# A Scalable Approach to Quantum Simulation via Projection-Based Embedding

Alexis Ralli,[1] Michael Williams de la Bastida,[1,*] and Peter V. Coveney[1,2,3]

[1] *Centre for Computational Science, Department of Chemistry, University College London, WC1H 0AJ, United Kingdom*[†]
[2] *UCL Centre for Advanced Research Computing, Gower Street , London WC1E 6BT, United Kingdom*
[3] *Informatics Institute, University of Amsterdam, Amsterdam, 1098 XH, Netherlands* [‡]

(Dated: September 22, 2023)

Owing to the computational complexity of electronic structure algorithms running on classical digital computers, the range of molecular systems amenable to simulation remains tightly circumscribed even after many decades of work. Many believe quantum computers will transcend such limitations although in the current era the size and noise of these devices militates against significant progress. Here we describe a chemically intuitive approach that permits a subdomain of a molecule's electronic structure to be calculated accurately on a quantum device, while the rest of the molecule is described at a lower level of accuracy using density functional theory running on a classical computer. We demonstrate that this approach produces improved results for molecules that cannot be simulated fully on current quantum computers but which can be resolved classically at a cheaper level of approximation. The algorithm is tunable, so that the size of the quantum simulation can be adjusted to run on available quantum resources. Therefore, as quantum devices become larger, this method will enable increasingly large subdomains to be studied accurately.

## I. INTRODUCTION

Quantum computing is anticipated to enable accurate simulation of chemical systems beyond the capabilities of classical methods. Whether this aim will be achieved with so-called Noisy Intermediate-scale Quantum (NISQ) processors is still to be seen[1–4]. While devices are improving rapidly, NISQ applications also require algorithmic tools to mitigate noise and reduce required qubit counts.

Embedding procedures work by first partitioning a system and then applying differing levels of theory to each region. An accurate but computationally expensive method is applied to a small *active region*[5, 6]. The surrounding *environment* is handled with a more efficient but approximate method. This allows some of the physically relevant detail to be captured while avoiding the computational cost of accurately simulating the entire system. However, even for fairly small active regions, exact classical simulation using the Full Configuration Interaction (FCI) method quickly becomes unfeasible due problem scaling exponentially with system size [7]. For such problems, the number of Slater determinants scales as $\binom{M}{N}$, for $N$ electrons and $M$ orbitals [4].

The current "gold standard" in conventional quantum chemistry is coupled cluster (CC) theory, which offers a good accuracy-to-cost ratio and reduces this factorial complexity [8, 9]. The CC single double (CCSD) method scales as $\mathcal{O}(M^6)$ [10]. The CCSD(T), which treats the triple excitations pertubatively, scales as $\mathcal{O}(M^7)$ in time [4]. This still imposes practical limitations on system size while imperfectly approximating the effects of correlation [11]. Therefore, classical embedding methods still inevitably inherit the shortcomings of such methods, even within a smaller active region. In short, accurately simulating quantum effects at large scale remains elusive.

Quantum computers can efficiently represent the state of general quantum systems and provide a practical way to perform quantum chemistry simulations in polynomial time [12]; however, this approach will only be possible in the fault tolerant regime, as it requires the quantum phase estimation (QPE) algorithm [13] which cannot be implemented on current NISQ quantum computers [14, 15]. There are however, still open questions regarding the advantage of using quantum computers for chemistry simulations [16]. Quantum algorithms designed for NISQ devices, such as the variational quantum eigensolver (VQE) [17], allow quantum systems to be studied using present day hardware; however, this is limited by the current quality and quantity of qubits. To date, the largest chemical VQE simulations are a 12 qubit study of a $H_{12}$ chain [18] and a 12 qubit simulation of $Li_2O$ [19].

By embedding a wave function simulation calculated on a quantum computer into a larger classical simulation, we can mitigate some of the shortfalls of classical hardware in describing quantum systems, while requiring fewer qubits and shorter quantum circuits than full-system quantum simulation. This will allow systems normally too large to study at the wave function level to be modelled via a multi-scale approach. In this way, embedding can serve as an algorithmic tool to mitigate the shortcomings of quantum and classical processors, thereby providing novel results. Additionally, as embedding methods may utilise fault-tolerant quantum

* michael.williams.20@ucl.ac.uk
† alexis.ralli.18@ucl.ac.uk
‡ p.v.coveney@ucl.ac.uk

simulation methods, they will continue to facilitate the study of systems larger than would otherwise be possible. Hybrid embedding methods published to date include wave function-in-DFT [20, 21], Density Matrix Embedding Theory (DMET) [22–24] and Dynamical Mean-Field Theory (DMFT) [25–27] approaches.

We reformulate the projection-based embedding method, which enables the application of quantum algorithms to molecules of arbitrary size while consistently improving on the results of full-system Density Functional Theory (DFT). This method outputs a Hamiltonian which can be solved using any suitable NISQ or fault-tolerant quantum algorithm, thus augmenting the usefulness of quantum processors in general. We anticipate that by targeting quantum processors at regions with strong correlation, hybrid embedding will enable novel results. This concept has been utilized in other works, where a smaller embedded qubit Hamiltonian is produced and solved; however, the underlying theory and assumptions are different for each hybrid model [20, 28, 29]. The method presented in this work is distinct from these.

## II. BACKGROUND

In this section, we cover the necessary background on the projection-based embedding technique and discuss the molecular orbital (MO) localization methods considered in this work.

### A. Projection-Based Embedding

Quantum embedding schemes for electronic structure problems aim to reduce the computational cost of a problem by dividing a molecular problem into smaller (and so less-costly) subsystems [30]. In this paper, we employ the projection-based embedding (PBE) method, first proposed by Manby *et al.* [31], that offers a practical way to implement formally exact quantum embedding. At a high level, a molecular problem is split into an active and environment region. The active system is then solved at a more accurate (and thus computationally more expensive) level of theory than the environment. The PBE method allows rigorous embedding of either a wavefunction subsystem into a self-consistent field (SCF) environment (WFT-in-SCF embedding) or an SCF subsystem description in an SCF environment (SCF-in-SCF embedding). Here the SCF method should be thought of as Hartree-Fock or Kohn-Sham DFT (KS-DFT). We summarise the important details of this model here, where the environment SCF calculation is performed using KS-DFT.

To begin, an initial KS-DFT calculation of the entire system is carried out using a low (cheap) level of theory. This yields a set of molecular orbitals (MOs) $\{\psi_i(\vec{r})|i = 1, 2, \ldots, N\}$. Each MO is formed from a linear combination of $K$ known atomic orbital (AO) basis functions $\{\phi_j(\vec{r})|j = 1, 2, \ldots, K\}$:

$$\psi_i(\vec{r}) = \sum_{j=1}^{K} \mathbf{C}_{ji}\phi_j(\vec{r}), \tag{1}$$

where $\mathbf{C}$ is a matrix of MO coefficients. In general, the AO basis functions $\phi_j(\vec{r})$ are not orthonormal. We can see this by the $(K \times K)$ overlap matrix:

$$\mathbf{S}_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle = \int d\vec{r}\, \phi_\mu(\vec{r})^*\phi_\nu(\vec{r}). \tag{2}$$

If $\mathbf{S}$ is the identity matrix then all the AO basis functions are orthonormal; however, in general this is not the case. However, linear combinations of these non-orthogonal AOs, given by the columns of $\mathbf{C}$, construct orthogonal MOs $\psi_i(\vec{r})$ - i.e. $\mathbf{C}^\dagger\mathbf{S}\mathbf{C} = \mathbf{I}$.

