Supplementary Information for "Elucidating Reaction Mechanisms on Quantum Computers"

Markus Reiher[a,c,1], **Nathan Wiebe**[b,1,2], **Krysta M Svore**[b], **Dave Wecker**[b], **and Matthias Troyer**[b,c,d]

[a]Laboratorium für Physikalische Chemie, ETH Zurich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland; [b]Station Q Quantum Architectures and Computation Group, Microsoft Research, Redmond, WA 98052, USA; [c]Station Q, Microsoft Research, Santa Barbara, CA 93106-6105, USA; [d]Theoretische Physik and Station Q Zurich, ETH Zurich, 8093 Zurich, Switzerland

1. Introduction to Quantum Computing

Here we provide a brief review of quantum computing for those that are not experts in quantum computing. The aim of this review is to introduce the basic concepts and notations that are used in quantum computing as well as introduce quantum algorithms for simulating chemistry and also quantum error correction. For a more detailed exposition see [1]. These concepts will be used in section 2 where we discuss quantum processes such as phase estimation.

Qubits and Quantum Gates. In quantum computation, quantum information is stored in a quantum bit, or *qubit*. Whereas a classical bit has a state value $s \in \{0, 1\}$, a qubit state $|\psi\rangle$ is a *linear superposition* of states:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right], \qquad [1]$$

where the $\{0, 1\}$ basis state vectors are represented in Dirac notation (called *ket* vectors) as $|0\rangle = \begin{bmatrix}1 & 0\end{bmatrix}^T$, and $|1\rangle = \begin{bmatrix}0 & 1\end{bmatrix}^T$, respectively. The *amplitudes* α and β are complex numbers that satisfy the normalization condition: $|\alpha|^2 + |\beta|^2 = 1$. Upon *measurement* of the quantum state $|\psi\rangle$, either state $|0\rangle$ or $|1\rangle$ is observed with probability $|\alpha|^2$ or $|\beta|^2$, respectively.

Dirac notation is used largely because it contains an implicit tensor product structure that makes expressing qubit states much easier. For example, that the four-qubit state $|0000\rangle$ is equivalent to writing the tensor product of the four states: $|0\rangle \otimes |0\rangle \otimes |0\rangle \otimes |0\rangle = |0\rangle^{\otimes 4} = [1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0]^T$.

A general n-qubit quantum state lives in a 2^n-dimensional Hilbert space and is represented by a $2^n \times 1$-dimensional state vector whose entries represent the amplitudes of the basis states. A superposition over 2^n states is given by:

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle, \text{ such that } \sum_i |\alpha_i|^2 = 1, \qquad [2]$$

where α_i are complex amplitudes and i is the binary representation of integer i. The ability to represent a superposition over exponentially many states with only a linear number of qubits is one of the essential ingredients of a quantum algorithm — an innate massive parallelism.

In a quantum computation, unitary transformations are used to transform quantum states into other quantum states. In particular, the quantum state $|\psi_1\rangle$ of the system at time t_1 is related to the quantum state $|\psi_2\rangle$ at time t_2 by $|\psi_2\rangle = U|\psi_1\rangle$ for a unitary operator U.

In general we cannot expect that a quantum computer can implement every unitary transformation on n qubits exactly because there are an infinite number of such transformations. Instead, gate model quantum computers use a discrete set of quantum gates (which can be represented as $2^n \times 2^n$ unitary matrices) to approximate these continuous transformations. Any such set of gates is known as universal if any unitary transformation can be expressed, within arbitrarily small error, as a sequence of such gates.

The gate set that we consider in this work includes the following gates:

- the `X` gate which is the classical NOT gate that maps $|0\rangle \to |1\rangle$ and $|1\rangle \to |0\rangle$.
- the *Hadamard* gate `H` maps $|0\rangle \to \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|1\rangle \to \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.
- the `Z` gate maps $|1\rangle \to -|1\rangle$ and `T` is the fourth–root of Z, and can be used to interconvert Z and X gates via $\texttt{HZH} = \texttt{X}$.
- the `Y` gate maps $|1\rangle \to -i|0\rangle$ and $|0\rangle \to i|1\rangle$.
- the identity gate is represented by $\mathbf{I}$.
- the two-qubit *controlled-NOT* gate, `CNOT`, maps $|x, y\rangle \to |x, x \oplus y\rangle$. The corresponding unitary matrices are:

$$\texttt{H} = \left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right], \ \texttt{X} = \left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right], \ \texttt{Y} = \left[\begin{smallmatrix}0 & i\\ -i & 0\end{smallmatrix}\right], \ \texttt{Z} = \left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right],$$

$$\texttt{T} = \left[\begin{smallmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{smallmatrix}\right], \mathbf{I} = \left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right], \texttt{CNOT} = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}.$$

Measurement is an exception to this rule; it collapses the quantum state to the observed value, thereby erasing any information about the amplitudes α and β. This means that extracting information from a quantum system irreversibly damages a state. Consequently, although the exponential parallelism that quantum computers naturally possess is mitigated by the fact that most quantum computations will have to be repeated many times to extract the necessary information about the output of the quantum algorithm.

Quantum Circuit Synthesis. In order to understand how the cost estimates of our algorithms are found it is important to understand how arbitrary unitaries can be converted into discrete gate sequences. For our purposes, we consider the Clifford + `T` gate library discussed above, but compilation into other gate sets is also possible [2, 3]. The simplest case to consider is that of single qubit unitary synthesis where $U \in \mathbb{C}^{2\times 2}$. In such cases, an Euler angle decomposition exists such that, up to an irrelevant overall phase,

$$U = e^{i\alpha Z}\texttt{H}e^{i\beta Z}\texttt{H}e^{i\gamma Z}. \qquad [3]$$

Thus the problem of implementing a single qubit transformation reduces to the problem of performing the rotation $e^{i\theta Z} = cos(\theta)\mathbf{I} + i\sin(\theta)Z$. The general case of $U \in \mathbb{C}^{2^n \times 2^n}$ similarly reduces to instances of single qubit synthesis interspersed with entangling gates such as `CNOT`. Thus implementing single qubit rotations can be seen as an atomic operation on which compilation process for U is based.

There are a host of methods known for decomposing rotations into Clifford and `T` gates. Perhaps the earliest such algorithm is the Solovay–Kitaev algorithm [4], which provided an efficient method for performing this decomposition based on Lie–algebraic techniques. In recent years, much more efficient methods for decomposition have been discovered that are based on number theory. These methods require near–quartically fewer operations than the Solovay–Kitaev algorithm would need [4] and are useful, if not necessary, for reducing the gate counts of the simulation algorithm to a palatable level.

In the absence of ancillae, the cost required to approximate an axial rotation within error ϵ for the worst possible input angle, $C(\epsilon)$, is known to lie within the interval [5]

$$4\log_2(1/\epsilon) - 9 \leq C(\epsilon) \leq 4\log_2(1/\epsilon) + 11. \quad [4]$$

Subsequent work [6] has provided a method that has an average case complexity roughly 3/4 of this worst case bound and also has made the compilation process efficient [7].

More recent methods have introduced the use of ancillae to reduce the costs of synthesis [8]. These methods can substantially reduce the number of `T` gates required to perform the synthesis. This approach can reduce the scaling of the average number of `T` gates needed to implement an axial rotation

$$C(\epsilon) \approx 1.15\log_2(1/\epsilon) + 9.2. \quad [5]$$

This approach requires one additional ancilla qubit.

Asymptotically better methods exist, such as repeat-until-success (RUS) synthesis with fallback [8], but these methods do not outperform this method given the range of ϵ that we require for the simulation to reach chemical accuracy. Further methods such as PAR rotations [9] or gearbox circuits [10] can be used to substantially reduce the `T`–depth of the simulation circuits at the price of requiring greater parallel width. While we do not consider the impact of gearbox synthesis here, we will examine the costs and benefits of PAR rotations.

An important issue arises when using such methods for parallel execution. If several rotations are implemented simultaneously on different qubits and RUS synthesis is used then the dominant contribution to the cost is given by the longest sequence of gates used in each block of parallelized rotations. In particular, while RUS synthesis reduces the expectation value by a factor of roughly 4 from the worst case bounds, it is clear that when performing many of them in parallel it is very likely that at least one of them will saturate the bound. For this reason we use upper bounds on deterministic synthesis given by Eq. (4), which have a better worst case scaling.

Simulating Quantum Chemistry. Our approach to quantum chemistry simulations closely follows the strategy proposed in Ref. [11]. We begin by representing the quantum chemistry Hamiltonian in a second quantized form, keeping track of the locations of the electrons by using the occupations of a discrete basis of spin-orbitals. Since a spin orbital can only contain one electron, such states are very natural to express in a quantum computer. Each spin orbital is assigned a qubit where the state $|1\rangle$ corresponds to an occupied orbital and $|0\rangle$ an unoccupied orbital.

These states are often described using creation and annihilation operators which obey $a^\dagger|0\rangle = |1\rangle$, $a^\dagger|1\rangle = 0$, $a|1\rangle = |0\rangle$ and $a|0\rangle = 0$. Since these operators correspond to creating electrons, they must respect the symmetries appropriate for Fermions. The most important property is the anti–commutation property $\{a_i^\dagger, a_j\} = \delta_{i,j}$. This means that while it may be tempting to identify $a^\dagger = (\mathtt{X} - i\mathtt{Y})/2$, such a representation does not satisfy the anti–commutation relation for Fermionic operators. Instead, these creation and annihilation operators can be converted into Pauli operators using the Jordan–Wigner or Bravyi–Kitaev [12] transformations.

In order to simulate the dynamics of the system we need to emulate the unitary dynamics that the system undergoes using a sequence of gate operations. This dynamics is of the form e^{-iHt} where

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2}\sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \quad [6]$$

Once these fermionic creation and annihilation operators are translated into Pauli operators using, for example, the Jordan–Wigner transformation then the Hamiltonian can be translated into a sequence of operations that individually could be performed on a quantum computer. Well known circuits exist for performing the exponentials yielded by the Trotter–Suzuki formulas [1, 13, 14].

The most direct method for estimating the ground-state energy is phase estimation (see Section 2 for more detail). Phase estimation uses quantum interference between controlled executions of $\mathbf{I}, e^{-iHt_1}, e^{-iHt_2}, \ldots$, to infer the eigenvalues of e^{-iHt}. If t is smaller than $\pi/\|H\|$ this also directly yields the eigenvalues of H. Apart from the ability to directly sample from the eigenvalues, a further advantage to phase estimation is that it requires $O(1/\epsilon)$ applications of e^{-iHdt} to learn its eigenphase within error ϵ with high probability. This is quadratically better than bounds on the variance that would be seen from optimal classical sampling methods using an unbiased estimator.

It is necessary to apply phase estimation on an initial state that has large overlap with the ground state in order to find the ground-state energy with high probability,. Specifically, if applied to an initial state $|\psi\rangle = a_g|E_G\rangle + \sqrt{1-|a_g|^2}\left|E_G^\perp\right\rangle$ then PE returns an estimate of ground-state energy E_G with probability $|a_g|^2$. The simplest state to use is the Hartree–Fock state, which only requires applying a sequence of NOT gates to the state $|0\rangle^n$, however configuration interaction with sufficiently high excitations may be required to achieve high overlap for systems that have strong correlations in their ground states. We do not consider the cost of preparing such a state here, since such a cost of preparing a sufficiently accurate approximation to the ground state of FeMoco is difficult to determine in absentia of large scale quantum computers. We discuss this issue in more detail in Section 6.

Quantum error correction. Quantum hardware is far less robust to errors as classical hardware. Quantum error correction provides a way to reduce the errors in the device without sacrificing the quantum nature of the system. Quantum error correcting codes require that the physical error rates, say of

qubits, quantum gates, and measurements, are less than a given threshold value. This threshold value depends on the error correcting code being used and the type of noise of the system. If error rates are below the threshold, then errors in the computation, referred to as the logical gates and logical qubits, can be made arbitrarily small at only a polylogarithmic overhead.

The surface code is currently the most popular code due in part to its relatively high threshold of 1% [15]. Much of this enthusiasm has arisen because of evidence that existing superconducting quantum computers may already have error rates near this threshold [16]. This raises the hope that a fault-tolerant, scalable quantum computer may be just over the horizon.

While quantum error correction promises the ability to perform arbitrarily long quantum computations on a noisy device, the resources required to execute such a fault-tolerant computation can be large if the system operates close to the threshold. The majority of the cost of quantum error correction arises from the need to perform a universal set of quantum gates. While protecting so-called Clifford gates requires very little overhead in the surface code, in comparison protecting a non-Clifford gate, such as a T or Toffoli gate, requires substantial resources. While several techniques for producing fault-tolerant non-Clifford gates exist [17–20], here we focus on the use of magic state distillation in conjunction with the surface code [15]. Magic state distillation [21] takes as input a set of noisy resource states and outputs a cleaner resource state. For the surface code, we must distlll the T state as it cannot be implemented directly in the code. Here we consider the $15-1$ distillation scheme of Bravyi and Kitaev [21], where 15 noisy input states with error rate p produce a single magic resource state with error rate roughly $35p^3$. Magic state distillation consists only of Clifford operations, which can be implemented easily within the surface code.

Further details about the architectural issues that arise when implementing quantum error correction in future devices can be found in [22–24].

2. Implementing Phase Estimation

In this section, we discuss how to implement the phase estimation protocol in quantum computers. This is important to our subsequent estimates of the complexity of simulating nitrogenase's FeMoco because phase estimation constitutes the outer most loop of the quantum simulation and hence is a major driver of the cost of the simulation.

Phase Estimation. Phase estimation is one of the most critical components of the quantum simulation algorithm. Without phase estimation, the amount of simulation time needed to estimate the ground-state energy would grow quadratically as ϵ^{-2} with the precision ϵ required. Since ϵ is on the order of 0.1mHartree, the phase must be estimated for one time step of the evolution operator within an error $0.1/r$ mHartree where r is the number of Trotter steps required. If statistical sampling, rather than phase estimation, were used to estimate the phase then on the order of 10^{12} experiments would be needed to make the variance in the estimate sufficiently small. This would be impractical and so phase estimation is crucial for most large scale applications in quantum chemistry.

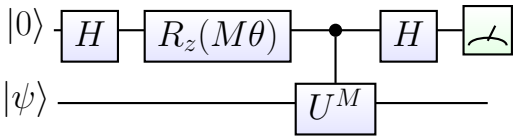

Fig. 1. Iterative phase estimation. This circuit is used in iterative phase estimation algorithms wherein the eigenvalues are inferred from measurement statistics.

The standard quantum phase estimation algorithm [1] allows eigenvalues to be learned within error ϵ with probability at least $1/2$ uses a number of applications of the unitary circuit, $M(\epsilon)$, that is bounded above by

$$M(\epsilon) \leq \frac{16\pi}{\epsilon}. \quad [7]$$

This value is far from optimal time scaling of π/ϵ, but has the advantage of requiring neither measurements of the quantum system nor a classical computer to infer the most likely eigenvalue. We will also use this result below because it provides an upper bound on the cost of the optimal phase estimation algorithm.

An alternative approach is to use iterative phase estimation. Iterative phase estimation forgoes storing the phase in a quantum register and instead uses a classical inference algorithm to learn the eigenphase from measurements of an auxiliary probe qubit that is iteratively measured and re-entangled with the system. Kitaev proposed the first variant of iterative phase estimation. The circuit used in this process is given in Fig 1.

The ultimate limit that can be achieved, in terms of the number of times the unitary is applied as a function of desired error tolerance, is given by [25–27]

$$M(\epsilon) \approx \frac{\pi}{\epsilon}. \quad [8]$$

Furthermore, Gaussian strategies such as rejection filter phase estimation (RFPE) [28] yield Bayesian Cramer–Rao bounds that come close to saturating this: $3.3/\epsilon$. As these bounds are often saturated for likelihood functions of this form [29] and because the Gaussian assumption causes the user to throw away substantial information from higher moments in the posterior distribution, we expect that Eq. (8) represents the ultimate limit for phase estimation and that based on previous studies the use of optimized policies [30–32] will enable performance that comes close to saturating this limit. As a result we use $M(\epsilon) \approx \pi/(2\epsilon)$ as a surrogate for the expected performance of these optimized approaches, where the factor of 2 difference follows from an important optimization that holds for the case of quantum simulation algorithms that we discuss in detail in the following section.

