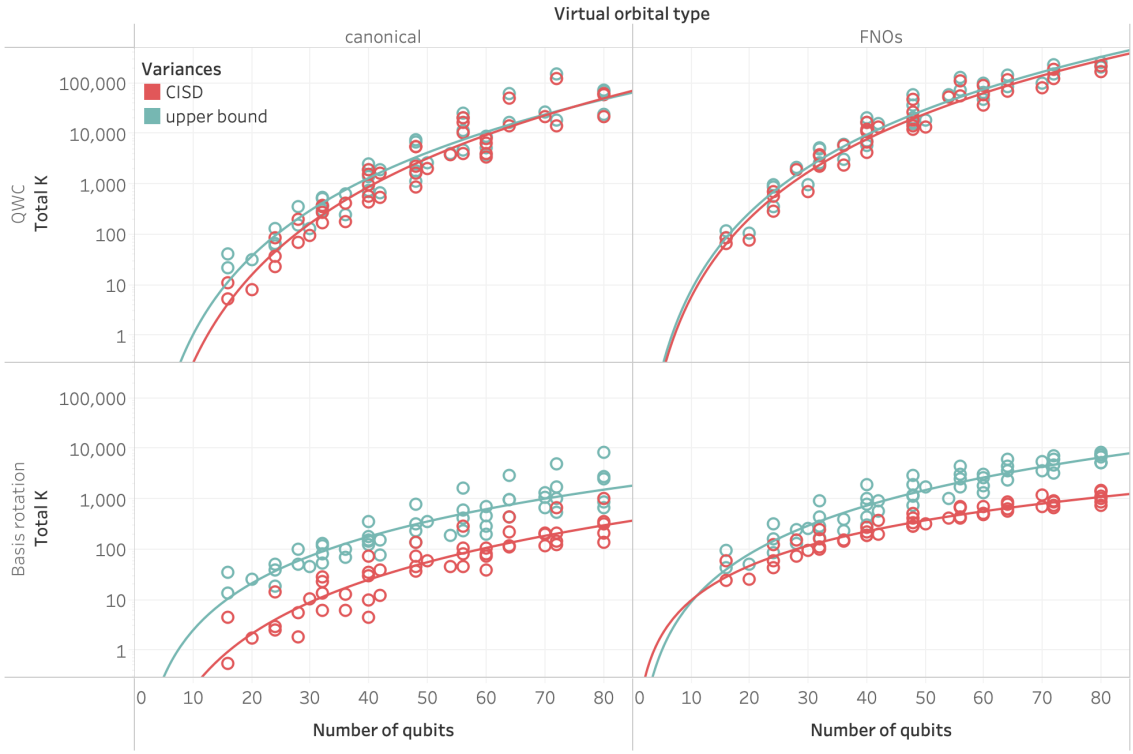

FIG. S5. Values of K computed for molecules in our benchmark set approximating variances using upper bounds (teal) and CISD data (red). Covariances are set to zero in both cases. The top row uses QWC grouping and the bottom row basis set grouping, while the left column represents the Hamiltonians in the canonical orbital basis and the right column in the FNO basis. A power law is fit through the data for each variance approximation method to help visualization.