In order to partition the molecular problem into an active and environment part, these canonical MOs $\psi_i$ must first be localized and assigned to a subsystem. This can be done by different localisation methods - described in further detail in the Supplementary Material. In effect, we use a unitary transform $\mathbf{U}$ (defined by the localization procedure) to spatially localize each $\psi_i$ as much as possible. The reasoning behind this is introduced next.

Lehtola and Jónsson noted that "total energy in both Hartree-Fock and KS-DFT is invariant under a unitary transformation of the occupied-occupied and virtual-virtual blocks" [32]. This is true because the Hartree-Fock and KS-DFT wavefunctions are approximated as a single Slater determinant. From linear algebra, it is well known that the determinant of a matrix product is given by the product of their determinants - i.e. for general matrices $det(\mathbf{AB}) = det(\mathbf{A})det(\mathbf{B})$. Using this property and the fact that the determinant of the identity matrix is $det(\mathbf{I}) = 1$. Given a unitary matrix $\mathbf{V}$, where $\mathbf{V}^\dagger\mathbf{V} = I$, we obtain the following:

$$\begin{aligned} 1 = det(\mathbf{I}) &= det(\mathbf{V}^\dagger\mathbf{V}) = det(\mathbf{V}^\dagger)det(\mathbf{V}) \\ &= det(\mathbf{V})^* det(\mathbf{V}) = |det(\mathbf{V})|^2. \end{aligned} \tag{3}$$

This implies that the determinant of any unitary matrix must have a value of $e^{i\theta}$, as $|e^{i\theta}| = 1$. Therefore, acting with a unitary will leave the Slater determinant unchanged up to a global phase and so observable quantities of the wavefunction will be unchanged. Thus the solution of an SCF problem can be described by a set of different (unitarily) rotated orbitals. Such a unitary rotation $\mathbf{U}$ can be used to spatially localize each MO $\psi_i$ as much as possible. The form of $\mathbf{U}$ is defined by a particular localization procedure and there are many methods based on different localization criteria. We denote these orbitals as localized molecular orbitals (LMOs) or $\psi_i^{LMO}$. The matrix of orbital coefficients for these localized orbitals are given by the columns of $\mathbf{C}^{LMO}$ defined as:

$$\mathbf{C}^{LMO} = \mathbf{C}\mathbf{U}. \tag{4}$$

This construction ensures the orthonormality condition of each molecular orbital is still conserved - i.e. $\left(\mathbf{C}^{LMO}\right)^\dagger \mathbf{S}\mathbf{C}^{LMO} = \mathbf{I}$. We see this via the following proof:

$$\begin{aligned}\left(\mathbf{C}^{LMO}\right)^\dagger \mathbf{S}\mathbf{C}^{LMO} &= \mathbf{U}^\dagger\mathbf{C}^\dagger\mathbf{S}\mathbf{C}\mathbf{U} \\ &= \mathbf{U}^\dagger\mathbf{I}\mathbf{U} = \mathbf{I}\mathbf{U}^\dagger\mathbf{U} \\ &= \mathbf{I}\end{aligned} \tag{5}$$

This construction allows for $\mathbf{U}$ to be determined from $\mathbf{C}$ and $\mathbf{C}^{LMO}$:

$$\begin{aligned}\left(\mathbf{C}^{LMO}\right)^\dagger \mathbf{S}\mathbf{C}^{LMO} = \mathbf{U}^\dagger\mathbf{C}^\dagger\mathbf{S}\mathbf{C}^{LMO} = \mathbf{I} \\ \implies \mathbf{U} = \mathbf{C}^\dagger\mathbf{S}\mathbf{C}^{LMO}\end{aligned} \tag{6}$$

by multiplying on the left with $\mathbf{U}$. The reason we include equation 6 is sometimes quantum chemistry packages only return $\mathbf{C}^{LMO}$ without $\mathbf{U}$.

In summary, we get the following mapping from canonical to localised molecular orbitals:

$$\begin{aligned}\psi_i(\vec{r}) = \sum_{j=1}^{K} \mathbf{C}_{ji}\phi_j(\vec{r}) \mapsto \psi_i^{LMO} &= \sum_{j=1}^{K} [\mathbf{C}\mathbf{U}]_{ji}\phi_j(\vec{r}) \\ &= \sum_{j=1}^{K} \mathbf{C}_{ji}^{LMO}\phi_j(\vec{r}).\end{aligned} \tag{7}$$

Next we show how the charge density remains unchanged.

For a closed shell molecule, described by a single determinant wave function, each MO $\psi_i$ contains two electrons and thus the total charge density is [7]:

$$\begin{aligned}\rho(\vec{r}) &= 2\sum_{i=1}^{N/2} \psi_i^*(\vec{r})\psi_i(\vec{r}) \\ &= 2\sum_{i=1}^{N/2} \left( \underbrace{\sum_{\nu=1}^{K} \mathbf{C}_{\nu i}^*\phi_\nu^*(\vec{r})}_{\psi_i^*(\vec{r})} \underbrace{\sum_{\mu=1}^{K} \mathbf{C}_{\mu i}\phi_\mu(\vec{r})}_{\psi_i(\vec{r})} \right) \\ &= \sum_{\mu=1}^{K}\sum_{\nu=1}^{K} \left[ 2\sum_{i=1}^{N/2} \mathbf{C}_{\mu i}\mathbf{C}_{\nu i}^* \right] \phi_\nu^*(\vec{r})\phi_\mu(\vec{r}) \\ &= \sum_{\mu=1}^{K}\sum_{\nu=1}^{K} \gamma_{\mu\nu}\phi_\nu^*(\vec{r})\phi_\mu(\vec{r}).\end{aligned} \tag{8}$$

Here the square brackets define the density matrix $\gamma_{\mu\nu}$ (defined in the AO basis):

$$\gamma_{\mu\nu} = 2\sum_{i=1}^{N/2} \mathbf{C}_{\mu i}\mathbf{C}_{\nu i}^\dagger, \tag{9}$$

that for a set of basis function $\{\phi_j(\vec{r})|j = 1, 2, \ldots, K\}$ fully specifies the charge density $\rho(\vec{r})$ [7]. The sum runs over $N/2$, as these are the occupied MOs of a closed shell calculation. The whole matrix can be obtained as $\gamma^{total} = 2\mathbf{C}_{occ}(\mathbf{C}_{occ})^\dagger$, where *occ* denotes only using the occupied columns of the $\mathbf{C}$ matrix (the first $N/2$ columns, indexed by $i$ in Equation 9). In the localized basis, the density matrix remains unchanged as:

$$\begin{aligned}\gamma^{total} &= 2\mathbf{C}_{occ}(\mathbf{C}_{occ})^\dagger \\ &= 2[\mathbf{C}_{occ}^{LMO}\mathbf{U}]\,[\mathbf{U}^\dagger(\mathbf{C}_{occ}^{LMO})^\dagger] \\ &= 2\mathbf{C}_{occ}^{LMO}(\mathbf{C}_{occ}^{LMO})^\dagger.\end{aligned} \tag{10}$$

Given a set of localised molecular orbitals, we partition them into two subsystems denoted *act* (active) and *env* (environment). There are different methods to do so and we summarise our approach in the Supplementary Material. Overall we generate a set of (occupied) LMO indices $\mathcal{K}$ and $\mathcal{L}$ for the active and environment subsystems respectively. The resulting charge density for each subsystem can then be written as:

$$\gamma_{\mu\nu}^{act} = 2\sum_{k\in\mathcal{K}} \mathbf{C}_{\mu k}^{LMO}(\mathbf{C}_{\nu k}^{LMO})^\dagger, \tag{11a}$$

$$\gamma_{\mu\nu}^{env} = 2\sum_{l\in\mathcal{L}} \mathbf{C}_{\mu l}^{LMO}(\mathbf{C}_{\nu l}^{LMO})^\dagger, \tag{11b}$$

for closed-shell calculations. The set $\mathcal{K}\cup\mathcal{L}$ contains all the occupied molecular orbital indices.