Controlled rotations. While the previous discussion only showed how to perform a `Z`–rotation, we need to perform controlled `Z`–rotations to perform the phase estimation algorithm. Fortunately, there are well known methods that can be used to perform such controlled rotations using a pair of rotations and two `CNOT` gates. A better approach, previously shown in unpublished work by Tsuyoshi Ito in 2012, is given in Figure 2 and the validity of the circuit is proven in the following lemma.

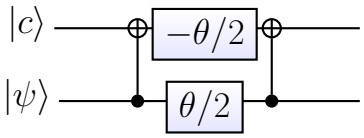

Fig. 2. Controlled Rotation. Low depth circuit for controlled Z–rotations where the θ gate represents $e^{i\theta Z}$.

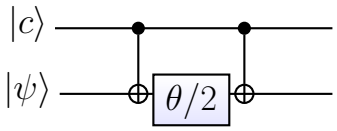

Fig. 3. Rotation with conditional direction. Circuit used to implement analogue of controlled Z–rotations used in Lemma 2, which is a rotation whose sign is conditioned on the control qubit.

Lemma 1. *The circuit of Fig. 2 implements the controlled operation $\Lambda(R_Z(2\theta))$ where the top most qubit is the control.*

Proof. Assume that $|c\rangle = |0\rangle$ or $|1\rangle$ then the CNOT gate performs

$$|c\rangle|\psi\rangle \mapsto (\mathbf{X}^c \otimes \mathbf{I})(a|00\rangle + b|11\rangle). \quad [9]$$

The rotation gates then prepare the state

$$(ae^{i(1-(-1)^c)\theta/2}|c\rangle|0\rangle + be^{-i(1-(-1)^c)\theta/2}|c\oplus 1\rangle|1\rangle). \quad [10]$$

Finally the CNOT gate yields the state

$$\begin{aligned}&|c\rangle(ae^{i(1-(-1)^c)\theta/2}|0\rangle + be^{-i(1-(-1)^c)\theta/2}|1\rangle)\\ &\quad = (\Lambda(R_z(2\theta))|c\rangle|\psi\rangle)\,. \end{aligned} \quad [11]$$

Therefore the circuit functions properly for either $|c\rangle = |0\rangle$ or $|c\rangle = |1\rangle$. By linearity it is also valid for any input. □

This form of a controlled rotation is well suited for many quantum simulation applications because it allows the controlled evolutions to be replaced with evolutions that control only on the single qubit rotations in the circuits for implementing the individual terms in the Hamiltonian. However, it is not optimal here because we can implement a variant of a controlled rotation gate to in effect double the impact that the eigenphase has on the measurement probabilities.

Lemma 2. *Let $U = \prod_{j=1}^{N_{\text{exp}}} B_j(\mathbf{I}\otimes R_z(\phi_j))B_j^\dagger$ where B_j are unitary transformations and assume that it is chosen such that $U^\dagger = \prod_{j=1}^{N_{\text{exp}}} B_j(\mathbf{I}\otimes R_z(-\phi_j))B_j^\dagger$ then if $U|\phi\rangle = e^{i\phi}$ then there exists a quantum protocol that uses an M rotations and samples from a Bernoulli distribution such that the probability of the protocol outputting 0 is*

$$P(0|\phi;\theta,M) = \frac{1+\cos(2M[\phi+\theta])}{2}.$$

Proof. Our proof is constructive and follows directly from the arguments presented informally in [33]. The idea behind the protocol is to replace every controlled rotation used in the circuit in Fig. 1 with the circuit in Fig. 3. Formally we denote this by replacing the controlled operation $\Lambda(U^M)$ with W^M, which is defined such that $W|0\rangle|\psi\rangle = |0\rangle U|\psi\rangle$ and $W|1\rangle|\psi\rangle = |1\rangle U^\dagger|\psi\rangle$. Since $XR_z(\theta_j)X = R_z(-\theta_j)$, the operation W can be written as

$$W = \prod_{j=1}^{N_{\text{exp}}} B_j\Lambda(X)(\mathbf{I}\otimes R_z(-\phi_j))\Lambda(X)B_j^\dagger, \quad [12]$$

where the controlled–not operation $\Lambda(X)$ uses the control qubit as its control and the qubit that R_z acts on as its target. This circuit is shown in Fig. 3.

Applying $(\mathtt{H}\otimes\mathbf{I})W(\mathtt{H}\otimes\mathbf{I})$ to the eigenstate $|0\rangle|\phi\rangle$, such that $U|\phi\rangle = e^{i\phi}|\phi\rangle$, yields

$$\frac{1}{2}\left(|0\rangle|\phi\rangle(1+e^{2i(\theta-\phi)M}) + |1\rangle|\phi\rangle(1-e^{2i(\theta-\phi)M})\right), \quad [13]$$

up to a global phase. The probability of measuring the ancilla qubit to be zero is $(1+\cos(2M(\phi-\theta)))/2$ as claimed. □

Lemma 2 shows that the number of rotations needed to perform phase estimation is 1/4 the value that would be expected if circuits such as that of Fig. 2 were used (or 1/2 the cost in parallel settings). As an example, for the case of RFPE numerical experiments show that we can learn the eigenphase within error ϵ with probability 1/2 using

$$M(\epsilon) \approx \frac{2.3}{\epsilon}. \quad [14]$$

In comparison, the ultimate lower limit given by previous studies becomes $\pi/(2\epsilon)$ and the Bayesian Cramer–Rao bound gives a lower limit of approximately $1.6/\epsilon$ under the assumptions of Gaussian priors made in RFPE [28]. We show in Section 8 that the lower bound on the depth can be reached using a new adaptive approach to phase estimation that uses a small cluster of quantum computers. Similarly the upper bounds on the cost of traditional QPE in [1] become $8\pi/\epsilon$ rather than $16\pi/\epsilon$ if these circuits are used.

While the assumption that the underlying Trotter–Suzuki formula can be inverted by simply inverting the sign of the evolution time applies for the $(2k+1)^{\text{th}}$ order Trotter-Suzuki formulas for all $k \geq 1$, it does not apply to the second order formula unless further assumptions are made. We can see this from the fact that

$$\left(\prod_{j=1}^{N} e^{-iH_j t}\right)\left(\prod_{j=1}^{N} e^{iH_j t}\right) = 1 + O(t^2).$$

If we assume that we are interested only in the ground-state energy and the Hamiltonian is real valued (like in the quantum chemistry applications that we consider) then the error in assuming that $U^\dagger$ can be formed by simply flipping the signs is $O(t^3)$ [34], which is by no means fatal but it could potentially contribute to the error in Trotter–Suzuki decompositions. Whereas if the third (or higher) order Trotter–Suzuki formula is used then no such danger exists. This, along with the superior bounds proven for the error in the third order formula, provide the justification for using this formula in preference to the asymptotically equivalent second–order formula.

3. Trotter errors

Here we discuss the issue of how errors from the use of Trotter-Suzuki formulas lead to systematic errors in the ground-state energy estimates output by phase estimation. We further discuss the methodology that we use to upper bound these errors and also give empirical estimates of the scaling of such errors.

Rigorous bounds. A major source of error in most quantum simulation algorithms arises from the use of Trotter formula expansions. Such errors can be made arbitrarily small but the need to make such errors smaller than chemical accuracy means that the cost of doing so can substantially impact the time required to perform the simulation. In this appendix we will provide a detailed discussion about how to bound the error in low–order Trotter formulas.

Trotter errors arise from the fact that the terms in the Hamiltonian used in the expansion do not commute. In principle, the Zassenhaus formula provides everything that is needed in order to understand the scaling of these errors:

$$e^{(A+B)t} = e^{At}e^{Bt}e^{\frac{1}{2}[A,B]t^2} + O(t^3), \quad [15]$$

and thus

$$\|e^{(A+B)t} - e^{At}e^{Bt}\| \in O(\|[A,B]\|t^2). \quad [16]$$

Although this expression is useful in estimating the scaling of simulation errors, the question that we are interested in is somewhat orthogonal to this. Instead, we are interested in the errors in estimated eigenvalues. The Baker–Campbell–Hausdorff formula, which is the dual to the Zassenhaus formula, can be used to estimate these errors. If $H = \sum_{\alpha=1}^{L} H_\alpha$ where the H_α correspond to terms in Eq. (6) then the second order Trotter–Suzuki formula (also known as the Strang splitting) gives

$$\prod_{\alpha=1}^{L} e^{-iH_\alpha t/2} \prod_{\alpha=L}^{1} e^{-iH_\alpha t/2} = e^{-iH_{\text{eff}}t}, \quad [17]$$

where $H - H_{\text{eff}}$ is

$$-\frac{1}{12}\sum_{\alpha\leq\beta}\sum_{\beta}\sum_{\alpha'<\beta}[H_\alpha(1-\frac{\delta_{\alpha,\beta}}{2}),[H_\beta,H_{\alpha'}]]t^2 + O(t^3). \quad [18]$$

This shows that, to leading order, the error in the Strang splitting can be estimated by the ground–state expectation value of a double commutator sum. Furthermore, since $L \in O(N^4)$ it is clear from this form that the error in the Trotter formula must scale at most as $O(N^{10}t^2)$. It is not $O(N^{12}t^2)$ because the commutator structure restricts two of the orbitals H_α and H_β act on. Although this estimate can be computed in polynomial time, there are too many terms for this to be reliably estimated for molecules on the scale of nitrogenase even using Monte Carlo methods [35].

An upper bound on the asymptotic scaling can be trivially found by applying the triangle inequality to Eq. (18). This approach is not appropriate for our purposes because we do not know a priori how large t must be for the leading order term to be dominant. As a result, we use the following result, which is provably an upper bound on the error in the energy of the evolution:

$$\Delta E_{\text{TS}}/t^2 \leq 4\sum_{\alpha,\beta,\alpha'}\|H_\alpha\|\|H_\beta\|\|H_{\alpha'}\| \times (\delta_{\alpha>\beta}\delta_{\alpha'>\beta} + \delta_{\beta>\alpha}\delta_{\alpha',\alpha})W(\alpha,\beta,\alpha'), \quad [19]$$

where $W(\alpha,\beta,\alpha')$ is an indicator function that takes the value 1 if and only if the corresponding double commutator is non–zero. Note that this is slightly tighter than the bound used in (16) of [36].

There are several criteria that we know a priori lead to a double commutator vanishing:

1. $H_{\alpha'}$ and H_β act on disjoint sets of qubits.
2. H_α acts on a disjoint set of qubits from the set of qubits that $H_{\alpha'}$ and H_β act on.
3. $H_{\alpha'}$ and H_β correspond to PP or PQQP terms.
4. $H_{\alpha'}$ and H_β correspond to PR and PQQR terms with the same P and R.
5. Only one of $[H_\alpha,[H_\beta,H_{\alpha'}]]$, $[H_\beta,[H_{\alpha'},H_\alpha]]$ and $[H_{\alpha'},[H_\alpha,H_\beta]]$ is non–zero according to the prior rules (Jacobi identity).

There are other symmetries to the terms that can be used to argue that even more terms are necessarily zero. Since we do not consider these properties, we overcount the contribution of any such terms and so our estimate remains an upper bound.

This bound is much more computable than the original expression, but is computationally challenging to compute exactly owing to the $O(N^{10})$ terms in the double commutator sum. The Cauchy–Schwarz inequality can be used to convert this expression into one that can be computed in $O(N^4)$ operations, but doing so can over-represent the influence of large terms in the Hamiltonian [36].

Empirical Estimates of Trotter Error. Given that computation of the matrix elements of the error operator in Eq. (18) is computationally challenging, we rely on Monte Carlo sampling to estimate the upper bound in Eq. (19). Monte Carlo sampling is much more effective here because the use of the triangle inequality removes the alternating sign that we see when summing the original series. We achieve this by drawing M samples uniformly at random for $\{\alpha_j : j = 1\ldots M\}$, $\{\beta_j : j = 1\ldots M\}$ and $\{\beta_j : j = 1\ldots M\}$. We then reject the sample if $(\delta_{\alpha>\beta}\delta_{\alpha'>\beta} + \delta_{\beta>\alpha}\delta_{\alpha',\alpha})W(\alpha,\beta,\alpha') = 0$, and otherwise compute the product of the products of the norms of the three corresponding Hamiltonian terms. If there are L terms in the Hamiltonian and define

$$\Gamma(\alpha,\beta,\alpha') := 4\|H_\alpha\|\|H_\beta\|\|H_{\alpha'}\| \times (\delta_{\alpha>\beta}\delta_{\alpha'>\beta} + \delta_{\beta>\alpha}\delta_{\alpha',\alpha})W(\alpha,\beta,\alpha') \quad [20]$$

then

$$\Delta E_{\text{TS}} \lessapprox \frac{L^3}{M}\sum_{j=1}^{M}\Gamma(\alpha_j,\beta_j,\alpha'_j)t^2 := ht^2. \quad [21]$$

This implies that, for a fixed h, if we wish to achieve an error ϵ in the eigenvalues of the Trotter–Suzuki expansion then it suffices to pick

$$t = \sqrt{\frac{\epsilon}{h}}. \quad [22]$$

The variance in this estimator is

$$\frac{L^6 \mathbb{V}_{\alpha,\beta,\alpha'}(\Gamma(\alpha,\beta,\alpha'))t^4}{M}, \quad [23]$$

which implies using Chebyshev's inequality that with probability greater than 75% the sample error is less than 2ϵ if

$$M \geq \frac{L^6 \mathbb{V}_{\alpha,\beta,\alpha'}(\Gamma(\alpha,\beta,\alpha'))t^4}{\epsilon^2} = \frac{L^6 \mathbb{V}_{\alpha,\beta,\alpha'}(\Gamma(\alpha,\beta,\alpha'))}{h^2}. \quad [24]$$

In practice, uniform sampling is not necessarily the best option because the importance of the different terms can vary wildly within a class. For example, the one-body terms tend to be much larger than the two-body terms but the two-body terms are far more numerous. This means that uniform sampling can underestimate the contributions of such terms because of their relative scarcity in the sample space.

We combat this by sampling from each *type* of double commutator. In particular, we sample over a class of double commutators such as [PQ, [PP, PQ]] and uniformly draw PQ, PP and PQ terms to estimate those terms contribution to the overall error. The total error is then the sum of the estimates over all such classes. We use a minimum of 10^8 samples per class, which renders the sample standard deviation in our estimates of the error less than 1%.

The resultant bounds can be seen in Fig. 4 wherein we examine the predicted Trotter numbers and the empirically observed Trotter numbers for small molecules. The rough scaling of the upper bound on the Trotter number that we see corresponds to $N^{2.5}$ which was noted in previous numerical studies [36], but owing to the scatter of the data due to the widely varying chemical properties of the molecules, this scaling should not be seen as definitive. We observe that the upper bounds seem to be roughly a factor of 10 000 times too loose for small molecules. For this reason we plot three reasonable extrapolations of the scaling based on the numerical results for small molecules. The most pessimistic bound rescales the scaling extracted from the upper bound such that all the data remains beneath the curve. The middle one rescales the upper bound data by the average discrepancy between the upper bound and the numerically computed examples. The most optimistic curve is simply a polynomial fit to the numerical data that ignores the upper bound. We expect the middle curve to be the most realistic estimates, but provide resource estimates for these three cases below as well as results that follow from using the upper bound.

4. Error propagation

Here we provide proofs of some basic results that we will use to propagate these errors through the quantum simulation. These results are crucial for the cost estimates in the subsequent section because they show how large the worst case errors can be in the eigenvalue estimation given errors of these magnitudes. It is worth noting that we expect these results to yield substantial overestimates of the error because they do not consider the natural cancellations that are likely to occur in practical eigenvalue estimation problems.