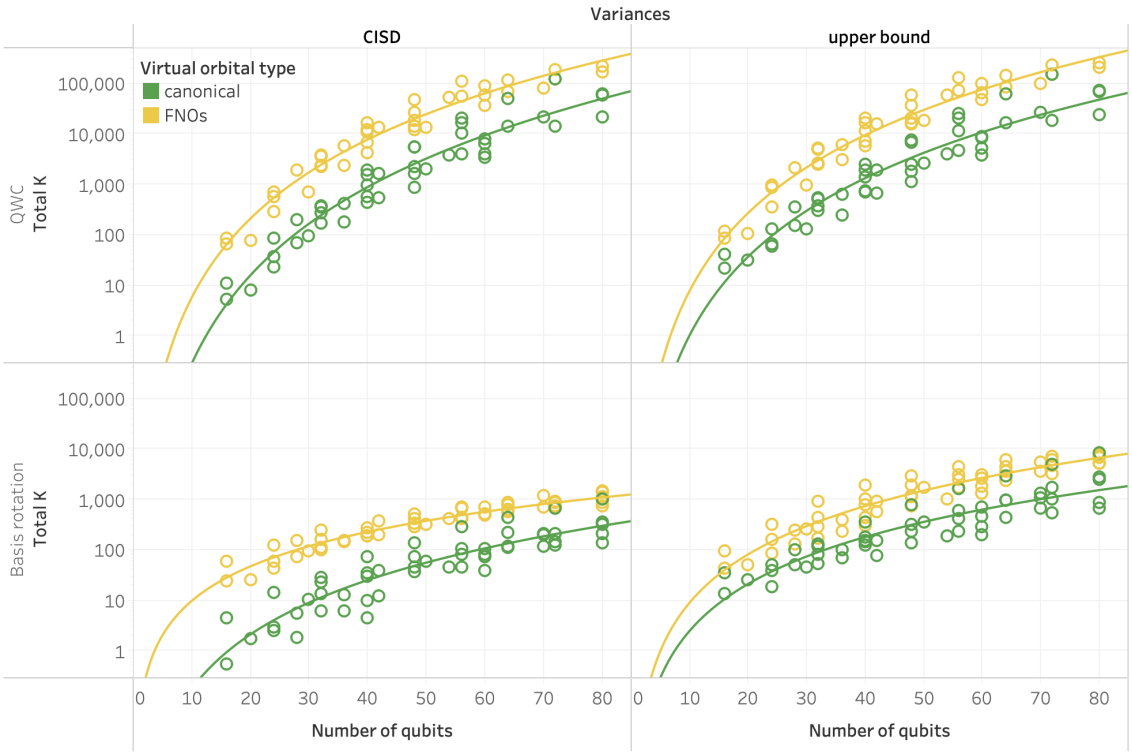

FIG. S6. Values of K computed for molecules in our benchmark set using canonical orbitals (green) and FNOs (yellow) to represent the Hamiltonian. The top row uses QWC grouping and the bottom row basis set grouping. Variances are estimated using CISD in the left column and using their upper bounds in the right column. Covariances are set to zero in both cases. A power law is fit through the data for each orbital basis to help visualization.

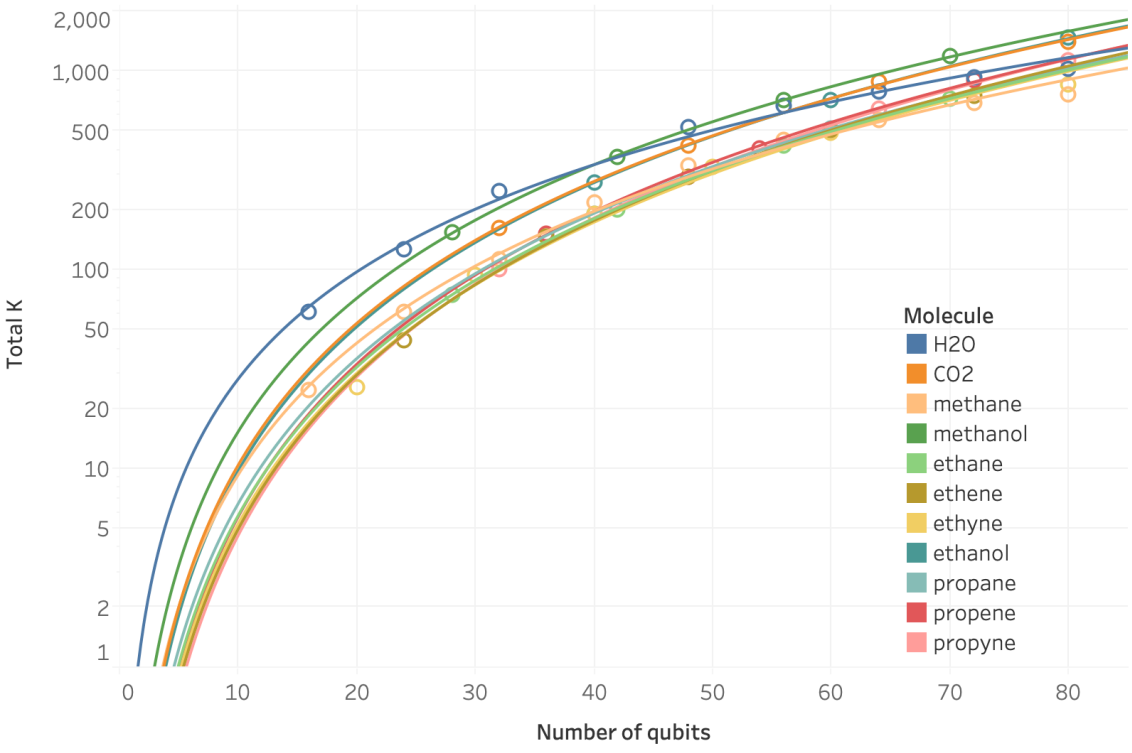

FIG. S7. Values of K computed for various molecules using basis rotation grouping with CISD variances and FNOs. For each molecule, a power law $K = aN^b$ is fitted through the data. Values of a and b obtained are reported in Table III.