The total system electron density is written as a sum of subsystem densities:

$$\begin{aligned}\gamma^{total} &= \gamma^{act} + \gamma^{env} \\ &= 2\mathbf{C}_{\mathcal{K}}^{LMO}(\mathbf{C}_{\mathcal{K}}^{LMO})^\dagger + 2\mathbf{C}_{\mathcal{L}}^{LMO}(\mathbf{C}_{\mathcal{L}}^{LMO})^\dagger \\ &= 2\mathbf{C}_{occ}^{LMO}(\mathbf{C}_{occ}^{LMO})^\dagger.\end{aligned} \tag{12}$$

The number of electrons will also be split according to $n_e^{total} = n_e^{act} + n_e^{env} = \mathrm{Tr}(\mathbf{S}\gamma^{act}) + \mathrm{Tr}(\mathbf{S}\gamma^{env}) = \mathrm{Tr}\big(\mathbf{S}\gamma^{total}\big)$, where Tr denotes the trace operation.

The energy of the full system can be found from its components via [33]:

$$\begin{aligned}E[\gamma^{act}, \gamma^{env}] = &\underbrace{\mathrm{Tr}\big(\gamma^{act}\mathbf{h}_{core}\big) + \mathbf{g}(\gamma^{act})}_{\text{energy of isolated } act \text{ system}} + \\ &\underbrace{\mathrm{Tr}(\gamma^{env}\mathbf{h}_{core}) + \mathbf{g}(\gamma^{env})}_{\text{energy of isolated } env \text{ system}} + \\ &\underbrace{\mathbf{g}(\gamma^{act}, \gamma^{env})}_{\text{nonadditive two-electron energy}}.\end{aligned} \tag{13}$$

Here $\mathbf{h}_{core}$ is the one-electron core Hamiltonian and $\mathbf{g}$ groups the two-electron terms - Coulomb and exchange

for Hartree-Fock and exchange-correlation for DFT. The nonadditive two-electron energy is given by:

$$\mathbf{g}(\gamma^{act}, \gamma^{env}) = \mathbf{g}(\gamma^{act} + \gamma^{env}) - \mathbf{g}(\gamma^{act}) - \mathbf{g}(\gamma^{env}), \quad (14)$$

and accounts for the interaction between subsystems [33].

Next we want to solve the active system using a higher (more accurate) level of theory. The effect of the interaction between the active and environment subsystems is accounted for by additional terms in the core Hamiltonian. The Fock matrix for the active system embedded in the environment system is [31]:

$$\begin{aligned} \mathbf{F}^{act}_{emb} &= \mathbf{h}_{core} + \mathbf{V}_{emb} + \mathbf{P}^{env}_{proj} + \mathbf{g}(\gamma^{act}_{emb}) \\ &= \mathbf{h}_{emb} + \mathbf{g}(\gamma^{act}_{emb}), \end{aligned} \quad (15)$$

where:

$$\mathbf{V}_{emb} = \mathbf{g}(\gamma^{act} + \gamma^{env}) - \mathbf{g}(\gamma^{act}), \quad (16a)$$

$$\mathbf{h}_{emb} = \mathbf{h}_{core} + \mathbf{V}_{emb} + \mathbf{P}^{env}_{proj}. \quad (16b)$$

The embedding potential $\mathbf{V}_{emb}$ describes all the interactions (nonadditive part) between the active and environment subsystems [34]. Due to the subsystem densities (Equation 12) being constructed from disjoint subsets of orthogonal orbitals, the normally difficult-to-evalute nonadditive kinetic potential (NAKP) terms [35] are exactly zero [31, 34, 36].

$\mathbf{P}^{env}_{proj}$ is a projection operator that enforces inter-subsystem (orbital) orthogonality. There are different ways to define this operator and we consider two in this work. The first definition was proposed by the Manby and Miller groups[31]. They use a parameter ($\mu$) to shift the orbital energies of the environment to high energies - effectively meaning they will never by occupied. This projector is defined as:

$$\begin{aligned} (\mathbf{P}^{env}_{\mu})_{ij} &= \mu \left\langle \psi^{LMO}_i \right| \mathbf{P}^{env} \left| \psi^{LMO}_j \right\rangle \\ &= \mu [\mathbf{S}\gamma^{env}\mathbf{S}]_{ij}, \end{aligned} \quad (17)$$

where $\mu$ is some large integer, $\mathbf{S}$ is the AO overlap matrix. $\mathbf{P}^{env}$ is a projector defined as:

$$\mathbf{P}^{env} = \sum_{l \in \mathcal{L}} \left| \psi^{LMO}_l \right\rangle \left\langle \psi^{LMO}_l \right|. \quad (18)$$

Here we use the notation $l \in \mathcal{L}$ to mean the sum over the set of occupied MO indices for the environment orbitals. It has been shown that $\mu$ is numerically robust and can usually be set to $\mu = 10^6$ [31, 33]. In the limit that $\mu \to \infty$ this method is exact, as the environment orbitals will be pushed to infinite energy and thus will never be occupied. The action of this operator with the Fock matrix is:

$$(\mathbf{F} + \mathbf{P}^{env}_{\mu}) \left| \psi^{LMO}_k \right\rangle = \epsilon^{act}_k \left| \psi^{LMO}_k \right\rangle, \quad (19a)$$

$$\begin{aligned} (\mathbf{F} + \mathbf{P}^{env}_{\mu}) \left| \psi^{LMO}_l \right\rangle &= (\epsilon^{env}_l + \mu) \left| \psi^{LMO}_l \right\rangle \\ &\approx +\mu \left| \psi^{LMO}_l \right\rangle. \end{aligned} \quad (19b)$$

Again, $k \in \mathcal{K}$ and $l \in \mathcal{L}$ are occupied LMOs of the active and environment subsystems respectively. Qualitatively the orbital energies of the active system are left unchanged and the orbitals for the environment are pushed to very high energies as $\mu >> \epsilon^{env}_l$ - effectively suppressing transitions to these states and stopping hybridisation.

The second approach, proposed by Kallay *et al.* [37], is to use the Huzinaga projector [38, 39]:

$$\begin{aligned} \mathbf{P}^{env}_{huz} &= -\big(\mathbf{F}\mathbf{P}^{env} + \mathbf{P}^{env}\mathbf{F}\big) \\ &= -\frac{1}{2}\big(\mathbf{F}\gamma^{env}\mathbf{S} + \mathbf{S}\gamma^{env}\mathbf{F}\big). \end{aligned} \quad (20)$$

Note that the $-\frac{1}{2}$ prefactor is needed for closed-shell systems. This operator enforces orthogonality of the occupied orbitals of each subsystem [40]. The form of this operator increases the orbital energy for the occupied environment orbitals and leaves the active system unchanged. We write its action formally as:

$$(\mathbf{F} + \mathbf{P}^{env}_{huz}) \left| \psi^{LMO}_k \right\rangle = \epsilon^{act}_k \left| \psi^{LMO}_k \right\rangle, \quad (21a)$$

$$\begin{aligned} (\mathbf{F} + \mathbf{P}^{env}_{huz}) \left| \psi^{LMO}_l \right\rangle &= (\epsilon^{env}_l - 2\epsilon^{env}_l) \left| \psi^{LMO}_l \right\rangle \\ &= -1\epsilon^{env}_l \left| \psi^{LMO}_l \right\rangle. \end{aligned} \quad (21b)$$

It is usually the case that for the occupied molecular orbitals $\epsilon^{env}_l$ is negative and so equation 21b shows how the environment $\epsilon^{env}_l$ orbitals are normally shifted to positive energies and thus will not be filled. This formalism guarantees that $[\mathbf{P}^{env}_{huz}, \mathbf{F}^{act}_{emb}] = 0$ and removes the need for the $\mu$ parameter shift [41]. If any occupied MO has a positive $\epsilon^{env}_l$, then equation 20 can be modified to treat such systems correctly - this operator is known as the "Fermi-shifted Huzinaga operator" [42].