Lemma 3. *Let A and B be Hermitian operators acting on finite dimensional Hilbert spaces such that $\|A-B\|_2 \leq \epsilon$ and $A|\psi_A\rangle = E_A|\psi_A\rangle$ and $B|\psi_B\rangle = E_B|\psi_B\rangle$ where E_A and E_B are the smallest eigenvalues of either operator then $|E_A - E_B| \leq \epsilon$.*

Proof. Because A and B are Hermitian they satisfy the variational property meaning that

$$E_B \leq \langle\psi_A|B|\psi_A\rangle. \quad [25]$$

Since $\|A-B\|_2 \leq \epsilon$ it follows that there exists C such that $\|C\|_2 \leq 1$ and $A = B + \epsilon C$. This implies that

$$E_B \leq \langle\psi_A|A+\epsilon C|\psi_A\rangle = E_A + \epsilon\langle\psi_A|C|\psi_A\rangle. \quad [26]$$

Thus

$$E_B - E_A \leq \epsilon\langle\psi_A|C|\psi_A\rangle. \quad [27]$$

If $E_A \leq E_B$ then we have from the definition of $\|\cdot\|_2$ that

$$|E_B - E_A| \leq \epsilon|\langle\psi_A|C|\psi_A\rangle| \leq \epsilon. \quad [28]$$

Now assume that $E_B > E_A$. We then have

$$E_A \leq \langle\psi_B|A|\psi_B\rangle. \quad [29]$$

Then by repeating the same argument we conclude that $|E_B - E_A| \leq \epsilon$ regardless of the sign of $E_A - E_B$. □

Now we will go beyond this bound to show that the error scaling in the eigenvalues of the unitary evolutions generated by two similar Hamiltonians is no more pathological than the scaling of errors in the ground-state energies.

Lemma 4. *Assume that for Hermitian bounded operators A and B acting on a finite dimensional Hilbert space $\|e^{-iAt} - e^{-iBt}\|_2 \leq t\gamma(t)$ for $\gamma(t)$ a non–decreasing continuous function of t on $[0,\infty)$ then $\|A-B\|_2 \leq \gamma(t)$.*

Proof. Using standard bounds [1], we have that $\|e^{-iAt} - e^{-iBt}\|_2 \leq \|A-B\|_2 t$ and furthermore from Taylor's theorem $\|e^{-iAt} - e^{-iBt}\|_2 = \|A-B\|_2 t + O([\|A-B\|_2 t]^2)$. Therefore the former upper bound is tight in the limit as $t \to 0$. By assumption $\|e^{-iAt} - e^{-iBt}\|_2 \leq t\gamma(t)$ for all t in a compact subinterval containing 0. Assume that $\lim_{t\to 0}\gamma(t)/\|A-B\|_2 < 1$. This implies that there exists a compact interval containing 0 such that for all t in this interval $\|e^{-iAt} - e^{-iBt}\|_2 > \|A-B\|_2 t$, which leads to a contradiction because we have already demonstrated that $\|A-B\|_2 t$ is a tight bound on the error in this limit. Therefore $\gamma(t) \geq \lim_{t\to 0}\gamma(t) \geq \|A-B\|_2$ under the assumptions of the lemma. □

Theorem 1. *Let $H = \sum_{j=1}^{M} H_j$ where each H_j is a bounded Hermitian operator acting on a finite dimensional Hilbert space. Furthermore, let $e^{-i\tilde{H}_j t/2}$ be exponentials of individual terms in the Hamiltonian yielded by a synthesis process that approximates them within error at most δt in $\|\cdot\|_2$ for any $t \geq 0$. Finally, let $\|e^{-iHt} - \prod_{j=1}^{M} e^{-iH_j t/2}\prod_{j=M}^{1} e^{-iH_j t/2}\|_2 \leq \Delta E_{TS}(t)t$. Then the difference in ground-state energies between $H(t)$ and $\tilde{H}(t) := i\log(\prod_{j=1}^{M} e^{-i\tilde{H}_j t/2}\prod_{j=M}^{1} e^{-i\tilde{H}_j t/2})/t$ is at most $\Delta E_{TS}(t) + (2M-1)\delta$.*

Proof. First, since $\tilde{H}_j$ is Hermitian

$$\prod_{j=1}^{M} e^{-i\tilde{H}_j t/2}\prod_{j=M}^{1} e^{-i\tilde{H}_j t/2}$$

is a unitary operator. The matrix logarithm is defined if and only if the matrix in question is invertible and hence

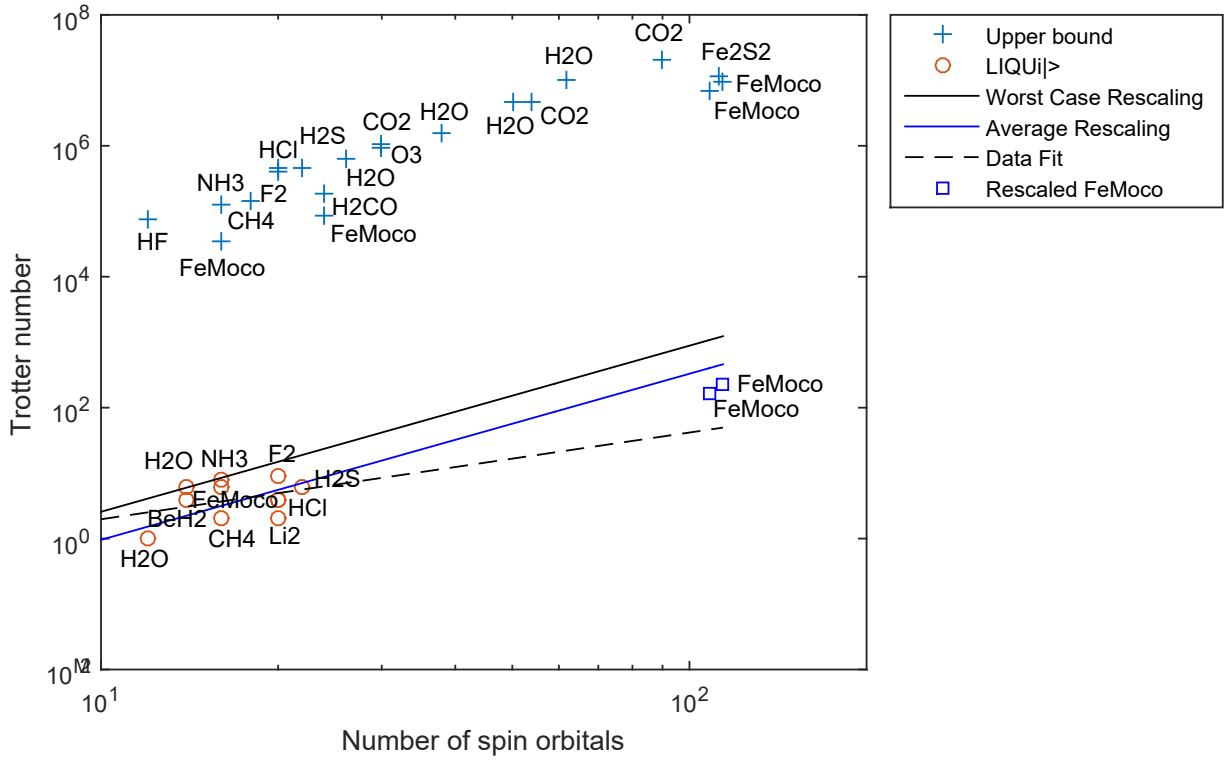

Fig. 4. Trotter number $(1/dt)$ needed to reach 0.1 mHartree of accuracy assuming no errors from synthesis or phase estimation. Lines represent projections based on current data.

the matrix logarithm exists because unitary matrices are invertible. The logarithm is then clearly an anti–Hermitian operator and hence $\tilde{H}(t)$ is Hermitian. This implies that $\prod_{j=1}^{M} e^{-iH_j t/2} \prod_{j=M}^{1} e^{-iH_j t/2} \equiv e^{-i\tilde{H}(t)t}$.

In order to use Lemma 4 to prove the result, we need to have constant operators. However, the particular $\tilde{H}_j$ applied at each step depends on the value of t used. This can be decoupled by introducing a parameter, s, such that $\tilde{H}_j(s=t)$ corresponds to the $\tilde{H}_j$ at $s=t$.

The triangle inequality implies that

$$\|e^{-iHt} - e^{-i\tilde{H}(t)t}\|_2 \leq \max_s \|e^{-iHt} - e^{-i\tilde{H}(s)t}\|_2 \leq \left\| e^{-iHt} - \prod_{j=1}^{M} e^{-iH_j t/2} \prod_{j=M}^{1} e^{-iH_j t/2} \right\|_2 + \max_s \left\| \prod_{j=1}^{M} e^{-iH_j t/2} \prod_{j=M}^{1} e^{-iH_j t/2} - e^{-i\tilde{H}(s)t} \right\|_2 . \quad [30]$$

Then using our assumptions about the error in each exponential for all s and standard inequalities for the errors in unitary operations [1] this error is at most

$$[\Delta E_{TS}(t) + (2M-1)\delta]t. \quad [31]$$

This is a non-decreasing function of t so applying Lemma 4 yields $\|H - \tilde{H}(s)\|_2 \leq \Delta E_{TS}(t) + (2M-1)\delta$ for all s. The result then follows from taking $s=t$ and applying Lemma 3. □

This result trivially extends to the case of quantum chemistry simulation where each H_j is an n–qubit Pauli operator. In such cases, it suffices to choose synthesis error that shrinks linearly with the timestep used in the Trotter decomposition. Since the cost of circuit synthesis of rotations in the Clifford + T gate library scales logarithmically with $\delta = \Delta_{\text{synth}}/t$ [5], where Δ_{synth} is the synthesis error in the quantum circuit.

5. Cost estimates for nitrogenase

Fundamentally, two factors contribute to the cost of the quantum simulation (assuming that the user can prepare an exact copy of the ground state at negligible cost). The first is the cost of implementing the Trotter decomposition of the Hamiltonian and the second is the number of times that the Trotter circuit must be repeated in the phase estimation algorithm.

One might object that the number of time steps required in the Trotter–Suzuki decomposition also is a driving factor in the cost. Of course the Trotter decomposition is a major driver of the cost, but it comes in only indirectly through the cost of phase estimation. This is because, in principle, the phase estimation algorithm learns the eigenphases of a single Trotter step. The Trotter error can be made arbitrarily small by choosing shorter evolution times, but this in turn requires the phase estimation algorithm to take more steps. As the cost of phase estimation scales inversely with the desired uncertainty, this causes the cost to scale inversely with the time step used in the Trotter–Suzuki decomposition. Thus the cost of the simulation can be thought of as arising from only two sources, the cost of each depending on the error tolerances allowed for all three contributions to the error.

If we then define ϵ_1 to be the error in phase estimation,

Upper Bound			LIQUi|⟩		
Molecule	Spin Orbitals	Basis	Molecule	Spin Orbitals	Basis
HF	12	sto6g	H2O (frozen core)	12	sto6g
FeMoco	16	tzvp	BeH2	14	sto6g
NH3	16	sto6g	H2O	14	sto6g
CH4	18	sto6g	CH4 (frozen core)	16	sto6g
F2	20	sto6g	FeMoco	16	tzvp
HCl	20	sto6g	NH3	16	sto6g
H2S	22	sto6g	Li2	20	sto6g
FeMoco	24	tzvp	HCl	20	sto6g
H2CO	24	tzvp	F2	20	sto6g
H2O	26	p321	CH4	20	sto6g
CO2	30	sto3g	H2S	22	sto6g
O3	30	sto6g			
H2O	38	dzvp			
H2O	50	p6311ss			
CO2	54	p321			
H2O	62	p6311ss			
CO2	90	dzvp			
FeMoco	108	tzvp			
Fe2S2	112	sto3g			
FeMoco	114	tzvp			

Table 1. Table contains the identities of each molecule sorted first by the number of *spin* orbitals, which is twice the number of spatial orbitals, and then by the actual, or upper bounded, Trotter number.

$\epsilon_2 := \Delta E_{\mathrm{TS}}$ to be the error in the Trotter–Suzuki expansion and $\epsilon_3 := (2M-1)\Delta_{\mathrm{synth}}/t$ to be the error in circuit synthesis then it follows from the triangle inequality and Thoerem 1 that the error in the ground-state energy is at most 0.1 mHartree if

$$\epsilon_1 + \epsilon_2 + \epsilon_3 \leq \epsilon := 10^{-4}\mathrm{Ha}. \qquad [32]$$

In this section we will focus on the target of 0.1 mHa level of accuracy, which is appropriate for quantitative calculations of the reaction rates.

We estimate the cost of the circuit using the number of T gates required in the algorithm, which is a function of the form

$$C = 2M \left\lceil \frac{\alpha}{\epsilon_1} \right\rceil \left\lceil \beta \sqrt{\frac{\epsilon}{\epsilon_2}} \right\rceil \left(\gamma \log_2 \left(\frac{2M}{\epsilon_3} \left\lceil \beta \sqrt{\frac{\epsilon}{\epsilon_2}} \right\rceil \right) + \delta \right), \quad [33]$$

and then optimize over ϵ_1, ϵ_2 and ϵ_3 to minimize C subject to the constraint in Eq. (32). This functional form follows directly from the phase estimation algorithm. If iterative phase estimation requires K experiments, each of which requires a simulation with R rotations and each rotation requires S T gates then the overall cost is $C = 2MKRS$. The factor of $2M$ comes from the number of exponentials in the Trotter–Suzuki decomposition. The functional forms for K, R and S then come from Eq. (7), Eq. (22) and Eq. (4) respectively. The log factor contains a factor of $2M\lceil\beta\sqrt{\epsilon/\epsilon_2}\rceil$ because we need to make sure that the sum of the errors in the eigenvalues due to synthesis synthesis add up to at most ϵ_3.

Here α is the scaling constant for the phase estimation algorithm used, β is the Trotter number (or multiplicative factor by which t is decreased from 1 Ha^{-1}) needed to achieve an error of $\epsilon = 0.1$ mHa in the ground-state energy estimate and γ and δ are the constants used in the quantum circuit synthesis algorithm. Here $M = 6.1 \times 10^6$ for nitrogenase in the 54 orbital basis and $M = 8.2 \times 10^6$ for nitrogenase in the 57 orbital basis using the circuits of [11].

The true optima of Eq. (33) are difficult to find because of the factor of $\sqrt{\epsilon/\epsilon_2}$ in the logarithm. In order to simplify our optimization we instead choose ϵ_1, ϵ_2 and ϵ_3 to minimize

$$\tilde{C} = 2M \left(\frac{\alpha}{\epsilon_1} \right) \left(\beta \sqrt{\frac{\epsilon}{\epsilon_2}} \right) \left(\gamma \log_2 \left(\frac{2M}{\epsilon_3} \beta \right) + \delta \right), \qquad [34]$$

subject to the same constraint. The global optimum of Eq. (34) can be found directly from calculus, which allows near optimal parameters for Eq. (33) to be found easily. The value of Eq. (33) at these parameters is then an upper bound on the minimum of Eq. (33) and so the estimates provided remain upper bounds (modulo assumptions about the Trotter error).

The "worst" case assumptions in Table 2 correspond to only using rigorously proven upper bounds on the cost. These lead to estimates that are clearly extremely pessimistic. Even given our optimistic assumptions about the target computer, the worst case bounds suggest between millions and tens of thousands of years depending on whether parallelism is used.

The "pessimistic" assumptions use empirical scalings for circuit synthesis and phase estimation and use the worst scaling supported from our bounds on the Trotter error, but rescaled by the ratios observed between the actual Trotter numbers required and the theoretically predicted ones.

The "rescaled" case takes the rigorous upper bound for nitrogenase and divides it by the average ratio observed between the exact Trotter numbers and their upper bounds for the tractable molecules. Rescaling the upper bound by a constant and applying least squares fitting to find the most consistent constant yields similar results. We suspect these rescaled estimates may provide the most realistic estimate of the Trotter number required to simulate nitrogenase.

The "optimistic" assumptions again use the same empirical scalings for PE and synthesis, but instead extrapolate the average scaling observed for the ensemble of molecules whose Trotter numbers we can compute and scaling up the result so that all of the data lies beneath the curve. This scaling is optimistic, as there is little evidence for a clear trend in

Case	Gates	Time (100 MHz T gates)
Rigorous bound	1.0×10^{21}	3.2×10^5 years.
Clifford	1.4×10^{21}	–
Rigorous + PAR	3.2×10^{22}	8500 years.
Clifford	3.3×10^{22}	–
Pessimistic bound	7.9×10^{15}	30 months
Clifford	1.2×10^{16}	–
Pessimistic + PAR	2.3×10^{17}	31 days
Clifford	2.3×10^{17}	–
Rescaled bound	1.2×10^{15}	135 days
Clifford	1.8×10^{15}	–
Rescaled + PAR	3.5×10^{16}	120 hours
Clifford	3.5×10^{16}	–
Optimistic bound	1.6×10^{14}	19 days
Clifford	2.4×10^{14}	–
Optimistic + PAR	4.7×10^{15}	17 hours
Clifford	4.7×10^{15}	–

Case	α	β	γ	δ
Rigorous	8π	7×10^6	4	11
Pessimistic	$\pi/2$	1075	1.15	9.2
Rescaled	$\pi/2$	166	1.15	9.2
Optimistic	$\pi/2$	24	1.15	9.2

Table 2. Resource estimates for simulation of nitrogenase's FeMoco in structure 1 which requires a small basis consisting of 108 spin orbitals. PAR uses $\gamma = 4$ and $\delta = 11$.