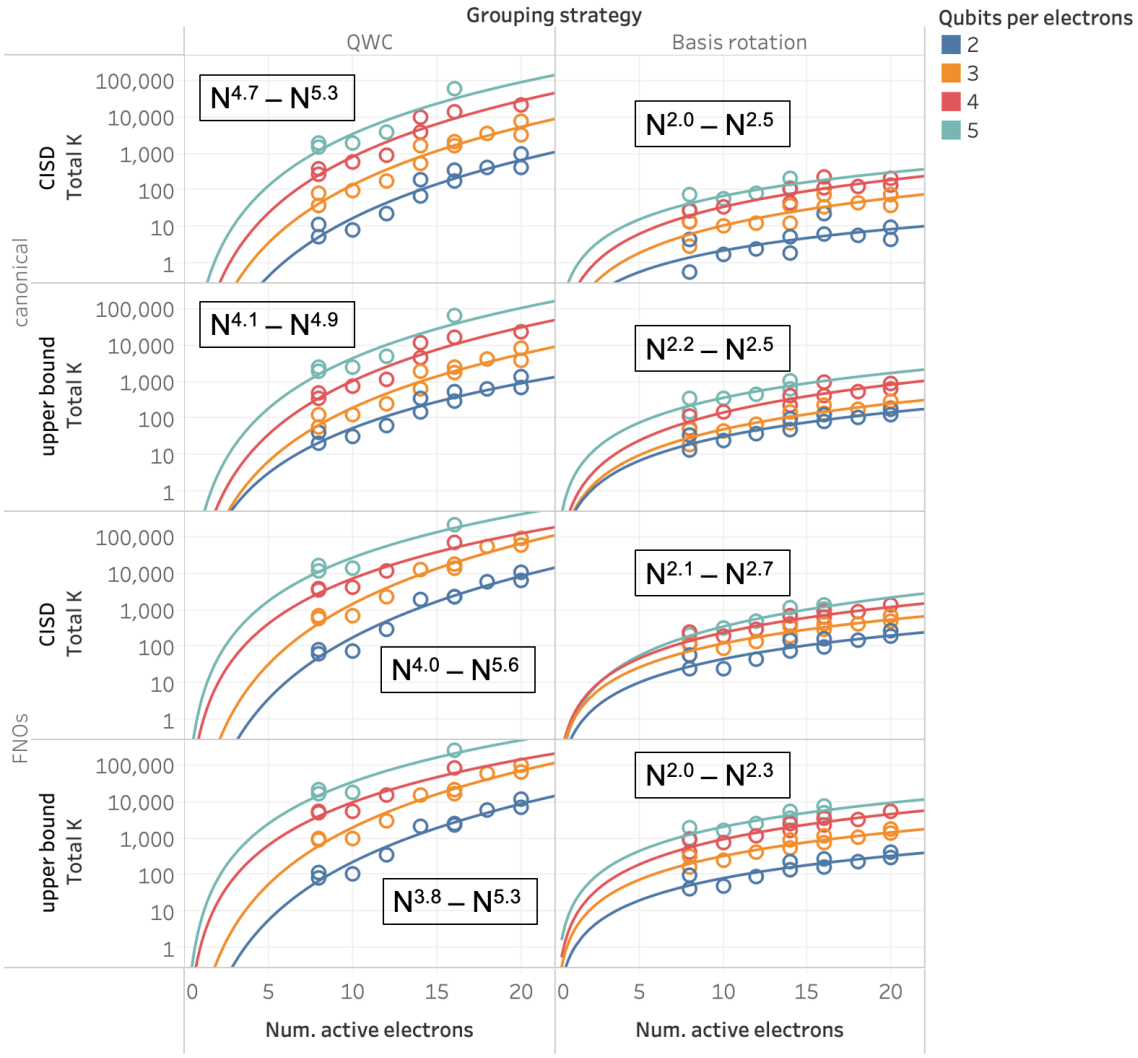

FIG. S8. Values of K computed for molecules of increasing size, while fixing the number of qubits per electrons to 2 (blue), 3 (orange), 4 (red) and 5 (teal). Left column uses QWC grouping while right column contains results with basis rotation grouping. The top two rows represent the Hamiltonian in the canonical basis and the bottom two in the FNO basis. Finally, variances are estimated by upper bounds or CISD as indicated. A power law is fit through the data for each number of qubits per electrons. The range of exponents obtained for the asymptotic scaling is reported on each plot.