The energy of the active system embedded in the environment is given by:

$$\begin{aligned} E[\gamma^{act}_{emb}; \gamma^{act}, \gamma^{env}] = \mathcal{E}[\gamma^{act}_{emb}] + E[\gamma_{env}] + \mathbf{g}(\gamma^{act}, \gamma^{env}) \\ + \mathrm{Tr}\left((\gamma^{act}_{emb} - \gamma^{act})(\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj})\right), \end{aligned} \quad (22)$$

colloquially denoted as a DFT-in-DFT calculation.

We use the same notation as Claudino and coworkers [33], where $\mathcal{E}$ differs from $E$ as it allows for different functionals to be applied and is computed from the embedded density matrix of the active system. Note that Equation 15 is solved self-consistently to give $\gamma^{act}_{emb}$. Equation 22

reduces to Equation 13 for the case that the active and environment regions are treated at the same level of theory [33].

Importantly $\mathcal{E}[\gamma^{act}_{emb}] = \mathrm{Tr}(\gamma^{act}_{emb}\mathbf{h}_{core}) + \mathbf{g}(\gamma^{act}_{emb})$ and does not involve $\mathbf{V}_{emb}$ or $\mathbf{P}^{env}_{proj}$. The final term in Equation 22 is a first-order correction that accounts for the difference between $\mathbf{g}(\gamma^{act}, \gamma^{env})$ and $\mathbf{g}(\gamma^{act}_{emb}, \gamma^{env})$, and corrects for the fact that in general $\gamma^{act} \neq \gamma^{act}_{emb}$ [43].

This projection based embedding approach then allows for the active system to be treated using some wave function level of theory and therefore studied using a quantum computer. The electronic energy for this is given by [33]:

$$E[\Psi^{act}_{emb};\gamma^{act},\gamma^{env}] = \left\langle \Psi^{act}_{emb} \right| \mathbf{H}_{emb} \left| \Psi^{act}_{emb} \right\rangle + E[\gamma_{env}] + \mathbf{g}(\gamma^{act},\gamma^{env}) - \mathrm{Tr}\left(\gamma^{act}(\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj})\right). \tag{23}$$

Importantly $\mathbf{H}_{emb} = \mathbf{h}_{emb} + \mathbf{g}(\Psi^{act}_{emb})$, where $\mathbf{g}(\Psi^{act}_{emb})$ is the two-electron operator for a given wave function method and $\mathbf{h}_{emb}$ is the embedded core Hamiltonian (Equation 16b) which depends on $\gamma^{act}$ and $\gamma^{env}$ [44]. As the embedding terms have been included in $\mathbf{H}_{emb}$, the final correction term is therefore slightly different to Equation 22 [43]. The wave function calculation in Equation 23 includes contributions from $(\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj})$ - similar to: $\mathrm{Tr}\big(\gamma^{act}_{emb}(\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj})\big)$. The correction therefore only requires subtracting $\mathrm{Tr}\big(\gamma^{act}(\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj})\big)$, unlike in Equation 22, where $\mathcal{E}$ does not use $(\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj})$ to calculate the energy of the active system.

### B. MO localisation methods

In this work, we only use the Subsystem Projected AO DEcomposition (SPADE) and intrinsic bonding orbitals (IBO) localized molecular orbitals [33, 45]. The motivation for using SPADE is primarily that it does not require a parameterised heuristic to determine the active and environment subsystems. Futhermore, unlike most localization schemes that attempt to localize orbitals in atoms or bonds, SPADE orbitals are local only in the sense that they remain in their native subsystems, which is one of the requirements for successful embedding [33]

IBOs were used as they only depend on the intrinsic atomic orbital charges, rather than Mulliken charges which change erratically depending on the basis set used [45]. IBOs are therefore always well-defined, whereas other localization methods - such as Pipek-Mezey orbitals [46], which depend on the Mulliken charges [47] - are unphysically tied to the basis set used [45]. For the IBO approach, we calculate the percentage of the $i$-th LMO over atoms a user defines as the active subsystem. Any LMO that has a percentage higher than a threshold (here 95 %) was assigned to the active region. The SPADE approach does not require this predefined threshold hyper-parameter. However, it does use a function of the molecular orbital coefficient matrix. Further details on both approaches are given in the Supplementary Material.

The effect of different localisation methods for this embedding method, such as: Pipek-Mezey (PM) [46], Foster-Boys (FB) [48], Edmiston-Ruedenberg (ER)[49] and fourth moment (FM) localization [50], would be an interesting area to explore. Our software package Nbed can run any method given by PySCF, and users can also build their own localisation strategies themselves [51].

## III. METHOD

We studied the performance of our wave function projection based embedding method on a selected set of molecular systems. We have developed a python package, Nbed, that utilizes the PySCF and Openfermion quantum chemistry packages to build each embedded model [52, 53]. The package outputs a qubit Hamiltonian for the wave function portion of an embedded problem and the classical energy corrections from density functional theory. This is freely available for use on GitHub [51].

For all calculations presented, the minimal *STO-3G* basis set was employed. Each global DFT calculation performed, prior to orbital localisation, used the *B3LYP* functional. The Intrinsic Bonding Orbitals (IBO) or Subsystem Projected AO DEcomposition (SPADE) localisation procedures are used in order to isolate the molecular orbitals to the active and environment subsystem from pre-selected active atoms [33, 45]. For the IBO procedure, in order to assign the active and environment molecular orbitals, we calculate the percentage of the $i_{th}$ LMO over atoms a user defines as the active subsystem. Any LMO that has a percentage higher than a predefined threshold (we used 95 %) is assigned to the active region. This paper's Supplementary Material goes into further detail on each localisation strategy. We performed both the $\mu$-shift and Huzinaga methods for each. A Hartree-Fock calculation for the active system, using the modified core Hamiltonian (equation 16b), was performed for each molecular system. The second quantized molecular Hamiltonian was then constructed with Openfermion and converted to a qubit Hamiltonian using the in-built Jordan-Wigner transformation [54]. Post Hartree-Fock methods were performed with PySCF. The frozen core approximation was not used and all virtual orbitals were included in the wave function calculations. Only the occupied environment molecular orbitals were removed from the wave function calculations of the active systems. To achieve this, the columns of $\boldsymbol{C}$ (the matrix of MO coefficients) associated with the environment were masked and not considered by further post Hartree-Fock (HF) methods on the embedded active system (the WF part of a WF-in-DFT calculation). The removal of these (occupied) environment orbitals is what gives a qubit reduction (when constructing the second quantized molecular Hamiltonian). We note that this

removal approach is slightly different to the implementation of Goodpaster and coworkers [41], where these orbitals are left in the embedded calculation. Our approach is justified as the environment MOs have been projected out of the SCF problem. More qualitatively, in Goodpaster's approach the results of the Huzinaga and $\mu$-shifted approaches are very similar [41]. For the $\mu$-shifted approach, the occupied environment orbitals are shifted to such high energies that they remain unoccupied in subsequent post-HF calculations. This is not the case for the Huzinaga method but, as it gives similar results to the $\mu$-shifted technique, removing the associated (occupied environment) orbitals of the Huzinaga method follows inline with them not being able to be occupied in the $\mu$-shifted approach.