Case	Gates	Time (100 MHz T gates)
Rigorous bound	1.8×10^{21}	5.5×10^5 years.
Clifford	2.5×10^{21}	–
Rigorous + PAR	5.9×10^{22}	1.5×10^4 years.
Clifford	6.0×10^{22}	–
Pessimistic bound	1.2×10^{16}	3.8 years
Clifford	1.8×10^{16}	–
Pessimistic + PAR	3.5×10^{17}	48 days
Clifford	3.5×10^{17}	–
Rescaled bound	2.2×10^{15}	250 days
Clifford	3.1×10^{15}	–
Rescaled + PAR	6.3×10^{16}	9 days
Clifford	6.3×10^{16}	–
Optimistic bound	2.3×10^{14}	27 days
Clifford	3.5×10^{14}	–
Optimistic + PAR	6.6×10^{15}	23 hours
Clifford	6.6×10^{15}	–

Case	α	β	γ	δ
Rigorous	8π	9.5×10^6	4	11
Pessimistic	$\pi/2$	1233	1.15	9.2
Rescaled	$\pi/2$	225	1.15	9.2
Optimistic	$\pi/2$	25	1.15	9.2

Table 3. Resource estimates for simulation of nitrogenase's FeMoco in structure 2 which requires a small basis consisting of 114 spin orbitals. PAR uses $\gamma = 4$ and $\delta = 11$.

the empirical data and the range provided is insufficient to meaningfully extrapolate out to 108 spin orbitals (54 spatial orbitals) or more. We provide this estimate because it provides the best scaling that could reasonably be claimed to be supported by the data.

Variance-based estimates. Such errors arise from three sources: the systematic error in the TS decomposition ϵ_{TS}, the statistical error tolerance for phase estimation ϵ_{QPE}, and the statistical error in synthesizing rotations from Clifford and T gates ϵ_{Rot}. We then require the total error to be $\epsilon_{\mathrm{TS}} + \sqrt{\epsilon_{\mathrm{QPE}}^2 + \epsilon_{\mathrm{Rot}}^2} = 0.1$ mHa. The three uncertainties are then chosen such that the number of T gates required for the simulation is minimized given the target accuracy.

The previous analysis for the estimates in the error can be used within this expression for the error under the assumptions that the errors in QPE and synthesis are not adversarial. This approach was taken with the estimates in the main body, wherein the three dominant costs are optimized against each other to minimize the resources needed to achieve the 0.1 mHartree target. The optimization process is exactly the same as that used to minimize the cost given in Eq. (33), however a different constraint linking the three errors is used. This leads to modest reductions in the costs relative to the worst case bounds, which we provide in Tables 2 and 3.

PAR circuits. There are several approaches that can be taken to parallelize rotations. The first, often coined nesting, is discussed in [34, 37]. It involves taking terms that commute with each other in the Hamiltonian and grouping them together so that they can be executed simultaneously. In principle, this can lead to substantial reductions in the depth but in practice it is difficult to assess the performance of these schemes here because of the size of the molecule and the fact that we have chosen to restrict ourselves to lexicographic ordering. This means that if we are to estimate the impact that parallelization can bring to these calculations we need to introduce a method that can reduce the T–depth without changing the ordering of terms in the Trotter–Suzuki decomposition.

The PAR method gives a way to achieve this goal [9]. It works by teleporting a rotation into a state with probability 1/2 using only Clifford operations and a pre–rotated ancilla. In the event that this method fails then instead of performing $R_z(\theta)$ it performs $R_z(-\theta)$. This can be corrected by teleporting a rotation $R_z(2\theta)$ into the qubit in question. Should this fail (and it will half the time) the rotation can be corrected by teleporting a rotation of $R_z(4\theta)$ and so on. This creates a geometric distribution of the number of pre–cached qubits needed to perform a given rotation. These rotation angles are known before hand and so can be prepared offline in parallel. Hence a logarithmic multiplicative overhead in space is needed to guarantee that enough ancilla qubits are prepared to perform the rotations with such high probability that it is unlikely that the cache of ancillae will ever be depleted.

In order to bound the number of ancillae needed to parallelize M rotations with high probability consider the following protocol.

- Divide the terms in the Trotter expansion of the Hamiltonian into blocks consisting of M sequential terms.
- For each rotation angle θ_j in a given block and each $j = 0, \ldots, n-1$, prepare the states $R_z(2^j\theta_j)|+\rangle$ in parallel for a predetermined value of n.
- Implement each of the M (possibly sequential operations) using programmable ancilla rotations using at most n attempts, if the PAR circuit fails in each attempt then a failure is said to have occurred.

- If a failure occurs then implement the correct rotation and proceed to the next precached rotation.

This protocol can be used to perform the desired rotation and its performance is summarized below.

Theorem 2. *There exists a protocol for implementing PAR rotations that caches M rotations of the form $R_z(\theta_j), R_z(2\theta_j), \ldots, R_z(2^{n-1}\theta_j)$ for $j = 1, \ldots, M$ such that the expected number of rotations performed before a failure occurs is*

$$2^n(1-(1-2^{-n})^M).$$

Proof. The proof is constructive. To see this consider the protocol discussed above. Such a protocol fails when all n PAR circuits fail for any of the M rotations cached, and a failure occurs when all n attempts at the rotation that have been precomputed are expended. The probability of such a failure is clearly 2^{-n} because the PAR circuit's success probabilities are independent and each attempt has success probability $1/2$ [9].

From the geometric distribution, the probability that no failure occurs in M trials is then simply $(1-2^{-n})^M$. Similarly, the probability that a failure occurs after precisely k attempts is $2^{-n}(1-2^{-n})^{k-1}$. Since there are only M rotations in the cache any branch that has more than M successes can only yield M rotations. This implies that the mean is

$$\sum_{k=1}^{M} k2^{-n}(1-2^{-n})^k + M(1-2^{-n})^M = 2^n(1-(1-2^{-n})^M). \quad [35]$$

□

A consequence of Theorem 2 is that a PAR cache of M rotations that further caches the correction operations for n failures can reduce the T–depth by a factor of $2^n(1-(1-2^{-n})^M)$. We can therefore adjust these parameters to substantially reduce the T–depth without adding a prohibitive number of rotations.

Note that as $n \to \infty$ Eq. (35) approaches M as expected. This suggests that taking large n allows the PAR rotations to be parallelized more efficiently, but this comes at the price of requiring more T gates and hence increases the overheads of quantum error correction.

Factory approach. A major challenge with costing PAR in a fault tolerant setting is that all the non–Clifford operations can be prepared simultaneously and offline. This means that if we take the cost model where only `T` gates are considered then we come to the absurd conclusion that the costs of all quantum simulation algorithms can be reduced to that of synthesizing a single rotation. In order to prevent such absurd tradeoffs we assume here that a delay of time equal to 1 `T` gate is included to model the measurement and feed-forward step that is needed for the programmable ancilla rotation. This fixed cost means that even if all of the `T`–gates are precached before hand then the time required for the remainder of the simulation will never be zero.

The above assumptions lead to an alternative approach to implementing PAR rotations, which we follow in the PAR costs in the following as well as the main body. Rather than constructing a large cache of rotations offline, it makes sense to produce the rotations just in time. Specifically if the cost of synthesizing a rotation is C `T` gates then it makes sense to have nC factories constantly producing rotation states of the form $R_z(2^j\theta_j)|+\rangle$ for $j = 1, \ldots, n$. Each of these nC factories is staggered such that one set of factories finishes with their n states at least 1 cycle before the next rotation is needed. As soon as a set of factories finishes it then proceeds on to the next set of rotations needed (excluding those that are currently being generated). If a failure occurs, then the factory approach halts just like the traditional approach and waits for a rotation to be synthesized that applies the correct rotation online.

Imagine for the moment that the PAR circuit succeeds in C consecutive attempts. Since each attempt requires time equivalent to a single invocation of a `T` gate and the cost of synthesis is C `T` gates, the first set of factories will have finished producing their states before the last set applies theirs. This means that with C such factories rotation states can be continuously generated even in the "worst case scenario" where each PAR circuit succeeds on the first attempt. This is why in this setting it does not make sense to use more than nC rotation factories given the assumption that the cost of each PAR attempt is 1 `T` gate and that the rotations for at most n failures are to be pre-cached.

Theorem 3. *The average time required per rotation to apply the factory–based PAR strategy, assuming each PAR application requires time at most equal to* 1 `T` *gate and all remaining Clifford operations are free, is*

$$\left(2-\frac{n+2}{2^n}\right)+\frac{C+n}{2^n}.$$

Proof. The probability of success in any PAR attempt is $1/2$ therefore the expected number of `T` gates that need to be applied before a solution is found is

$$\sum_{j=1}^{n}\frac{j}{2^n}+\frac{C+n}{2^n}. \quad [36]$$

The latter term gives the expected impact of a failure, which occurs with probability $1/2^n$, makes n PAR attempts and incurs a cost of C `T` gates in applying the fallback rotation. The result then follows from summing the geometric series. □

We use the above theorem to compute the expected time required by PAR under our assumptions of the costs of feed forwarding. If such costs are neglected then the online cost of performing a PAR rotation is simply $C/2^n$ `T` gates.

Nesting estimates. Finally, we would like to reiterate that in both the serial and PAR cases, the costs are found by building the circuit corresponding to the TS formula and counting the gates that compose it. The estimates for nesting given in the main body are upper bounds based on empirical estimates of the number of terms that can be simultaneously executed. We estimate this by taking FeMoco and greedily grouping terms that act on distinct sets of spin orbitals. We specifically find that there exists a grouping that can simultaneously execute 26.43 terms simultaneously for strucure 1 and 27.83 for structure 2. These values roughly correspond to the optimal scaling of $N/4$ that can be achieved through this nesting strategy. We then assume the quantum computer can simultaneously execute each of these commuting groups and find the corresponding time by dividing the `T` count by these factors.

Rather than executing the grouping in LIQ$Ui|\rangle$ we use a simple upper bound on the number of Clifford gates that

could arise from the grouping. We do this because the grouping strategy breaks the lexicographic ordering that leads to dramatic cancelation of the `CNOT` strings that arise from the Jordan–Wigner decomposition.

For the nested data, the number of timesteps needed is assumed to be the same as for the other cases, which is reasonable given that operator ordering tends to not have a dramatic impact on the error. Subsequent work will investigate the precise interplay that operator ordering has in nesting.

6. State Preparation

We discuss in this section the issue of state preparation, which is a major unanswered question that impacts the cost of quantum simulations. While we do not discuss these costs in detail in the rest of the paper, we discuss below the issues that arise when using elementary state preparation methods based on coupled cluster, configuration interaction or Hartree–Fock states as well as adiabatic state preparation. We also provide numerical results showing that the ground state overlap with Hartree–Fock states scales for small molecules and discuss the costs involved in adiabatic state preparation.

Elementary state preparation methods. Although we cannot rigorously prove that elementary state preparation methods such as Hartree–Fock states, unitary coupled cluster or truncated configuration interaction states (such as CISD or difference dedicated CI (DDCI) [38] states) will suffice for preparing a state with large overlap with the ground state, it is still important to ask how good simple ansatzes perform for numerically tractable cases. We provide some numerical evidence for small molecules showing that the overlaps of the true ground state with the Hartree–Fock ground state is not necessarily small. We leave similar studies of the overlap for unitary coupled cluster and truncated configuration interaction ansatzes for subsequent work.

We see in Figure 5 that there is substantial overlap between the Hartree–Fock state and the true ground state of the molecules calculated using LIQ$Ui|\rangle$. In particular, the smallest overlap that we see is 89%. Other studies that have looked at chains of hydrogen atoms that are near disassociation show very small overlaps with the Hartree–Fock states and propose methods to address such problems [39]. Thus there are small molecules that can be constructed that are not well described by such ansatzes.

We find that roughly 50% of the data points are well approximated by $107.75\% \times e^{-0.0076n}$ where n is the number of spin orbitals. This scaling would suggest that the overlap with the Hartree–Fock state for a molecule typical of this ensemble of the scale of FeMoco may be roughly 43%, we cannot say whether nitrogenase is indeed typical of this ensemble. Indeed, one may expect this to be a dramatic over-estimate for molecules that contain atoms with d–electrons because the resultant correlations are much greater than those examined in Figure 5 and multi-reference states are likely needed to achieve good over lap. Such molecules are frequently outside of our ability to simulate classically so finding an appropriate ensemble of molecules to use for such benchmarks remains an open research problem; however, the success of methods such as CI dynamically extended active space (CI-DEAS) [40, 41] in DMRG suggests that elementary truncated configuration interaction states may also suffice for quantum simulation [42].

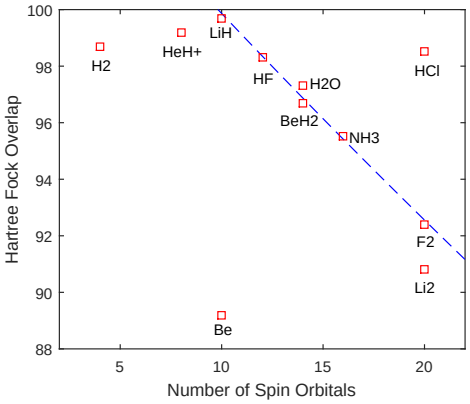

Fig. 5. Accuracy of Hartree Fock. Percent overlap ($|\langle\psi|\psi_{\rm HF}\rangle|^2 \times 100$) of the Hartree–Fock state with the electronic ground state computed by LIQ$Ui|\rangle$. All integrals are computed in a sto6g basis except for H_2 and HeH+. The integrals for those molecules are computed using sto3g and 3-21g bases respectively. The blue dashed line shows a possible extrapolated trend from the data.

We further see no compelling evidence for scaling with the maximum nuclear charge of the constituent atoms for the molecules in this set. This can clearly be seen from BeH_2 which has substantially better fidelity with the ground state than Be does. Similarly, HF and HCl have nearly identical overlaps despite the fact that Cl has nearly twice the nuclear charge of F. More study is needed in order to understand how these overlaps scale for strongly statically correlated molecules, however it is obvious that orbitals optimized for a multi-configurational wavefunction will be required to accurately model highly correlated ground states.

The choice of orbitals is crucial for accurate calculations. As orbital optimization requires the one- and two-body reduced density matrices, whose calculation will not be efficient on a quantum computer with present-day algorithms, it is decisive to start from suitable molecular orbitals which do not require further optimizations. This unfortunate situation our scheme shares with traditional approaches. (Restricted open-shell) Hartree-Fock orbitals are likely to be not the best choice, and hence one may exploit a small-CAS CASSCF calculation for the generation of suitable orbitals. We have observed [43] in DMRG calculations that such orbitals are well suited and yield results similar to fully optimized DMRG-SCF calculations. Whereas this result might not hold in general, it shows that suitable a priori orbital preparation is a way out of the self-consistent optimization of orbitals in a multi-configurational framework.

Adiabatic State Preparation. Adiabatic state preparation [44] provides an alternative approach state preparation wherein the Hamiltonian to be simulated is replaced with a time–dependent Hamiltonian whose final ground state coincides with the target ground state and whose initial ground state is an easily preparable state. An example of such a Hamiltonian is

$$H(s) = \sum_p h_{pp} a_p^\dagger a_p + \sum_{p,q} h_{pqqp} a_p^\dagger a_q^\dagger a_p a_q + s\left(H - \sum_p h_{pp} a_p^\dagger a_p - \sum_{p,q} h_{pqqp} a_p^\dagger a_q^\dagger a_p a_q\right), \quad [37]$$

where $s \in [0,1]$ is a dimensionless time that is 0 at the beginning of the evolution and 1 at the end. It is then easy to see that the ground state of $H(0)$ is the Hartree–Fock state and the ground state of $H(1)$ is the full configuration interaction (FCI) ground state.