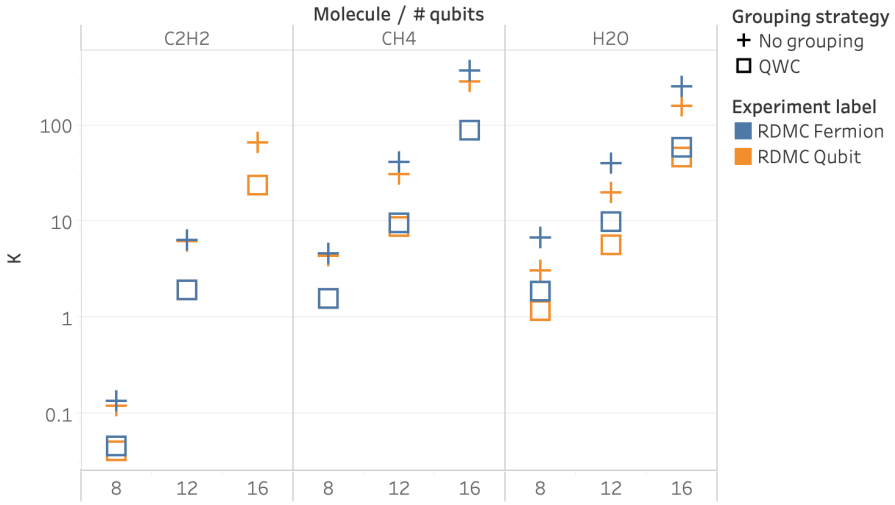

FIG. S9. Values of K computed for various molecules using no grouping and QWC grouping, combined with RDM constraints in the fermionic picture (blue) or in the qubit picture (orange).