For the single point electronic structure calculations we perform a CCSD-in-DFT calculation (active subsystem treated at CCSD level). Each result is then compared to full system CCSD(T) calculations. Each molecular geometry was obtained from PubChem [55]. The potential energy surface of a OH bond stretching in water was calculated using FCI-in-DFT, where the embedded molecular Hamiltonian at each geometry was diagonalized to find the ground state energy of the active system. This was compared to a full system FCI calculation at each step. As the PBE model requires a full system DFT calculation to determine the active and environment subsystems, we also report these DFT results.

## IV. RESULTS AND DISCUSSION

In the following subsections, we apply our model to different molecular systems.

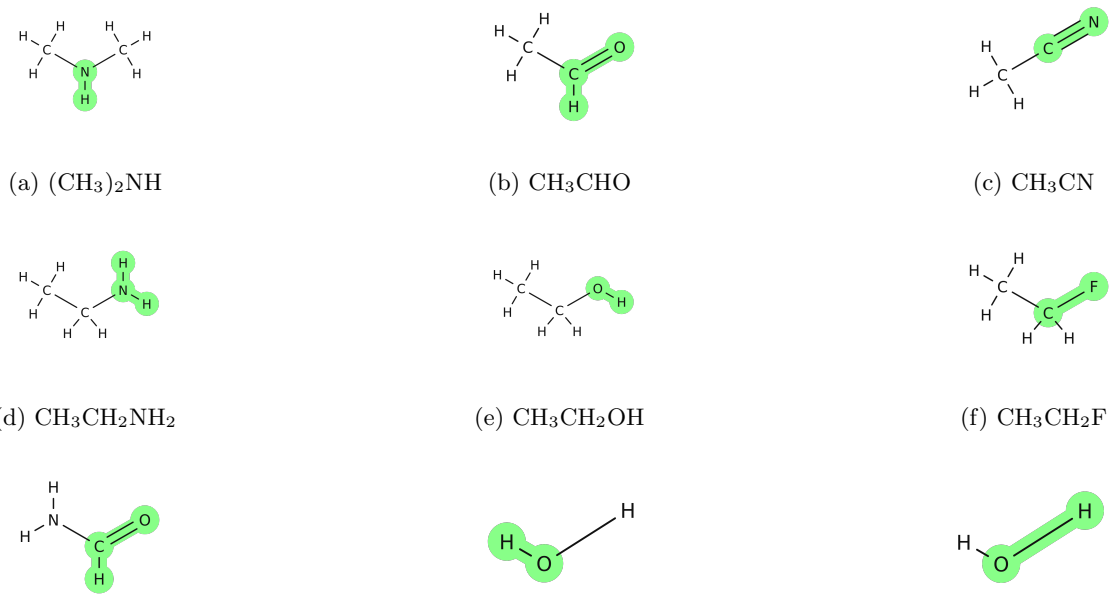
### A. Molecular Ground State Energy

In order to assess the ability of the embedding procedure, we selected a test bed of molecular structures - which are summarised in Figure 1. The active atoms considered at a more expensive level of theory are highlighted in green. The choice of these molecules was motivated by selecting compounds commonly encountered by chemists. To date, most quantum computing studies consider only the smallest molecular systems (often $H_2$, LiH, $BeH_2$) [17, 57, 58], due to current quantum computing constraints - low numbers of qubits numbers and high error rates. The goal of this paper is to show this embedding approach will allow larger systems to be studied on such devices. Figure 2 reports that the results for SPADE localized CCSD-in-DFT embedding molecular ground state energy calculations for the molecules in Figure 1. Numerical values of these results are available in this paper's Supplementary Material. The results for the same calculations using IBO localized orbitals can also be found in the Supplementary Material.

The embedded FCI Hamiltonians (describing the active region) output using both localisation methods were significantly reduced in number of terms and qubit counts compared to the full system FCI Hamiltonian. However, they still exceed the limit of what is practical to exactly solve using classical computers. We therefore performed classical CCSD-in-DFT calculations, the results of which are given in Figure 1(a). Our results show increased accuracy in CCSD-in-DFT calculated molecular ground state energies, compared to full system DFT. We benchmarked these approaches compared to full system CCSD(T), as full system FCI was not possible. The following metric was used $|\Delta E| = |E_{exp} - E_{CCSD(T)}|$ to approximate the true error $|\Delta E_{true}| = |E_{exp} - E_{FCI}|$, where $E_{exp}$ is the ground state energy calculated calculated via different procedures as specified and $E_{CCSD(T)}$ is a full system CCSD(T) reference ground state energy that is used to approximate each (full system) FCI ground state energy $E_{FCI}$. Our results show that PBE embedding gives ground state energies closer to the reference value, namely the full system CCSD(T) energy. It is known that different DFT functionals will give different ground state densities and thus energies [59]; however, there will always be a true ground state wavefunction and energy. The results in Figure 2 show that we can improve on the results of a DFT calculation by getting closer to the ground state energy - which in our case we approximate with CCSD(T). Qualitatively, we attribute the improvement of our PBE method to it including different correlation effects not captured by the DFT calculation. Further evidence of this will be seen in the next section on strong correlation.

Typically, results for the $\mu$-shift and Huzinaga projectors are very similar. This is expected as the active and environment subsystems were described in the supermolecular basis in our implementation of PBE [37, 41]. If the subsystems were described in an alternate basis, it has been shown that the Huzinaga operator outperforms the $\mu$-shift approach [41]. In our results, the Huzinaga projector usually produces marginally more accurate energies compared to the $\mu$-shifted implementation. We attribute this to the Huzinaga approach being based on a formally exact embedding, while the $\mu$-shift embedding is approximate due to a finite shift value being used.

The number of qubits describing the embedded FCI Hamiltonian (of a FCI-in-DFT problem) will be the same between the embedding methods - Figure 1b. This is due to the number of qubits depending on how many spin orbitals are considered in the embedded active system. In this work, only the occupied molecular orbitals of the environment are removed from the embedded active subsystem WF calculations (which leads to a qubit reduction). To further reduce the qubit count, virtual (unoccupied) molecular orbitals should also be removed from the embedded active subsystem calculations. This requires partitioning of the virtual space between the active and environment subsystems. Recent work by Yuan and coworkers, shows that truncating the virtual space can still give



(a) $(CH_3)_2NH$ (b) $CH_3CHO$ (c) $CH_3CN$

(d) $CH_3CH_2NH_2$ (e) $CH_3CH_2OH$ (f) $CH_3CH_2F$

(g) $HCONH_2$ (h) $H_2O$ (i) $H_2O$

FIG. 1: Planar representations of the molecules used in embedding calculations. Atoms shaded in green were selected as active for localisation procedures. Images were generated using MolView [56]. 1a N-methylmethanamine; 1b acetaldehyde; 1c acetonitrile; 1d ethanamine; 1e ethanol; 1f flouroethane; 1g formamide; 1h water (fixed bond active); 1i water (stretching bond active)

reliable estimates of both energies and molecular properties and we anticipate that this result will also be found if virtual environment orbitals are removed from the active embedded subsystem [60]. We leave this to future work, but note that it could lead to a significant further reduction in the number of qubits.

The number of terms in the Jordan-Wigner encoded qubit FCI Hamiltonian of the embedded WF problem (FCI-in-DFT problem), $|H|$, is typically very similar between the two projection methods, as shown in Figure 1c. This is expected as the number of molecular orbitals used to describe the embedded problem is the same between the different projection methods.