For such Hamiltonians (or more generally for those that are differentiable at least three times and whose resulting derivatives are $O(1)$ [45]) a sufficiently slow evolution under the Hamiltonian will cause the Hartree–Fock state to be transformed into the FCI ground state of H. In other words if $\mathcal{T}$ is the time–ordering operator which is defined such that

$$\partial_s \mathcal{T}(e^{-i\int_0^s H(s')\mathrm{d}s'T}) := -iH(s)\mathcal{T}(e^{-i\int_0^s H(s')\mathrm{d}s'T}) \quad [38]$$

and if we define $\mathcal{P}$ to be a projector onto the FCI ground state and $\Delta(s)$ to be the smallest eigenvalue gap between the ground state and the rest of the spectrum of $H(s)$ then

$$|(\mathbf{I}-\mathcal{P})\mathcal{T}(e^{-i\int_0^1 H(s')\mathrm{d}s'T})|\psi_{\mathrm{HF}}\rangle| \in O\left(\frac{\max_s \|\dot{H}(s)\|}{\min_s \Delta^2(s)T}\right). \quad [39]$$

Here we take $T \gg 1$ and treat the other parameters to be bounded above by a constant in this asymptotic expansion. For such Hamiltonians the triangle inequality clearly shows that $\|\dot{H}(s)\| \in O(N^4)$ (although on physical grounds we expect $\|\dot{H}(s)\|$ to be in $O(\eta^2) \in O(N^2)$ because the potential energy scales quadratically with the number of constituent particles). Thus if we want to have $O(1)$ probability of preparing the FCI ground state it suffices to simulate the time–dependent Hamiltonian $H(s)$ for time

$$T \in O\left(\frac{N^4}{\min_s \Delta^2(s)}\right). \quad [40]$$

Prima facie, the best known bounds on the costs of Trotter–Suzuki based simulation give the circuit size for such a simulation to be [46, 47]

$$N_{\mathrm{operations}} \in \left(N^4\left[\frac{hN^8}{\min_s \Delta^2(s)}\right]^{1+o(1)}\right), \quad [41]$$

where $h \geq \max\{|h_{pq}|, |h_{pqrs}|\}$. If we take $h \in O(1)$, this rigorous bound suggests that the cost of adiabatic state preparation may be prohibitively expensive even if the eigenvalue gap is constant, it is important to note that this upper bound on the norm of the derivative of H is expected to be extremely loose and the scaling of the error in the Trotter–Suzuki formulas is expected to be much better than the scaling quoted above.

If we take the depth of second-order Trotter–Suzuki formula simulations of real molecules to scale as $O(N^{5.5})$, as observed in previous studies [36], and take the norm of the Hamiltonian to scale as $O(N^2)$ as anticipated asymptotically for a local basis [48] then the scaling that arises from using the second–order Trotter formula for time–dependent Hamiltonians [46] would be

$$O\left(\frac{h^{3/2}N^{8.5}}{\min_s \Delta^3(s)}\right). \quad [42]$$

If the effective value of h scales as $h \in O(N^{-2})$ then this scaling reduces to $N^{5.5}/\min_s \Delta^3(s)$. This scaling of the effective h (more accurately the root-mean-square value of $|h_{pqrs}|$ and $|h_{pq}|$ is emperically observed to scale as $O(N^{-2}$. Such scaling is expected if the molecule is at constant filling fraction, two body terms dominate the cost of the Trotter–Suzuki decomposition and $\|H\| \in \Theta(N^4 h)$. If we further assume that the Trotter-Suzuki error is dominated by the $pqrs$-terms and thus h_{pp}, h_{pq} etc can be neglected in this expression for the error. Since $\|H\| \in O(N^2)$ under these assumptions it follows that $h \in O(N^{-2})$. This means that even after making strong empirical assumptions, highly gapped adiabatic paths are likely to be necessary for adiabatic state preparation to be useful.

It is worth mentioning that adiabatic state preparation has been investigated for other systems and in these settings it has been found to be a highly practical method of state preparation [33, 49]. Although it should be noted that the paths used from the initial Hamiltonian to the final Hamiltonian are often non-trivial. These observations suggest that the above complexity analysis may be quite loose. Further work is needed to better estimate the cost of adiabatic state preparation for realistic molecules and also the cost of learning optimal adiabatic paths from easily preparable Hamiltonians to the FCI Hamiltonian.

7. Cost estimates for topological qubits

In this section we will examine the impact topological quantum computing may have on these numbers. This is important because the fault tolerant overheads quoted in the main body depend sensitively on the error rate. Topological quantum computing promises to provide physical error rates that are orders of magnitude beyond what is achievable in alternative platforms; however there is a catch. Much of the current research underway focuses on topological quantum computing using Ising anyons, which provide topological protection for Clifford operations but do not provide protection for `T` gates. This means that even if we achieve very low error rates using this technology then the costs of magic state distillation may not be reduced dramatically despite the quality of the Clifford gates that topological quantum computing affords. We examine this by modeling the errors in generating the `T` states to be 10^{-4} and then consider the costs of using the surface code to distill the necessary gates. We provide the data for this scenario in Table 4.

The data in Table 4 shows that even in this setting assuming low quality magic states does not remove the benefit of topological protection for the Clifford operations. In particular, if we look at the savings in physical qubits that occur from going from error rate 10^{-6} to 10^{-9} we see in the data in the main body that roughly an order of magnitude separates the two numbers of physical qubits. In contrast, if we assume the lower quality magic states given above then the reductions are more modest: they are roughly a factor of 5. This shows that while having high fidelity magic states is ideal, even if such states are low quality then a topological quantum computer with protected Clifford operations can nonetheless see advantages from having topologically protected gates at the level of accuracy required to simulate nitrogenase.

8. Further optimizations for quantum simulation

In many architectures, such as ion traps and existing superconducting qubits, the ability to perform `T` gates every 10 ns is beyond the capability of existing implementations. While some of these issues can be addressed by incorporating cold classical logic within the cryostat to reduce speed of light issues, it

	Serial rotations		PAR rotations		Nested rotations	
Clifford Error Rate	10^{-6}	10^{-9}	10^{-6}	10^{-9}	10^{-6}	10^{-9}
Required code distance	9,3	5,3	9,5	5,3	9,3	5,3
	Quantum processor					
Logical qubits	111		110		109	
Physical qubits per logical qubit	1013	313	1013	313	1013	313
Total physical qubits for processor	1.1×10^5	3.5×10^4	1.1×10^5	3.4×10^4	1.1×10^5	3.4×10^4
	Discrete Rotation factories					
Number	0		1872		26	
Physical qubits per factory	–	–	1013	313	1013	313
Total physical qubits for rotations	–	–	1.9×10^6	5.9×10^5	2.6×10^4	8.1×10^3
	T factories					
Number	64	30	41110	23248	1427	813
Physical qubits per factory	2.7×10^4	2.7×10^4	7.5×10^4	2.7×10^4	7.5×10^4	2.7×10^4
Total physical qubits for T factories	1.7×10^6	8.1×10^5	3.1×10^9	6.4×10^8	1.1×10^8	2.2×10^7
Total physical qubits	1.8×10^6	8.5×10^5	3.1×10^9	6.4×10^8	1.1×10^8	2.2×10^7

Table 4. Fault Tolerance Overheads. This table gives the resource requirements including error correction for simulations of nitrogenase's FeMoco in a 54 (spatial) orbital basis within the times quoted in Table I in the main body using physical gates operating at 100 MHz. Here we use error rates that are appropriate for quantum computing with Ising anyons, wherein topological protection is granted to Clifford operations but not to non–Clifford rotations. The error rate used in the production of the raw magic states is taken to be 10^{-4} in all of the above cases.

may be possible that when quantum computers emerge that they will be initially slower than optimistic speeds envisioned in the main body. In such cases, further optimizations to the quantum simulation algorithms may be useful to allow the simulations to be performed using a modest amount of resources.

Here we present several optimizations that can be used to reduce the depth of such quantum simulations and parallelize the phase estimation component over several quantum computers connected with either quantum or classical interconnects. These optimizations can reduce the time by many orders of magnitude but incur space/time tradeoffs that need to be addressed on a case by case basis within the limitations of the hardware.

Improved circuits for pqrs terms. The general two–body terms for the second quantized Hamiltonian involve, when expressed using the Jordan–Wigner transformation, can be written as tensor product of `X` and `Y` operators tensored with `Z` operators that appear due to the Jordan–Wigner strings. Conventional circuits, such as those in [13], break these terms in the Hamiltonian into a group of at most 8 commuting operators that act on the qubits in question. These circuits then diagonalize each term individually and perform the evolution based on this.

Here we take a different strategy. Rather than diagonalizing each term individually, we transform to the simultaneous eigenbasis of all 8 terms. This allows us to parallelize all 8 rotations while reducing the number of Clifford gates. The price is that this method requires ancillae to store the eigenvalues of all 8 terms. This approach can also be trivially applied to groups of only 2 or 4 commuting Hamiltonians but here we focus on the most complicated case of 8 terms.

We use the following convention to enumerate the Pauli operators that appear in the pqrs circuit.

$$P_s := \begin{cases} \mathtt{X} & s = 0 \\ \mathtt{Y} & s = 1 \end{cases}.$$

We will further ignore the `Z` operators that arise from the Jordan–Wigner because they are irrelevant here. Using this definition, we present a simultaneous eigenbasis for all such commuting operators in the following lemma.

Lemma 5. *Let $W = P_i \otimes P_j \otimes P_k \otimes P_\ell$ then the eigenstates of W, for any i, j, k, ℓ in $\mathbb{Z}_2$ such that $(i + j + k + \ell) = 0 \mod 2$, can be expressed as*

$$|a, v\rangle := \frac{|0\rangle|v_1 v_2 v_3\rangle + (-1)^a |1\rangle|\bar{v}_1 \bar{v}_2 \bar{v}_3\rangle}{\sqrt{2}}.$$

where $a \in \{0, 1\}$ and

$$W|a, v\rangle = (\sqrt{-1})^{i+j+k+l} (-1)^{a + v \cdot [j,k,l]} |a, v\rangle.$$

Proof.

$$\begin{aligned} W|0\rangle|v_1 v_2 v_3\rangle &= \sqrt{-1}^{i+j+k+l} (-1)^{j v_1 + k v_2 + \ell v_3} |1\rangle|\bar{v}_1 \bar{v}_2 \bar{v}_3\rangle. \\ W|1\rangle|\bar{v}_1 \bar{v}_2 \bar{v}_3\rangle &= \sqrt{-1}^{i+j+k+l} (-1)^{j \bar{v}_1 + k \bar{v}_2 + \ell \bar{v}_3 + i} |0\rangle|v_1 v_2 v_3\rangle. \\ &= \sqrt{-1}^{i+j+k+l} (-1)^{j v_1 + k v_2 + \ell v_3} |0\rangle|v_1 v_2 v_3\rangle, \end{aligned} \quad [43]$$

under our assumption that $i + j + k + \ell \mod 2 = 0$. Therefore the eigenvectors of W exist in the span of these two states as can be seen by

$$\begin{aligned} & W \left(\frac{|0\rangle|v_1 v_2 v_3\rangle + (-1)^a |1\rangle|\bar{v}_1 \bar{v}_2 \bar{v}_3\rangle}{\sqrt{2}} \right) \\ &= \sqrt{-1}^{i+j+k+l} (-1)^{j v_1 + k v_2 + \ell v_3 + a} \\ &\quad \times \left(\frac{|0\rangle|v_1 v_2 v_3\rangle + (-1)^a |1\rangle|\bar{v}_1 \bar{v}_2 \bar{v}_3\rangle}{\sqrt{2}} \right). \end{aligned} \quad [44]$$

The result then follows from $j v_1 + k v_2 + \ell v_3 = v \cdot [j, k, \ell]$, and observing that the eigenvectors corresponding to different a and $v = (v_1, v_2, v_3)$ are orthogonal. □

Now that we know the form of the eigenbasis of these terms we can find an operator that transforms between the

computational basis and the eigenbasis of W. We can then discuss a diagonalizing transformation for the commuting Hamiltonians.

Definition 1. *Let U_W be the circuit consisting of one Hadamard gate on the first qubit followed by $N-1$ CNOT opperations that maps*

$$|0\rangle^N \mapsto \frac{|0\rangle^N + |1\rangle^N}{\sqrt{2}}.$$

Corollary 1. *Let $W = P_i \otimes P_j \otimes P_k \otimes P_\ell$ for i,j,k,ℓ in $\mathbb{Z}_2$ then*

$$\sqrt{-1}^{i+j+k+l} U_W \left(\mathtt{Z}\otimes \mathtt{Z}^j \otimes \mathtt{Z}^k \otimes \mathtt{Z}^\ell\right) U_W^\dagger = W.$$

Proof. We prove the theorem by showing that the two operators have equivalent actions on the eigenstates of W. First it is easy to see by applying the inverse of U_W that

$$U_W^\dagger |a,v\rangle = |a\rangle|v_1v_2v_3\rangle. \quad [45]$$

$$\left(\mathtt{Z}\otimes \mathtt{Z}^j \otimes \mathtt{Z}^k \otimes \mathtt{Z}^\ell\right) U_W^\dagger |a,v\rangle = (-1)^{a+v\cdot[j,k,\ell]}|a\rangle|v_1v_2v_3\rangle. \quad [46]$$

Hence

$$U_W \left(\mathtt{Z}\otimes \mathtt{Z}^j \otimes \mathtt{Z}^k \otimes \mathtt{Z}^\ell\right) U_W^\dagger |a,v\rangle = (-1)^{a+v\cdot[j,k,\ell]}|a,v\rangle. \quad [47]$$

Thus every eigenvector of W is also an eigenvector of $\sqrt{-1}^{i+j+k+l} U_W \left(\mathtt{Z}\otimes \mathtt{Z}^j \otimes \mathtt{Z}^k \otimes \mathtt{Z}^\ell\right) U_W^\dagger$ with the same eigenvalue. Therefore both operators are equivalent. □

Definition 2. *For $\{a_0,\ldots,a_7\} \subset \mathbb{R}$ define*

$$\begin{aligned} H_8 := \quad & a_0 \mathtt{X}\otimes\mathtt{X}\otimes\mathtt{X}\otimes\mathtt{X} + a_1 \mathtt{Y}\otimes\mathtt{Y}\otimes\mathtt{X}\otimes\mathtt{X} \\ & + a_2 \mathtt{Y}\otimes\mathtt{X}\otimes\mathtt{Y}\otimes\mathtt{X} + a_3 \mathtt{Y}\otimes\mathtt{X}\otimes\mathtt{X}\otimes\mathtt{Y} \\ & + a_4 \mathtt{X}\otimes\mathtt{Y}\otimes\mathtt{Y}\otimes\mathtt{X} + a_5 \mathtt{X}\otimes\mathtt{Y}\otimes\mathtt{X}\otimes\mathtt{Y} \\ & + a_6 \mathtt{X}\otimes\mathtt{X}\otimes\mathtt{Y}\otimes\mathtt{Y} + a_7 \mathtt{Y}\otimes\mathtt{Y}\otimes\mathtt{Y}\otimes\mathtt{Y}. \end{aligned}$$

Theorem 4. *Given $\{a_0,\ldots,a_7\} \in \mathbb{R}$, the circuit of Figure 6 simulates $e^{-iH_8 t}$ exactly for all $t \in \mathbb{R}$.*

Proof. Corollary 1 implies, after noting that $i+j+k+\ell = 0, 2, 4$ for terms with 0,2 and 4 Y operators respectively, that

$$\begin{aligned} e^{-iH_8 t} &= e^{-i(a_0 \mathtt{XXXX} + \cdots + a_7 YYYY)t} \\ &= e^{-i(a_0 U_W(\mathtt{Z}\otimes\mathtt{I}\otimes\mathtt{I}\otimes\mathtt{I})U_W^\dagger - \cdots + a_7 U_W(\mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z})U_W^\dagger)t)} \\ &= U_W e^{-i(a_0 \mathtt{Z}\otimes\mathtt{I}\otimes\mathtt{I}\otimes\mathtt{I} - \cdots + a_7 \mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z})t} U_W^\dagger. \\ &= U_W \left(e^{-ia_0 \mathtt{Z}\otimes\mathtt{I}\otimes\mathtt{I}\otimes\mathtt{I} t} \cdots e^{-ia_7 \mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z}t}\right) U_W^\dagger. \end{aligned} \quad [48]$$

Thus the problem of simulating the unitary evolution reduces to the problem of sequentially simulating evolution under the above 8 terms after diagonalizing them using $U_W^\dagger$. It is clear that the first 5 steps of the circuit in Figure 6 performs $U_W^\dagger$ on the first four qubits and then prepares the following state in the quantum computer

$$\begin{aligned} |a,v\rangle|0\rangle^4 \mapsto & |a\rangle|a+v_1\rangle|a+v_2\rangle|a+v_3\rangle|a+v_1+v_2\rangle \\ & \times |a+v_1+v_3\rangle|a+v_2+v_3\rangle|a+v_1+v_2+v_3\rangle. \end{aligned} \quad [49]$$

Therefore the next step of our circuit implements

$$\begin{aligned} |a,v\rangle|0\rangle^4 \mapsto & e^{-ia_0 \mathtt{Z}t}|a\rangle e^{ia_1 \mathtt{Z}t}|a+v_1\rangle e^{ia_2 \mathtt{Z}t}|a+v_2\rangle e^{ia_3 \mathtt{Z}t} \\ & \times |a+v_3\rangle e^{ia_4 \mathtt{Z}t}|a+v_1+v_2\rangle e^{ia_5 \mathtt{Z}t}|a+v_1+v_3\rangle \\ & \times e^{ia_6 \mathtt{Z}t}|a+v_2+v_3\rangle e^{-ia_7 \mathtt{Z}t}|a+v_1+v_2+v_3\rangle, \end{aligned} \quad [50]$$

which after resetting the parity bits to 0 is equivalent to

$$\left(e^{-ia_0 \mathtt{Z}\otimes\mathtt{I}\otimes\mathtt{I}\otimes\mathtt{I} t} \cdots e^{-ia_7 \mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z}\otimes\mathtt{Z}t}\right) |a\rangle|v_1v_2v_3\rangle\left|0^4\right\rangle. \quad [51]$$

The remaining circuit serves to transform the ancillae back to their initial state and hence the circuit simulates $e^{-iH_8 t}$ as claimed. □

In principle, this optimization can reduce the depth by a factor of 8 at the price of requiring 4 extra qubits. In practice, however, few terms consist of all 8 rotations. For nitrogenase, these circuits offer roughly a factor of 4 reduction in the gate depth. Further optimizations along these lines are possible and their application could allow many more commuting terms to be applied simultaneously than current nesting strategies permit.