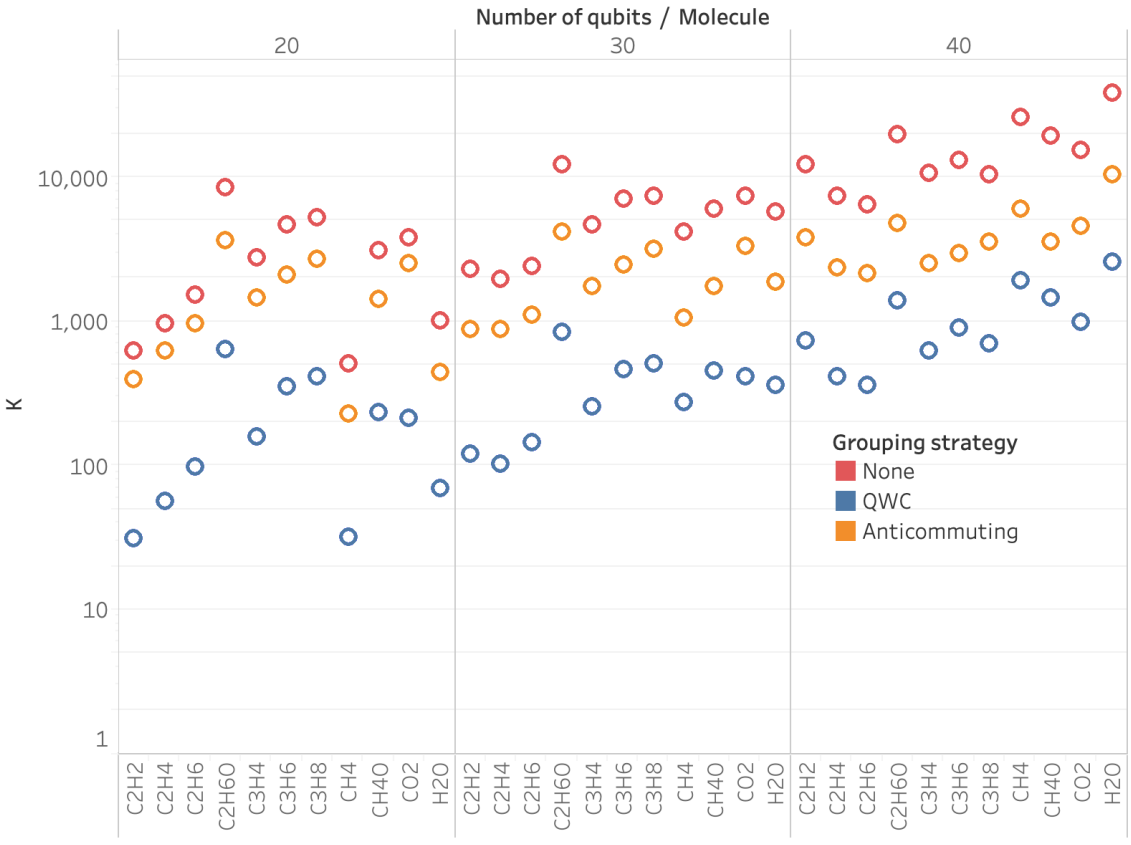

FIG. S10. Values of K computed for various molecules in active spaces of 20, 30 and 40 qubits, using no grouping (red), QWC grouping (blue) and anticommuting grouping (orange) with canonical orbitals, upper bounds for the variances and zero covariances. Core orbitals were frozen in all cases when computing the Hamiltonians.

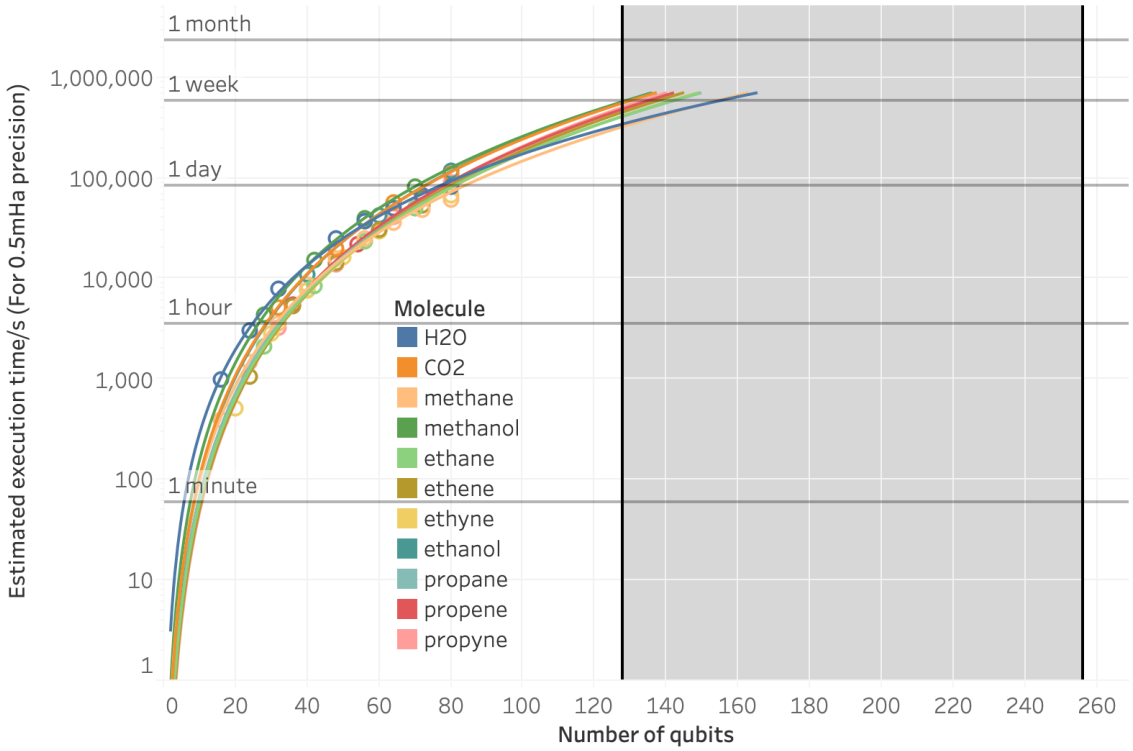

FIG. S11. Extrapolated runtimes (in s) for molecules in our benchmark set, using the values of K computed for the FNO representation of Hamiltonians, variances from CISD data and basis rotation grouping. An additional factor of 2 improvement is assumed from the RDMC technique, and the number of measurements is computed for a precision of 0.5 mHa to leave room for other uncertainties. Circuit depth is assumed to be $5N_{\mathrm{q}} - 3$ two-qubit gates and the two-qubit gate time is assumed to be 100 ns.