In comparing the two localisation methods, we find that for acetonitrile and formamide, SPADE and IBO partition the active system in a similar way. This results in a similar number of active MOs and hence the ground state estimation and resource requirements are very similar for these systems. For the majority of the molecules we study, SPADE includes more MOs, resulting in significantly more accurate ground state energies while still reducing the size of the Hamiltonian. However, by reducing the threshold for assigning the localized MOs from IBO to the active region, additional MOs could be included giving a similar result. See the Supplementary Material for further details.

### B. Strong Correlation

The impact of active region selection is demonstrated by our results shown in Figure 3, where SPADE localization was used in the embedding calculations (results for IBO localization are provided in the Supplementary Material). We consider the bond dissociation of an OH bond in water, where at high bond lengths, a correlated state is created [61, 62]. We perform projection based FCI-in-DFT calculations, at different molecular geometries, for two different active regions. One has the atoms in the fixed OH bond set active and the other has the atoms in the changing OH bond set active. We show this pictorially in Fig. 1h and 1i. These results are compared to full system FCI calculations.

At near equilibrium bond lengths, we see a similar performance between the different active systems (Figure 3). This is due to the symmetrical structure of $H_2O$, hence at low bond lengths there is little difference between the two active regions. In fact, the third data point gives results for the scenario where both OH bonds are the same length and consequently is why the results for the different active regions are the same here. However, in the correlated regime - at large bond lengths - selecting the active region to encompass the stretched atoms leads to significant improvements in energy calculation over DFT alone. This is due to correlation being effectively captured in the wave function calculation. In

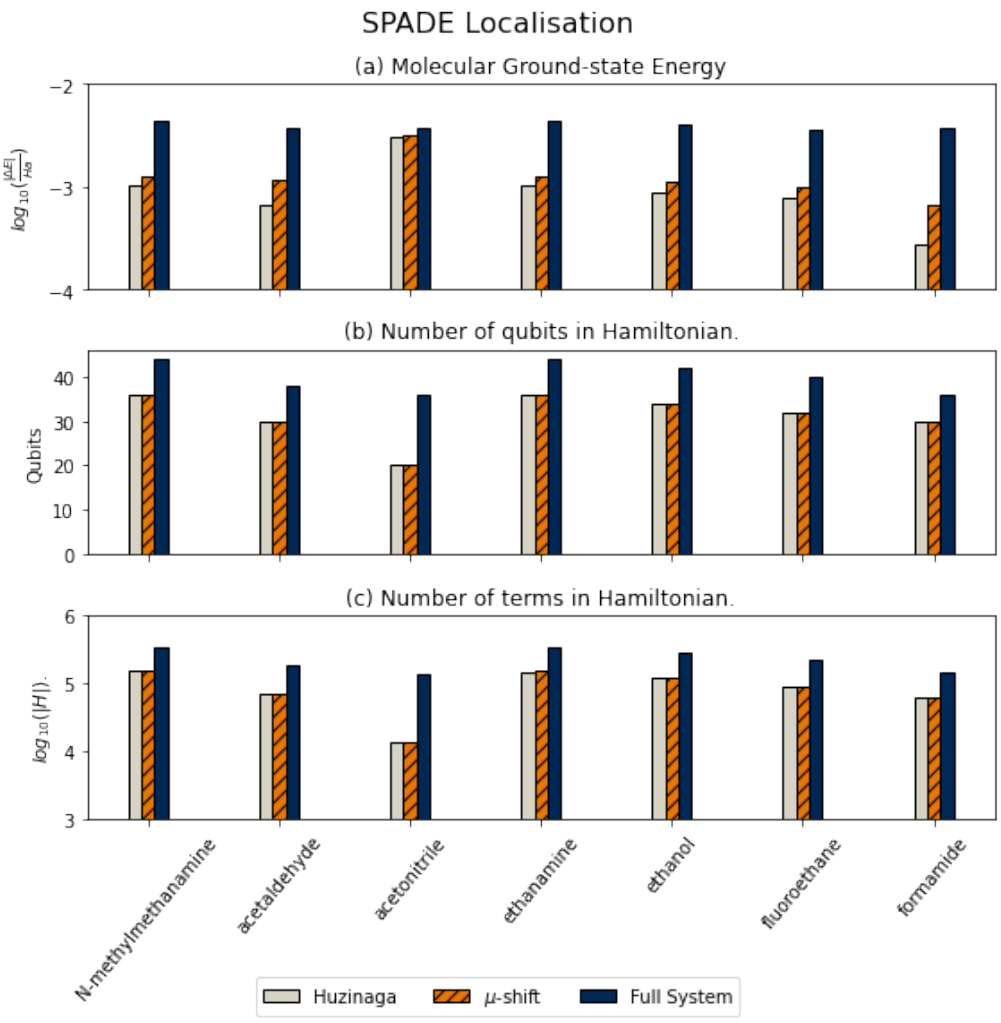



FIG. 2: Results for embedding of small molecules (Figure 1) using the SPADE localisation method. Bar chart (a) reports the ground state energy error for small molecules compared to full system CCSD(T): $|\Delta E| = |E_{exp} - E_{CCSD(T)}|$. For the blue result $E_{exp}$ is the full-system DFT ($B3LYP$) ground state energy, for the orange result $E_{exp}$ gives the $\mu$-shift CCSD-in-DFT embedding energy and for the grey result $E_{exp}$ is the Huzinaga CCSD-in-DFT embedding energy. Plot (b) shows the number of qubits needed to describe the embedded Hamiltonian, with the reference showing the number required for the full system FCI Hamiltonian. Plot (c) reports the number of terms in the Jordan-Wigner encoded embedded FCI-in-DFT Hamiltonian for each molecule with the blue bar representing the number of terms in the FCI Hamiltonian of the full system.

contrast, the full DFT calculation is plagued by deficiencies of current approximate exchange-correlation functionals [63, 64]. We see in Figure 3 that the global DFT calculation overestimates the bond dissociation energy. This problem is attributed to static correlation [63]. As there is no systematic way to improve the approximate exchange-correlation functionals, the way forward to describe such systems may be hybrid quantum-classical embedding. Here quantum processors could be exploited most effectively by application to only those regions of a molecule that are highly correlated. Finally, we note that all the DFT calculations performed in this work were restricted. At large bond lengths, it is known that spin-unrestricted calculations can often better describe bond dissociation; however, the major drawback of the method is spurious symmetry breaking [65]. Bearing this in mind, our results show that WF embedding combined with a restricted calculation can still capture major correlation effects.

## V. CONCLUSIONS

For a small collection of molecules, too large to study completely (full system) on currently available quantum hardware, we have shown that the PBE method allows a smaller active system to be studied using less resources on a quantum computer and the calculated energies of