Parallel phase estimation. A major driver of the cost of our algorithm is the use of phase estimation. Its use dramatically reduces the T count at the price of increased depth. Fortunately, there are strategies for parallelizing phase estimation. The first such strategy that we will consider is to network a number of quantum computers together by sharing GHZ states and then using these states to accelerate the phase estimation algorithm in concert with adiabatic state preparation.

The basis behind our technique for parallel phase estimation was invented by Knill, Ortiz and Somma [50]. The idea behind the method is the observation that if we have an eigenstate $|\lambda\rangle$ then

$$e^{-iHt}|\lambda\rangle|\lambda\rangle = e^{-iHt/2}|\lambda\rangle e^{-iHt/2}|\lambda\rangle. \quad [52]$$

Thus if we have r copies of an eigenstate then we can simulate the whole evolution using only evolutions of duration t/r.

By distributing a GHZ state over r quantum computers connected by quantum channels, we can implement this phase estimation protocol in $1/r$ the time. Entanglement distillation and teleportation can be used to make such a protocol scalable in the presence of noise in the limit of large r. Once such states have been distributed the phase estimation procedure takes the form of

$$\frac{|0\rangle^{\otimes r} + |1\rangle^{\otimes r}}{\sqrt{2}} |\lambda\rangle^{\otimes r} \rightarrow \frac{|0\rangle^{\otimes r}|\lambda\rangle^r + |1\rangle^{\otimes r}(e^{-iHt}|\lambda\rangle)^{\otimes r}}{\sqrt{2}}. \quad [53]$$

The two drawbacks of the Knill, Ortiz and Somma approach are that substantially more qubits are required to parallelize the phase estimation and that high-fidelity eigenstates are needed. This can be seen in the following lemma.

Lemma 6. *Assume $H|\lambda\rangle = \lambda|\lambda\rangle$ and let $\left|\tilde{\lambda}\right\rangle = \sqrt{1-\epsilon^2}|\lambda\rangle + \epsilon\left|\lambda^\perp\right\rangle$ where $\langle\lambda|\lambda^\perp\rangle = 0$ and $\epsilon \geq 0$. Then assuming a non–deterministic phase estimation algorithm that has non–zero success probability is applied to estimate the eigenvalues of e^{-iHt} for $t > 0$ and H a Hermitian matrix in $\mathbb{C}^{M\times M}$ using the state $\left|\tilde{\lambda}\right\rangle^{\otimes r}$, the expected error in the estimate of the eigenvalue*

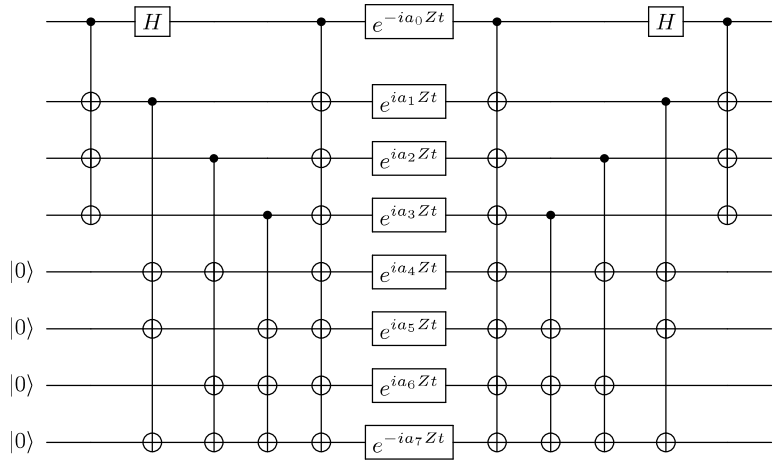

Fig. 6. Circuit for simulating e^{-iH_8t}. The four ancillae are used to store eigenvalues of H_8. The effects of the Jordan–Wigner strings can be included by simply applying the ladder of controlled-not gates to the top most qubit both before the first set of 7 controlled not operations and after the second set.

of $e^{-iH^{\otimes r}t}$ (post selected on success of the phase estimation algorithm) is at most $\delta > 0$ if

$$\epsilon \leq \frac{1}{r}\sqrt{\frac{\delta}{2\|H\|t}},$$

where $\|\cdot\|$ is the spectral norm.

Proof. First because the error probability is non–zero, we can meaningfully discuss the mean error conditioned upon the phase estimation algorithm succeeding. It is then straight forward to see that the inner product squared between the two states is $(1-\epsilon^2)^r \geq 1 - r\epsilon^2$. Thus the probability that phase estimation on $e^{-iH^{\otimes r}t}$ failing to provide the correct eigenvalue is at most $r\epsilon^2$. If the algorithm fails then it follows from the definition of the spectral norm that the largest error possible in the estimate of the eigenvalue is $2r\|H\|t$. Thus the expected error is at most $2r^2\epsilon^2\|H\|t$. If we set this error to be $\delta > 0$ then it suffices to pick $\epsilon \leq \frac{1}{r}\sqrt{\frac{\delta}{2\|H\|t}}$ to ensure that the post-selected mean error is at most δ. □

Thus as r increases, we should aim to have the error in state preparation must shrink at least linearly with the number of parallel copies of the state. The approximation error is typically considered to be fixed for many state preparation methods, such as Hartree Fock approximations or CCSD(T) states. This naturally leads to a problem that is known in chemistry as the Van Vleck catastrophe.

While the Van Vleck catastrophe is indeed catastrophic for fixed approximations, it is not necessarily for adiabatic state preparation. Adiabatic state preparation requires time that, under appropriate assumptions of continuity on the Hamiltonian [45, 51], scales as $T \in \tilde{O}(\max\|\dot{H}(s)\|/(\min gap^2\epsilon))$, where $H(s)$ interpolates from an elementary Hamiltonian with an easily preparable groundstate at $s = 0$ to the FCI Hamiltonian at $s = 1$. Thus if we take $\delta \in \Theta(\|H\|t)$ it follows that the time required for adiabatic state preparation is

$$T \in \tilde{O}\left(\frac{r\max\|\dot{H}(s)\|}{\min gap^2(s)}\right). \quad [54]$$

Thus even if a well gapped adiabatic path exists between the initial and final Hamiltonians, the time required for adiabatic state preparation scales at most linearly with r. Further, since the cost of quantum simulation scales at least linearly with T it follows that the depth reduction of a factor of r from parallelizing does not necessarily compensate for increased cost of adiabatic state preparation without further assumptions.

However, since we are primarily interested in high–accuracy state preparation much better strategies exist [52, 53]. If such approaches, known as boundary cancellation methods, are employed then the cost is dramatically reduced as we show below.

Theorem 5. *Assume that*

- *$H(s) : \mathbb{C} \mapsto \mathbb{C}^{M\times M}$ is analytic in a strip of width $\gamma \in \Theta(1)$ about the real-line,*
- *For all $q \in [1, \ldots, \infty)$ $\|H^{(q)}(s)\| \in O(1)$,*
- *for any positive integer q we can choose the above map such that, $H^{(p)}(0) = H^{(p)}(1) = 0$ for all $p \in \{1, \ldots, q\}$,*
- *for all $s \in [0, 1]$ the spectral gap between the groundstate of $H(s)$ and any other eigenstate is in $\Omega(1)$,*
- *all assumptions of Lemma 6 are satisfied.*

It then follows that the time required to perform adiabatic state preparation such that, post selected on success of the phase estimation algorithm, the mean error in the estimate of λ is at most δ for evolution time

$$T \in \Theta(\log(r\sqrt{2\|H\|t/\delta})).$$

Proof. Under our assumptions it follows that our time-dependent Hamiltonian is a one-parameter family of bounded Hamiltonians that has some fixed distance, $\gamma > 0$, such that any pole or branch point is distance at least γ from the real-axis. Thus we can apply Corollary 1 of [52], which states that for every $\epsilon > 0$ there exists an interpolation of the adiabatic path such that the error in the adiabatic approximation for time T (under our assumptions on the derivatives of $H(s)$) is at most

$$\epsilon \in T^{\Theta(1)}e^{-\Theta(T)} \subseteq e^{-\Theta(T)}, \quad [55]$$

under our assumptions on the gap, the derivatives of H and the constant γ. This implies that

$$T \in \Theta(\log(1/\epsilon)). \quad [56]$$

Then from Lemma 6 we see that the post selected mean error in phase estimation is at most δ for $T \in \Theta(\log(r\sqrt{2\|H\|t/\delta}))$. □

Since the cost of simulation scales at least linearly with time (and super-linearly for Trotter-Suzuki formulas [54]) it follows that the cost of adiabatic state preparation scales at most poly-logarithmically with r, which itself is $O(1/\sqrt{\delta})$ from previous discussions. This means that under the circumstances that a gapped adiabatic path exists that satisfies the above properties, the cost of preparing the state within sufficient error will not typically dominate the cost of phase estimation if each of the r Trotter steps used in the longest step in the phase estimation algorithm are parallelized over. This shows that quantum methods can be used to inexpensively parallelize the phase estimation algorithm.

If we apply this strategy to nitrogenase we can potentially reduce the depth by a factor of roughly 100 from Figure 4. However, this is predicated on the existence of a gapped adiabatic path connecting an elementary Hamiltonian (such as the second-quantized Hartree Fock Hamiltonian) and the FCI groundstate. Proving such a path exists is typically as hard, or harder, than the problem of groundstate estimation. Thus we cannot know how difficult adiabatic state preparation is for nitrogenase. So although this method is promising, as with other state preparation methods, its practicality may need to be assessed using quantum hardware. Regardless, this demonstrates that the depth of phase estimation can be substantially reduced given a rudimentary quantum network capable of distributing the qubits in a GHZ state over a cluster of quantum computers.

Phase estimation using clusters of quantum computers. The inference process used in iterative phase estimation algorithms provides a natural breakpoint to insert classical parrallelism. In particular, we can use the results from several quantum computers to infer the most likely eigenphase for a system. In this section we provide a concrete method for this based on approximate Bayesian inference. Although efficient schemes exist for approximate Bayesian inference, we do not use them here because chemical accuracy provides a fixed target precision for PE. This means that inefficient methods can be used provided that they need reasonable time.

Bayesian phase estimation works by positing a prior distribution that represents the user's subjective beliefs about the probabilities that certain hypotheses about the eigenphases are true. The next step involves performing one or more independent experiments, $\{E_1, E_2, \ldots, E_K\}$ which then inform these beliefs via Bayes' rule:

$$P(\phi|E_1, \ldots, E_k) = \frac{\left(\prod_j P(E_j|\phi)\right) P(\phi)}{\int \left(\prod_j P(E_j|\phi)\right) P(\phi) \mathrm{d}\phi}. \quad [57]$$

The probability density $P(\phi|E_1, \ldots, E_k)$ is known as the posterior distribution. The function $P(E_j|\phi)$ is known as the likelihood function, which is given by Lemma 2 for phase estimation. The probability density $P(\phi)$ is known as the prior distribution. In discrete cases the probability densities become probabilities and the normalizing integral in the denominator for Bayes' rule becomes a sum.

Unfortunately, exact Bayesian inference is typically intractable (cases where conjugate priors exist are a notable exception). Approximate methods therefore typically have to be employed. The method that we employ is reminiscent of information theory phase estimation (ITPE) [55], which divides the hypothesis space for the eigenphase into a fixed number of bins and then infers the most likely bit value based on the MLE estimate of the posterior probability of the eigenvalue. Here we modify this framework by replacing the maximum likelihood estimator by the posterior mean and use an adaptive heuristic to estimate the best times for each round of the experiment. The procedure we use, which is based on [28], is given below.

1. Divide the hypothesis space $[0, 2\pi)$ for the eigenphase into N discrete bins and select a target uncertainty, $\epsilon > 0$ such that $\epsilon > 2\pi/N$.

2. Choose $P(\phi)$ to be the uniform distribution over these discrete eigenphases.

3. For each of the C quantum computers in the cluster, choose the parameter $\theta_c : c = 1, \ldots, C$ by sampling C eigenphases from the prior distribution and choose $M = 1/(S\sqrt{C})$ where S is the circular standard deviation of the prior probability distribution.

4. Perform each experiment in parallel on a quantum computer.

5. Use Bayes' rule to update the prior probability distribution based on the C experiments.

6. Repeat until $S \leq \epsilon$ and return S and the posterior mean μ, which is our estimate of the eigenphase.

The guess heuristic used in the above discussion is known as the particle guess heuristic [28], which is known to be nearly optimal for non–parallelized phase estimation algorithms. We modify this heuristic to scale like $1/(S\sqrt{C})$ rather than $1/S$ which we justify based on the fact that the error in parallel phase estimation scales as $O(1/\sqrt{C})$ for a uniform prior distribution [27]. In principle, locally optimized methods can also be used for this such as those considered in [30]. We therefore choose less informative experiments at each step and compensate for this by performing many more of them.

We see from Figure 7 that the evolution time needed to perform phase estimation within chemical precision shrinks as more quantum computers are added to the cluster. We observe from the data that we see clear evidence that the mean time required to achieve an uncertainty in the phase of ϵ scales as $\pi/(2\epsilon\sqrt{C})$ (where $\epsilon = 10^{-4}$ Ha, 10^{-3} Ha) asymptotic scaling of the evolution time required to achieve a fixed uncertainty target. This scaling is asymptotically optimal [27] (because the Holevo variance coincides with the circular variance for narrow distributions), and as such we do not expect that an unbiased estimator of the eigenphase will be able to substantially outperform this without either the use of prior information or additional quantum resources at these fixed accuracy targets.