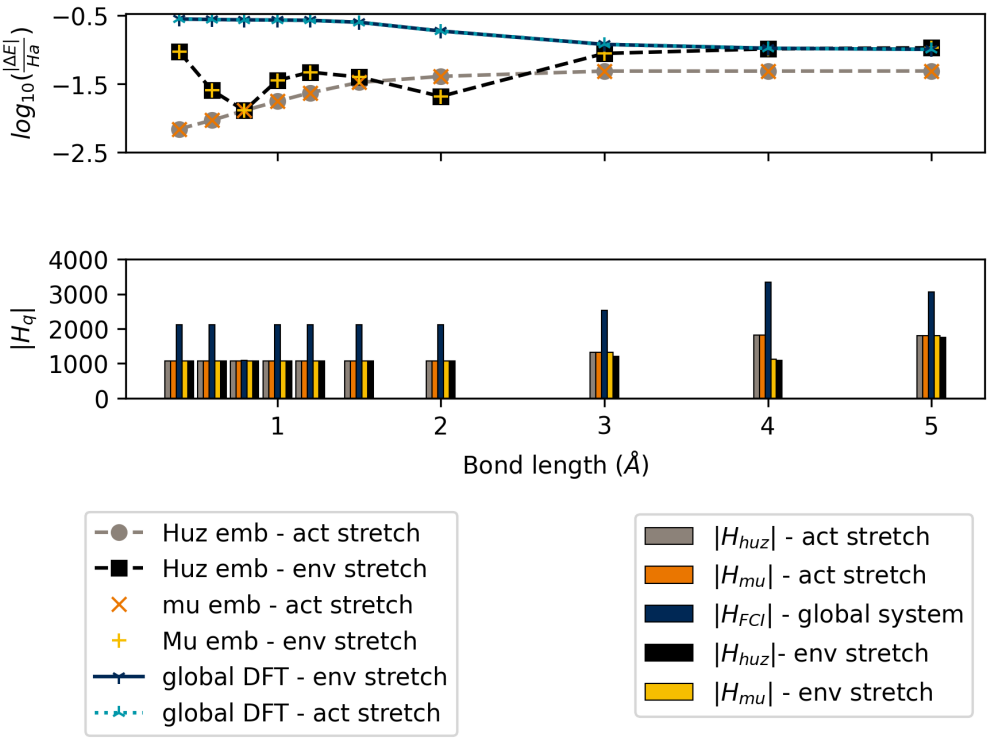



FIG. 3: Potential energy curve for $H_2O$, with changing OH bond length. *Active stretch* result has the changing OH bond as the active region and *environment stretch* result has the fixed OH bond selected as the active region. These results use SPADE localization. For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being four in all cases. The top plot reports the *log* base 10 error with respect to the exact FCI ground state energy ($E_{FCI}$) of the whole system, where $|\Delta E| = |E_{exp} - E_{FCI}|$. Here $E_{exp}$ is obtained from an FCI-in-DFT calculation. The bottom plot reports the number of terms in each Jordan-Wigner encoded qubit Hamiltonian. The blue result gives the size of the full system Hamiltonian, the orange and yellow results are for $\mu$-shifted embedded Hamiltonians while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are available in this paper's Supplementary Material.

such embedding calculations are closer to the "gold standard" CCSD(T) of the full system compared to full system DFT. Furthermore, we have shown its ability to capture the effects of strong correlation by investigating the bond dissociation of $H_2O$.

We use the projection-based embedding technique [31] to reduce the size of an electronic structure calculation studied at the wave function level. The molecular problem is split into active and environment parts, each solved using different levels of theory. The active part is treated using a wave function approach and an embedded qubit Hamiltonian is generated. Solving this provides $E^{WF}_{act} = \langle \Psi^{act}_{emb} | \mathbf{H}_{emb} | \Psi^{act}_{emb} \rangle$. The whole system and environment are treated using density functional theory and the overall electronic energy is found via an additive procedure [41, 43, 44]. This is similar to the own n-layered integrated molecular orbital and molecular mechanics (ONIOM) subtractive framework [66]. What is included in the active region can be modified and thus the size of the quantum problem varied. This allows users to tune their problem to available hardware and so it is possible to simulate large molecular problems on small quantum devices.

As this approach generates an embedded qubit Hamiltonian, it is agnostic to the quantum algorithm used to solve $\mathbf{H}_{emb}$. NISQ friendly approaches such as the VQE algorithm can therefore be used, but also fault-tolerant methods such as quantum phase estimation (QPE) [12, 17].

Moreover, as our method outputs a qubit Hamiltonian, different resource reduction techniques can be used in conjunction with it; for example, the contextual-subspace approach of Kirby and coworkers, [67] or the entanglement forging approach of Eddins [62]. Similarly, the $\mathbb{Z}_2$-symmetries of the problem can also be removed via qubit tapering [68].

As our method does not rely on imposing constraints on the system studied or costly parameter fitting, it may be reasonably combined with other hybridisation techniques which do[20, 69].

Further work is planned to develop this method. As significant resource reduction is achieved by localisation

of only the occupied orbitals, virtual orbital localisation could lead to a greater reduction in computational resources [70]. In the context of this work, if virtual LMOs are included in the active and environment subsystems respectively, then the number of qubits will reduce by how many are included in the environment. This will also have the effect of decreasing the total number of Pauli operators in the associated embedded qubit Hamiltonian.

We anticipate that our code will allow researchers to study molecules of real chemical interest on quantum computers. We welcome readers to make use of this, which is freely available on GitHub [51].

## ACKNOWLEDGMENTS

A. R. and M. I. W. acknowledge support from the Engineering and Physical Sciences Research Council (EPSRC) (EP/L015242/1 and EP/S021582/1 respectively). M.I.W. also acknowledges support from CBKSciCon Ltd. P. V. C. is grateful for funding from the European Commission for VECMA (800925) and EPSRC for SEAVEA (EP/W007711/1). We would like to thank Prof. Dieter Kranzlmüller at the Leibniz Supercomputing Centre (LRZ), who provided access to their ATOS Quantum Learning Machine simulator for some of the computations. The authors would also like to thank Dr. David A. Herrera-Martí for useful preliminary discussions on embedding. A.R. and M.I.W. contributed equally to this work.

[1] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).
[2] B. A. Cordier, N. P. Sawaya, G. G. Guerreschi, and S. K. McWeeney, Biology and medicine in the landscape of quantum advantages, arXiv preprint arXiv:2112.00760 (2021).
[3] H.-P. Cheng, E. Deumens, J. K. Freericks, C. Li, and B. A. Sanders, Application of quantum computing to biochemical systems: A look to the future, Frontiers in Chemistry **8**, 10.3389/fchem.2020.587143 (2020).
[4] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Reviews of Modern Physics **92**, 10.1103/RevModPhys.92.015003 (2020).
[5] Q. Sun, G. Kin, and L. Chan, Quantum embedding theories, Acc. Chem. Res **29**, 43 (2016).
[6] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum algorithms for quantum chemistry and quantum materials science, Chemical Reviews **120**, 12685 (2020), pMID: 33090772, https://doi.org/10.1021/acs.chemrev.9b00829.
[7] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed. (Dover Publications Inc., New York, 2012).
[8] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, Reviews of Modern Physics **79**, 291 (2007).
[9] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, Quantum Science and Technology **4**, 014008 (2018).
[10] G. D. Purvis III and R. J. Bartlett, A full coupled-cluster singles and doubles model: The inclusion of disconnected triples, Journal of Chemical Physics **76**, 1910 (1982).
[11] J. M. Herbert, Fantasy versus reality in fragment-based quantum chemistry, Journal of Chemical Physics **151**, 10.1063/1.5126216 (2019).
[12] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, Science **309**, 1704 (2005).
[13] A. Y. Kitaev, Quantum measurements and the abelian stabilizer problem, arXiv preprint quant-ph/9511026 (1995).
[14] P. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, *et al.*, Scalable quantum simulation of molecular energies, Physical Review X **6**, 031007 (2016).
[15] H. Mohammadbagherpoor, Y.-H. Oh, A. Singh, X. Yu, and A. J. Rindos, Experimental challenges of implementing quantum phase estimation algorithms on ibm quantum computer, arXiv preprint arXiv:1903.07605 (2019).
[16] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar, P. Helms, J. Gray, Z.-H. Cui, W. Liu, *et al.*, Is there evidence for exponential quantum advantage in quantum chemistry?, arXiv preprint arXiv:2208.02199 (2022).
[17] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, A variational eigenvalue solver on a photonic quantum processor, Nature Communications **5**, 1 (2014).
[18] G. A. Quantum, Collaborators*†, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, *et al.*, Hartree-fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).
[19] L. Zhao, J. Goings, K. Wright, J. Nguyen, J. Kim, S. Johri, K. Shin, W. Kyoung, J. I. Fuks, J.-K. K. Rhee, *et al.*, Orbital-optimized pair-correlated electron simulations on trapped-ion quantum computers, arXiv preprint arXiv:2212.02482 (2022).
[20] M. Rossmannek, P. K. Barkoutsos, P. J. Ollitrault, and I. Tavernelli, Quantum HF/DFT-embedding algorithms for electronic structure calculations: Scaling up to complex molecular systems, Journal of Chemical Physics **154**, 114105 (2021), https://doi.org/10.1063/5.0029536.
[21] H. Ma, M. Govoni, and G. Galli, Quantum simulations of materials on near-term quantum computers, npj Computational Materials **6**, 85 (2020).
[22] G. Knizia and G. K.-L. Chan, Density matrix embedding: A simple alternative to dynamical mean-field theory, Physical Review Letters **109**, 186404 (2012).
[23] N. C. Rubin, A hybrid classical/quantum approach for large-scale studies of quantum systems with density ma-