Note that while the mean uncertainty is small for $C \geq 3$, it is not small for $C = 1, 2$. The median data in Figure 8 shows that the algorithm still works well, with at least 50%

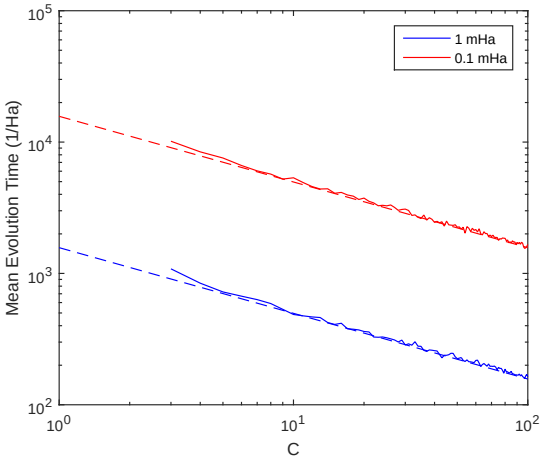

Fig. 7. The mean total time required (proportional to circuit depth) for PE in a cluster of C parallel quantum computers to estimate the phase within an uncertainty of $\epsilon = 1, 0.1$ mHa. Dashed lines correspond to $\pi/(2\epsilon\sqrt{C})$.

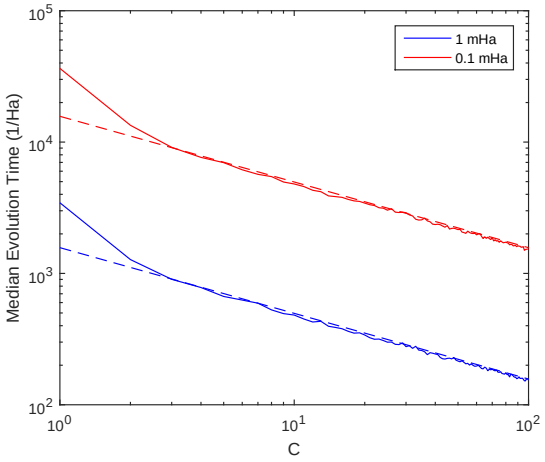

Fig. 8. The median total time required (proportional to circuit depth) for PE in a cluster of C parallel quantum computers to estimate the phase within an uncertainty of $\epsilon = 1, 0.1$ mHa. Dashed lines correspond to $\pi/(2\epsilon\sqrt{C})$.

probability for small C but suffers from rare failures. Also it is worth noting that small values of C do not allow optimal scaling because the guess heuristic is not always providing the best possible experiment. This tendancy dissapears as C increases because the algorithm can then explore a wider range of experimental parameters to compensate for guesses that are individually poor.

A crucial assumption in the above discussion is that an exact copy of the eigenstate is provided. If one is not provided then several options remain. The first is to use adiabatic state preparation to make increasingly accurate estimates of the groundstate, as discussed previously. The second option involves using lower bounds on the eigenvalue gap, $gap > 0$, and limitted precision phase estimation to project onto the groundstate. We will discuss this option below.

If the gap is sufficiently large, then the phase estimation algorithm can be run independently to learn the individual eigenstates stored in each of the C quantum computers. Since we know that the eigenvalue gap is g, this can be done in theory by learning each eigenvalue within error $\pm g$ and rejecting the states that do not fall within this range. In practice, this can be done by fixing a confidence level η and ensuring that the posterior distribution for ϕ assigns probability $1-\eta$ in a region of $(E_0 - g, E_0 + g)$ where E_0 is either an empirical or analytic approximation to the groundstate energy. This typically requires time on the order of $O(\log(1/\eta)/g)$ to attain [1]. Since $\eta \in O(1/C)$ in order to make sure that the probability that one of the quantum computers contains an erroneous eigenphase is small, the process requires time that scales as $O(\log(C)/g)$. Thus this approach can asymptotically provide an advantage in depth over traditional phase estimation provided $C \ll \exp(g/\Delta)$ where Δ is chemical accuracy.

One issue that remains is that in the worst case scenario this process requires $O(\sqrt{C})$ times as many quantum operations performed over *all the quantum computers* as traditional phase estimation would require. Thus the logical error rates must be reduced by a factor of $\sqrt{C}$ in order to ensure that the probability of failure remains small over the entire set. However, this is unlikely to result in substantially greater overheads for the computation owing to the fact that the overheads from error correction are at worst polylogarithmic.

Phase estimation can be realistically distributed over a small cluster of quantum computers. Roughly 100 quantum computers computing on the same eigenstate suffice to reduce the error in phase estimation by a factor of 10, which suggests that this approach can potentially be used to allow nitrogenase to be simulated on slower architectures with gate times on the order of 0.1μs as opposed to the 10ns assumed in previous discussions without increasing the run time or requiring a quantum network.

9. FeMoco — the active site of nitrogenase

In this section we provide more background on the importance of biological dinitrogen fixation and on the active site model of nitrogenase prepared in different charge and spin states applied in the feasibility analysis of this work.

For decades, a Holy Grail in chemistry has been the catalytic fixation of molecular nitrogen under ambient conditions. Less than half a dozen synthetic catalysts have been developed for this purpose [56–58] (after decades of fruitless efforts). All of them suffer from low turnover numbers and the synthetic dinitrogen-fixation problem under ambient conditions can thus be considered largely unsolved. The process is of tremendous importance for society as fertilizers are produced from ammonia, the final product of dinitrogen fixation. Industrially, ammonia is produced in the very efficient Haber–Bosch process, which, however, requires high temperature and pressure (and consumes up to 2% of the annual energy production) [59].

While it currently appears unrealistic that this simple heterogeneous process will be replaced by a sophisticated synthetic homogeneous catalyst, which is likely to be less stable and expensive to produce, a mono- or poly-nuclear iron-based catalyst working under ambient conditions and feeding on an easily accessible source for 'hydrogen' (and dinitrogen from air) could become important for local small-scale fertilizer-production concepts.

In any case, nitrogen fixation represents a tremendous chemical challenge to activate and break the strong triple bond in dinitrogen at room temperature and pressure. Nature found

an efficient way to achieve this goal. It is accomplished by the enzyme nitrogenase whose active site, the iron–molybdenum cofactor FeMoco, consists of seven iron atoms and one molybdenum atom, which are clamped together by bridging sulfur atoms [60]. The complete structure of the FeMoco was solved only very recently when a central main-group atom was discovered [61] that, surprisingly, turned out to be a carbon atom [62–64]. The complex electronic structure of this cluster of open-shell iron atoms, the possible charge, spin, and protonation states as well as the different ligand binding sites to be considered makes this active site a nightmare for electronic structure calculations, which is the basis of all theoretical approaches toward the elucidation of the mode of action of metalloenzymes such as nitrogenase.

It is thus no surprise that the specific mechanism of dinitrogen reduction at this active site has been elusive, especially given the fact that the mechanism of nitrogenase is difficult to study experimentally. Computational approaches suffer from the static electron correlation problem. For two ammonia molecules to be produced, the transfer of six protons and six electrons is required per dinitrogen molecule (in fact, eight protons and eight electrons are needed as one dihydrogen molecule is produced stoichiometrically). The transfer of these highly reactive agents leads to many stable intermediates and side products (see, for example, the analogous discussion of these steps in Ref. [65] for the first synthetic dinitrogen-fixating complex by Yandulov and Schrock). To elucidate the mechanism of nitrogenase, which is important for a better understanding of the activation of inert bonds by synthetic catalysts, therefore requires the consideration of many molecular structures.

While molecular structure of stable intermediates and transition states may be optimized within unrestricted Kohn–Sham DFT, the calculation of their energies demands an accurate wave-function-based approach. We therefore optimize molecular structures of a FeMoco model in the resting structure (Fig. 1 (right) in the main article) for varying charge and spin states in order to create different electronic situations that challenge the feasibility analysis presented in this work. For these structures, integrals in a molecular orbital basis have been obtained that parametrize the electronic structure of the cluster in the second-quantized quantum-chemical Hamiltonian of Eq. (6).

10. Exact diagonalization techniques in chemistry

The electronic structure of a molecular structure determines its reactivity. Predicting chemical reactions requires the solution of the electronic Schrödinger equation to obtain the electronic energy and wave function. Whereas the expansion of the many-electron wave function into a (quasi-) complete many-electron basis will produce the exact solution, called full configuration interaction in chemistry or exact diagonalization in physics, this approach is unfeasible for molecules of more than a few atoms. As all standard quantum-chemical solution approaches construct many-electron basis functions from orbitals, the number of the former is determined by the number of the latter. Exact diagonalization is therefore limited to about 18 electrons distributed among 18 spatial orbitals due to the exponential scaling of the many-electron basis states with the number of orbitals [66]. Unfortunately, the size of an orbital basis is already very large for moderately sized molecules. As a consequence, a restricted orbital space must be chosen.

Of all approximations developed in quantum chemistry to overcome this problem [67] the complete-active-space self-consistent-field (CASSCF) approach (and related models) utilizes exact diagonalization, but, because of the exponential scaling, in a reduced orbital space, the so-called CAS, that selects orbitals around the Fermi energy. To compensate for this approximation, the orbitals are relaxed self consistently, hence the name. Still, the CAS is limited by the 18-orbital wall and by the neglect of most of the (virtual) orbitals for the construction of the wave function. Considering the fact that molecules of a hundred atoms or more quickly require much more than a thousand one-electron basis functions for an accurate description of their electronic structure, most of the orbitals constructed from these basis functions are omitted in the construction of a CASSCF wave function. Even iterative techniques such as the density matrix renormalization group (DMRG), which can be understood as a polynomially scaling CASSCF approach, can push this wall only to about a hundred spatial orbitals.

As a result, a CASSCF-type wave function solves only the so-called static electron correlation problem. It is therefore particularly well suited for molecular structures with near-degenerate orbitals. The resulting electronic structure is only qualitatively well described. However, this feature is maintained throughout a reaction coordinate, which makes a CASSCF-type approach an appealing universal approach. For a quantitative description, the many virtual orbitals not considered for the CAS make a nonnegligible contribution to electronic-energy differences. To account for this so-called dynamic correlation is mandatory and typically achieved by a subsequent multi-reference perturbation theory calculation. Such perturbation theories to second order require elements of the three- and four-electron reduced density matrices, which are difficult to calculate and to store.

Instead of this 'diagonalize, then perturb' approach, also 'perturb, then diagonalize' ideas have been studied in chemistry. These latter approaches produce a 'dressed' many-electron Hamiltonian which is then better conditioned for a CASSCF-type approach. A new development in this area is the combination of density functional theory (DFT), known to describe dynamic electron correlation well with a CASSCF-type approach by spatial range separation introduced by Savin (see Refs. [68, 69] and references cited therein). Range separation is accomplished for the electron–electron interaction matrix elements and allows one to apply DFT at short range, whereas at long range the full flexibility of a CASSCF-type wave function can be exploited. This approach is very efficient and does not compromise the efficiency of a CASSCF-type approach (by contrast to perturbation theory). In combination with the polynomially scaling DMRG, this approach has delivered very promising results [69] for benchmark reactions involving transition metals [70].

As a consequence, a chemically sensible exact diagonalization technique is a CASSCF-type approach that captures dynamic correlation effects in the one-electron states. Such an approach could be directly implemented on a quantum computer with the specific exact diagonalization technology developed for such a machine.

Thermochemical data may be obtained in our framework by temperature and entropy corrections added to the electronic energy differences. These corrections may sufficiently reliably be calculated by standard (unrestricted) DFT. Their accuracy

may be enhanced by standard scaling factors that may even be subjected to rigorous Bayesian error estimation (for a detailed discussion see Ref. [71]). Such a patchwork approach is acceptable considering the fact that DFT geometry optimization will be the source of all molecular structures subjected to our quantum-computing framework. This is also the case for traditional methods as massive routine structure optimization with correlated ab initio approaches will not be possible in the near future and, at the same time, DFT structures are known to be surprisingly accurate, also in cases when the energy assignment to these structures is not reliable.

Considering dynamical electron correlation in a perturb-then-diagonalize approach may require the calculation of the one-body reduced density matrix (or its trace, the electron density) as it is the case for a combination with short-range DFT in a range-separated ansatz. This would then require a costly evaluation of the density within the quantum computing framework that may be circumvented by robust approximations. One may exploit model densities obtained by small-CAS CASSCF or DMRG-SCF calculations, by unrestricted density functional calculations with spin projection, or even by Hartree-Fock calculations, which has already proved reliable in traditional CASCI-srDFT implementations. Other perturb-then-diagonalize approaches such as PDFT may avoid this altogether by introducing an on-top pair density a posteriori to the multi-configurational calculation.

What chemical problems will benefit from such an implementation? Clearly, these will be problems that are dominated by strong static electron correlation (rather than by weak dispersion interactions originating from dynamic electron correlation). The electronic Schrödinger equation assigns an energy to a given molecular structure in the Born–Oppenheimer approximation, and this electronic energy (evaluated at zero Kelvin without vibrational, temperature, and entropy corrections) should dominate the energy change of a chemical process. Transition-metal catalysis is a field that presents such situations.

Many important chemical transformations are mediated by complicated electronic structures featured by transition-metal complexes. Especially late $3d$ transition metals are of this kind, most importantly iron, which is cheap and ubiquitous, often yields nontoxic compounds, and therefore represents an ideal catalytic center. Moreover, stable intermediates and, in particular, transition states of a reaction mechanism often represent typical static electron correlation problems as bonds are formed and broken on the way to stable products. To reliably predict a chemical transformation of this kind usually requires to study many more than one elementary reaction step — especially when reactive intermediates are involved. The number of stable intermediates also increases due to unwanted side reactions that need to be inspected. Therefore, the total number of molecular structures whose electronic energy is required for an understanding of a reaction mechanism is, in general, very large.

Clearly, all these structures must be optimized with an efficient quantum chemical method. While the accuracy of electronic energies obtained with present-day DFT approaches is often not satisfactory for predictive purposes (see, e.g., Refs. [70, 72]), molecular structures can be reproduced with remarkable accuracy (obviously, a geometry gradient would replace DFT structures, too). Hence, CASSCF-type methods will mostly be required for the validation of electronic energies of DFT-optimized molecular structures, which is the essential piece of information for establishing a reaction mechanism.

11. Computational Methodology

We optimized the structure of a FeMoco resting-state model that takes those residues of the protein backbone into account which anchor the metal cluster in the enzyme (see Fig. 1 (right) in the main article). Note that different spin and charge states were considered for the structure optimization in order to obtain a variety of electronic structures for the assessment of a solution algorithm on a quantum computer. These spin and charge states do not necessarily match the one of the resting state of nitrogenase. Structure **1** for three positive excess charges and an equal number of α- and β-spins, and structure **2** for an uncharged FeMoco model with one unpaired α-spin. The Cartesian coordinates of these structures are collected in Tables 6 and 7. In these unrestricted DFT calculations, the spin symmetry was broken [73] and only S_z remains as a good quantum number. For these structure optimizations we chose the TURBOMOLE program package (V6.4) [74] and employed the B3LYP density functional [75–78] with the def2-TZVP Ahlrichs triple-zeta basis set plus polarization functions on all atoms [79]. An effective core potential was chosen only for the molybdenum atom [80], which also takes care of all scalar-relativistic effects on this heavy atom.

Then, integrals in the molecular orbital (MO) basis were produced MO integrals for structure **1** were generated for a CAS of 54 electrons in 54 spatial CASSCF orbitals (108 spin orbitals), which were obtained from a singlet CASSCF calculation with 24 electrons in 16 orbitals. Accordingly, for structure **2** we generated MO integrals for a CAS of 65 electrons in 57 spatial CASSCF orbitals (114 spin orbitals) from a quartet CASSCF calculation with 21 electrons in 12 orbitals. The choice of the large active spaces was based on Pulay's UNO-CAS criterion [81, 82], in which the occupation number of the natural orbitals serves as a selection criterion. However, rather than unrestricted Hartree–Fock natural orbitals, we selected those small-CAS CASSCF natural orbitals in the occupation intervals [1.98,0.02] and [1.99,0.01], respectively. The molecular orbital integrals for the second-quantized electronic Hamiltonian in these small-CAS orbital bases were calculated with the MOLCAS program [66].

All settings for the calculations are summarized in Table 5.

1. Nielsen MA, Chuang IL (2010) *Quantum computation and quantum information.* (Cambridge university press).
2. Bocharov A, Gurevich Y, Svore KM (2013) Efficient decomposition of single-qubit gates into v basis circuits. *Physical Review A* 88(1):012313.
3. Forest S, Gosset D, Kliuchnikov V, McKinnon D (2015) Exact synthesis of single-qubit unitaries over Clifford-cyclotomic gate sets. *Journal of Mathematical Physics* 56(8):082201.
4. Dawson CM, Nielsen MA (2005) The solovay-kitaev algorithm. *arXiv preprint quant-ph/0505030.*
5. Selinger P (2015) Efficient Clifford+ t approximation of single-qubit operators. *Quantum Information & Computation* 15(1-2):159–180.
6. Kliuchnikov V, Maslov D, Mosca M (2013) Fast and efficient exact synthesis of single-qubit unitaries generated by Clifford and t gates. *Quantum Information & Computation* 13(7-8):607–630.
7. Ross NJ, Selinger P (2014) Optimal ancilla-free Clifford+ t approximation of z-rotations. *arXiv preprint arXiv:1403.2975.*
8. Bocharov A, Roetteler M, Svore KM (2015) Efficient synthesis of probabilistic quantum circuits with fallback. *Physical Review A* 91(5):052317.
9. Jones NC et al. (2012) Faster quantum chemistry simulation on fault-tolerant quantum computers. *New Journal of Physics* 14(11):115023.
10. Wiebe N, Kliuchnikov V (2013) Floating point representations in quantum circuit synthesis. *New Journal of Physics* 15:093041.
11. Lanyon BP et al. (2010) Towards quantum chemistry on a quantum computer. *Nature Chemistry* 2(2):106–111.

Structure	Structure opt./B3LYP		small-CAS CASSCF orbitals			
	Total spin S	Charge	Act. electrons	Act. orbitals	Total spin S	Charge
1	0	3	54	54	0	3
2	$1/2$	0	65	57	$3/2$	0

Table 5. The structures optimized for FeMoco and the settings for the small-CAS CASSCF orbital optimization. S is the total spin quantum number and the charge is measured in units of the elementary charge.

Atom	Coordinates Structure **1**		
S	0.193509	-1.756174	-6.077728
FE	-0.073029	0.147462	-7.060398
S	-0.155670	-0.304014	-9.055299
FE	-1.194342	-0.633669	-4.857848
C	0.097884	0.293121	-3.822079
FE	0.330922	1.682738	-2.639449
S	0.144858	3.427365	-3.756994
FE	0.000866	1.806617	-5.056623
S	-1.759709	1.058486	-6.097784
FE	1.393308	-0.388901	-4.927653
S	2.935549	-1.083502	-3.683191
FE	1.581418	-0.451806	-2.290305
S	1.643565	1.242708	-0.922047
MO	-0.046077	-0.172415	0.259925
O	-0.081763	1.100331	1.760010
FE	-0.867802	-0.634367	-2.449654
S	-1.333714	1.160052	-1.286619
S	-2.410505	-1.577384	-3.437041
S	1.630797	1.240870	-6.345822
S	0.353345	-1.960563	-1.213579
N	1.652155	-0.910661	1.409953
O	-1.541962	-0.693648	1.167835
C	-2.111338	-0.028674	2.280713
C	-0.152453	1.144086	-10.129068
C	-1.083149	1.057101	2.683832
H	-3.060803	0.437015	2.003587
H	-2.276775	-0.745389	3.087505
O	-1.167081	1.740174	3.642482
H	0.625836	0.971552	-10.878592
H	-1.119129	1.154099	-10.642807
H	0.015059	2.077284	-9.598730
C	2.339711	-0.178076	2.302234
N	3.244459	-0.944382	2.902881
C	3.160159	-2.227901	2.407388
C	2.176178	-2.202712	1.475167
H	1.794821	-3.019854	0.890693
H	3.781670	-3.031530	2.764468
H	3.882526	-0.630293	3.624464
H	2.182421	0.864873	2.517705

Table 6. The coordinates for Structure 1 of FeMoco in Å.

Atom	Coordinates Structure **2**		
S	0.032866	-2.093214	-6.099450
FE	0.017622	-0.056816	-7.223052
S	0.143644	-0.401241	-9.315852
FE	-1.471854	-0.721074	-4.983076
C	-0.152873	0.225677	-3.664235
FE	-0.118135	1.845485	-2.405214
S	-0.036661	3.550618	-3.771677
FE	-0.049780	1.771611	-5.047025
S	-1.760858	1.077437	-6.479794
FE	1.306370	-0.657731	-4.880018
S	2.846629	-1.294232	-3.368478
FE	1.196355	-0.482357	-2.281465
S	1.566535	1.288016	-0.916398
MO	-0.186653	-0.006921	0.115165
O	-0.140683	1.329614	1.698525
FE	-1.529490	-0.648712	-2.302651
S	-1.922731	1.225686	-0.982284
S	-3.001156	-1.537533	-3.665135
S	1.731662	1.109781	-6.342619
S	-0.092449	-1.984660	-1.157533
N	1.598971	-0.845206	1.346832
O	-1.365000	-0.810743	1.369699
C	-1.748263	-0.161994	2.555155
C	0.321553	1.195240	-10.180658
C	-0.891249	1.088353	2.753014
H	-2.802446	0.128482	2.480632
H	-1.641761	-0.837543	3.407640
O	-0.904893	1.741420	3.768510
H	0.360712	0.987604	-11.249659
H	-0.529456	1.838638	-9.965894
H	1.243686	1.686635	-9.874888
C	2.191192	-0.163147	2.313386
N	3.212741	-0.875946	2.820353
C	3.284255	-2.074995	2.147363
C	2.276098	-2.042456	1.234236
H	1.996684	-2.787079	0.511609
H	4.022709	-2.823292	2.371900
H	3.814462	-0.573873	3.568856
H	1.907357	0.820486	2.645710

Table 7. Coordinates for Structure 2 of FeMoco in Å.

12. Seeley JT, Richard MJ, Love PJ (2012) The bravyi-kitaev transformation for quantum computation of electronic structure. *The Journal of chemical physics* 137(22):224109.
13. Whitfield JD, Biamonte J, Aspuru-Guzik A (2011) Simulation of electronic structure hamiltonians using quantum computers. *Molecular Physics* 109(5):735–750.
14. Wecker D, Bauer B, Clark BK, Hastings MB, Troyer M (2014) Gate-count estimates for performing quantum chemistry on small quantum computers. *Physical Review A* 90(2):022305.
15. Fowler AG, Stephens AM, Groszkowski P (2009) High-threshold universal quantum computation on the surface code. *Physical Review A* 80(5):052312.
16. Barends R et al. (2014) Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* 508(7497):500–503.
17. Paetznick A, Reichardt BW (2013) Universal fault-tolerant quantum computation with only transversal gates and error correction. *Physical review letters* 111(9):090505.
18. Bombin H, Martin-Delgado M (2009) Quantum measurements and gates by code deformation. *Journal of Physics A: Mathematical and Theoretical* 42(9):095302.
19. Bravyi S, Cross A (2015) Doubled color codes. *arXiv preprint arXiv:1509.03239*.
20. Jones C (2013) Multilevel distillation of magic states for quantum computing. *Physical Review A* 87(4):042305.
21. Bravyi S, Kitaev A (2005) Universal quantum computation with ideal Clifford gates and noisy ancillas. *Physical Review A* 71(2):022316.
22. Metodi TS, Thaker DD, Cross AW, Chong FT, Chuang IL (2005) A quantum logic array microarchitecture: Scalable quantum data movement and computation in *38th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO'05)*. (IEEE), pp. 12–pp.
23. Van Meter R, Devitt SJ (2016) Local and distributed quantum computation. *arXiv preprint arXiv:1605.06951*.
24. Taylor J et al. (2005) Fault-tolerant architecture for quantum computation using electrically controlled semiconductor spins. *Nature Physics* 1(3):177–183.
25. Wiseman H, Killip R (1997) Adaptive single-shot phase measurements: A semiclassical approach. *Physical Review A* 56(1):944.
26. Van Dam W, D'Ariano GM, Ekert A, Macchiavello C, Mosca M (2007) Optimal phase estimation in quantum networks. *Journal of Physics A: Mathematical and Theoretical* 40(28):7971.
27. Berry DW et al. (2009) How to perform the most accurate possible phase measurements. *Physical Review A* 80(5):052114.
28. Wiebe N, Granade C (2016) Efficient bayesian phase estimation. *Phys. Rev. Lett.* 117:010503.
29. Wiebe N, Granade C, Ferrie C, Cory D (2014) Hamiltonian learning and certification using quantum resources. *Physical review letters* 112(19):190501.
30. Hentschel A, Sanders BC (2010) Machine learning for precise quantum measurement. *Physical review letters* 104(6):063603.
31. Granade CE, Ferrie C, Wiebe N, Cory DG (2012) Robust online hamiltonian learning. *New Journal of Physics* 14(10):103013.
32. Bonato C et al. (2015) Optimized quantum sensing with a single electron spin using real-time adaptive measurements. *Nature nanotechnology*.
33. Wecker D et al. (2015) Solving strongly correlated electron models on a quantum computer. *Physical Review A* 92(6):062318.
34. Hastings MB, Wecker D, Bauer B, Troyer M (2015) Improving quantum algorithms for quantum chemistry. *Quantum Information & Computation* 15(1-2):1–21.
35. Babbush R, McClean J, Wecker D, Aspuru-Guzik A, Wiebe N (2015) Chemical basis of trotter-suzuki errors in quantum chemistry simulation. *Physical Review A* 91(2):022311.
36. Poulin D et al. (2015) The trotter step size required for accurate quantum simulation of quantum chemistry. *Quantum Information & Computation* 15(5-6):361–384.
37. Raeisi S, Wiebe N, Sanders BC (2012) Quantum-circuit design for efficient simulations of many-body quantum dynamics. *New Journal of Physics* 14(10):103017.
38. Garcíaa VM, Castell O, Caballol R, Malrieu JP (1995) An iterative difference-dedicated configuration interaction. proposal and test studies. *Chem. Phys. Lett.* 238:222–229.
39. Sugisaki K et al. (2016) Quantum chemistry on quantum computers: A polynomial-time quantum algorithm for constructing the wave functions of open-shell molecules. *The Journal of Physical Chemistry A* 120(32):6459–6466.
40. Legeza Ö, Sólyom J (year?) International workshop on recent progress and prospects in density-matrix renormalization. Lorentz Center, Leiden University, The Netherlands, **2004**.
41. Legeza Ö (year?) Cecam workshop for tensor network methods for quantum chemistry. ETH Zürich, **2010**.
42. Keller SF, Reiher M (2014) Determining factors for the accuracy of dmrg in chemistry. *CHIMIA International Journal for Chemistry* 68(4):200–203.
43. Stein CJ, Reiher M (2016) Automated selection of active orbital spaces. *J. Chem. Theory Comput.* 12:1760–1771.
44. Wu LA, Byrd M, Lidar D (2002) Polynomial-time simulation of pairing models on a quantum computer. *Physical Review Letters* 89(5):057904.
45. Cheung D, Høyer P, Wiebe N (2011) Improved error bounds for the adiabatic approximation. *Journal of Physics A: Mathematical and Theoretical* 44(41):415302.
46. Wiebe N, Berry D, Høyer P, Sanders BC (2010) Higher order decompositions of ordered operator exponentials. *Journal of Physics A: Mathematical and Theoretical* 43(6):065203.
47. Wiebe N, Berry DW, Høyer P, Sanders BC (2011) Simulating quantum dynamics on a quantum computer. *Journal of Physics A: Mathematical and Theoretical* 44(44):445308.
48. McClean JR, Babbush R, Love PJ, Aspuru-Guzik A (2014) Exploiting locality in quantum computation for quantum chemistry. *The journal of physical chemistry letters* 5(24):4368–4380.
49. Bauer B, Wecker D, Millis AJ, Hastings MB, Troyer M (2015) Hybrid quantum-classical approach to correlated materials. *arXiv preprint arXiv:1510.03859*.
50. Knill E, Ortiz G, Somma RD (2007) Optimal quantum measurements of expectation values of observables. *Physical Review A* 75(1):012328.
51. Elgart A, Hagedorn GA (2012) A note on the switching adiabatic theorem. *Journal of Mathematical Physics* 53(10):102202.
52. Lidar DA, Rezakhani AT, Hamma A (2009) Adiabatic approximation with exponential accuracy for many-body systems and quantum computation. *Journal of Mathematical Physics* 50(10):102106.
53. Wiebe N, Babcock NS (2012) Improved error-scaling for adiabatic quantum evolutions. *New Journal of Physics* 14(1):013024.
54. Berry DW, Ahokas G, Cleve R, Sanders BC (2007) Efficient quantum algorithms for simulating sparse hamiltonians. *Communications in Mathematical Physics* 270(2):359–371.
55. Svore KM, Hastings MB, Freedman M (2014) Faster phase estimation. *Quantum Info. Comput.* 14(3-4):306–328.
56. Yandulov DV, Schrock RR (2003) Catalytic reduction of dinitrogen to ammonia at a single molybdenum center. *Science* 301:76–78.
57. Arashiba K, Miyake Y, Nishibayashi Y (2011) A molybdenum complex bearing pnp-type pincer ligands leads to the catalytic reduction of dinitrogen into ammonia. *Nature Chem.* 3:120–125.
58. Anderson JS, Rittle J, Peters JC (2013) Catalytic conversion of nitrogen to ammonia by an iron model complex. *Nature* 501:84–87.
59. Leigh GJ, ed. (2002) *Nitrogen Fixation at the Millenium.* (Elsevier Science, Amsterdam).
60. Kim J, Rees DC (1992) Structural Models for the Metal Centers in the Nitrogenase Molybdenum-Iron Protein. *Science* 257:1677–1682.
61. Einsle O et al. (2002) Nitrogenase MoFe-Protein at 1.16 ÅResolution: A Central Ligand in the FeMo-Cofactor. *Science* 297:1696–1700.
62. Spatzal T et al. (2011) Evidence for interstitial carbon in nitrogenase femo cofactor. *Science* 334:940.
63. Lancaster KM et al. (2011) X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science* 334:974–977.
64. Ribbe MW, Hu Y, Hodgson KO, Hedman B (2014) Biosynthesis of nitrogenase metalloclusters. *Chem. Rev.* 114:4063–4080.
65. Bergeler M, Simm GN, Proppe J, Reiher M (2015) Heuristics-guided exploration of reaction mechanisms. *J. Chem. Theory Comput.* 11:5712–5722.
66. Aquilante F et al. (2016) Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* pp. 506–541.
67. Helgaker T, Jørgensen P, Olsen J (2000) *Molecular Electronic-Structure Theory*. (John Wiley & Sons).
68. Fromager E, Toulouse J, Jensen HJÅ (2007) On the universality of the long-/short-range separation in multiconfigurational density-functional theory. *J. Chem. Phys.* 126:074111.
69. Hedegård ED, Knecht S, Kielberg JS, Jensen HJA, Reiher M (2015) Density matrix renormalization group with efficient dynamical electron correlation through range separation. *J. Chem. Phys.* 142(22):224108.
70. Weymuth T, Couzijn EPA, Chen P, Reiher M (2014) New benchmark set of transition-metal coordination reactions for the assessment of density functionals. *J. Chem. Theory Comput.* 10:3092–3103.
71. Proppe J, Husch T, Simm GN, Reiher M (2016) Uncertainty quantification for quantum chemical models of complex reaction networks. *Faraday Discussions* 195:497–520.
72. Liu C, Liu T, Hall MB (2015) Influence of the Density Functional and Basis Set on the Relative Stabilities of Oxygenated Isomers of Diiron Models for the Active Site of [FeFe]-Hydrogenase. *J. Chem. Theory Comput.* 11:205–214.
73. Jacob CR, Reiher M (2012) Spin in density-functional theory. *Int. J. Quantum Chem.* 112:3661–3684.
74. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Electronic Structure Calculations on Workstation Computers: The Program System TURBOMOLE. *Chem. Phys. Lett.* 162:165–169.
75. Lee C, Yang W, Parr RG (1988) Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37:785–789.
76. Becke AD (1988) Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* 38:3098–3100.
77. Becke AD (1993) Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.* 98:5648–5652.
78. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* 98:11623–11627.
79. Weigend F, Ahlrichs R (2005) 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.* 7:3297–3305.
80. Andrae D, Häußermann U, Dolg M, Stoll H, Preuß H (1990) Energy-adjusted $ab\ initio$ Pseudopotentials for the Second and Third Row Transition Elements. *Theor. Chim. Acta* 77:123–141.
81. Pulay P, Hamilton TP (1988) Uhf natural orbitals for defining and starting mc-scf calculations. *J. Chem. Phys.* 88(8):4926–4933.
82. Keller S, Boguslawski K, Janowski T, Reiher M, Pulay P (2015) Selection of active spaces for multiconfigurational wavefunctions. *J. Chem. Phys* 142(24):244104.