trix embedding theory, arXiv preprint arXiv:1610.06910 (2016).
[24] T. Yamazaki, S. Matsuura, A. Narimani, A. Saidmuradov, and A. Zaribafiyan, Towards the practical application of near-term quantum computers in quantum chemistry simulations: A problem decomposition approach, arXiv preprint arXiv:1806.01305 (2018).
[25] B. Bauer, D. Wecker, A. J. Millis, M. B. Hastings, and M. Troyer, Hybrid quantum-classical approach to correlated materials, Physical Review X **6**, 031045 (2016).
[26] J. M. Kreula, L. García-Álvarez, L. Lamata, S. R. Clark, E. Solano, and D. Jaksch, Few-qubit quantum-classical simulation of strongly correlated lattice fermions, EPJ Quantum Technology **3**, 11 (2016).
[27] T. Steckmann, T. Keen, A. F. Kemper, E. F. Dumitrescu, and Y. Wang, Simulating the mott transition on a noisy digital quantum computer via cartan-based fast-forwarding circuits, arXiv preprint arXiv:2112.05688 (2021).
[28] N. C. Rubin, A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory, arXiv preprint arXiv:1610.06910 (2016).
[29] H. Ma, M. Govoni, and G. Galli, Quantum simulations of materials on near-term quantum computers, npj Computational Materials **6**, 1 (2020).
[30] L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, Embedding methods for quantum chemistry: applications from materials to life sciences, Journal of the American Chemical Society **142**, 3281 (2020).
[31] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller, A simple, exact density-functional-theory embedding scheme, Journal of Chemical Theory and Computation **8**, 2564 (2012), pMID: 22904692.
[32] S. Lehtola and H. Jónsson, Unitary optimization of localized molecular orbitals, Journal of chemical theory and computation **9**, 5365 (2013).
[33] D. Claudino and N. J. Mayhall, Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories, Journal of Chemical Theory and Computation 10.1021/acs.jctc.8b01112 (2019).
[34] S. J. R. Lee, M. Welborn, F. R. Manby, and T. F. Miller, Projection-based wavefunction-in-DFT embedding, Accounts of Chemical Research **52**, 1359 (2019).
[35] O. Roncero, M. de Lara-Castells, P. Villarreal, F. Flores, J. Ortega, M. Paniagua, and A. Aguado, An inversion technique for the calculation of embedding potentials, Journal of Chemical Physics **129**, 184104 (2008).
[36] T. A. Barnes, J. D. Goodpaster, F. R. Manby, and T. F. Miller III, Accurate basis set truncation for wavefunction embedding, Journal of Chemical Physics **139**, 024103 (2013).
[37] B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay, Exact density functional and wave function embedding schemes based on orbital localization, Journal of Chemical Physics **145**, 064107 (2016).
[38] S. Huzinaga and A. A. Cantu, Theory of separability of many-electron systems, Journal of Chemical Physics **55**, 5543 (1971).
[39] E. Francisco, A. Martín Pendás, and W. Adams, Generalized huzinaga building-block equations for nonorthogonal electronic groups: Relation to the Adams–Gilbert theory, Journal of chemical physics **97**, 6504 (1992).
[40] T. Shimazaki, K. Kitaura, D. G. Fedorov, and T. Nakajima, Group molecular orbital approach to solve the huzinaga subsystem self-consistent-field equations, Journal of Chemical Physics **146**, 10.1063/1.4976646 (2017).
[41] D. V. Chulhai and J. D. Goodpaster, Improved accuracy and efficiency in quantum embedding through absolute localization, Journal of Chemical Theory and Computation **13**, 1503 (2017).
[42] D. V. Chulhai and J. D. Goodpaster, Projection-based correlated wave function in density functional theory embedding for periodic systems, Journal of chemical theory and computation **14**, 1928 (2018).
[43] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, Accurate and systematically improvable density functional theory embedding for correlated wavefunctions, Journal of chemical physics **140**, 18A507 (2014).
[44] D. S. Graham, X. Wen, D. V. Chulhai, and J. D. Goodpaster, Huzinaga projection embedding for efficient and accurate energies of systems with localized spin-densities, Journal of Chemical Physics **156**, 054112 (2022).
[45] G. Knizia, Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts, Journal of Chemical Theory and Computation **9**, 4834 (2013).
[46] J. Pipek and P. G. Mezey, A fast intrinsic localization procedure applicable for *ab initio* and semi-empirical linear combination of atomic orbital wave functions, Journal of Chemical Physics **90**, 4916 (1989).
[47] R. S. Mulliken, Electronic population analysis on lcao–mo molecular wave functions. i, Journal of Chemical Physics **23**, 1833 (1955), https://doi.org/10.1063/1.1740588.
[48] J. Foster and S. Boys, Canonical configurational interaction procedure, Reviews of Modern Physics **32**, 300 (1960).
[49] C. Edmiston and K. Ruedenberg, Localized atomic and molecular orbitals, Reviews of Modern Physics **35**, 457 (1963).
[50] I.-M. Høyvik, B. Jansik, and P. Jørgensen, Orbital localization using fourth central moment minimization, Journal of Chemical Physics **137**, 224114 (2012).
[51] M. Williams and A. Ralli, Nbed, `https://github.com/UCL-CCS/Nbed` (2022).
[52] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, Pyscf: the python-based simulations of chemistry framework, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).
[53] J. R. McClean, K. J. Sung, I. D. Kivlichan, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. Rubin, N. P. D. Sawaya, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush, Openfermion: The electronic structure package for quantum computers, Quantum Science and Technology **5** (2017).
[54] P. Jordan and E. Wigner, Uber das Paulische Aquivalenzverbot (1928), Zeitschrift für Physik **47**, 631 (1928).
[55] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, Pubchem in 2021:

new data content and improved web interfaces, Nucleic Acids Research **49**, D1388 (2021).

[56] H. Bergwerf, Molview, `https://molview.org/` (2022), accessed: 2022-02-03.

[57] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, *et al.*, Quantum chemistry calculations on a trapped-ion quantum simulator, Physical Review X **8**, 031022 (2018).

[58] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[59] A. D. Bochevarov and R. A. Friesner, The densities produced by the density functional theory: Comparison to full configuration interaction, The Journal of chemical physics **128**, 034102 (2008).

[60] X. Yuan, L. Visscher, and A. S. P. Gomes, Assessing mp2 frozen natural orbitals in relativistic correlated electronic structure calculations, The Journal of Chemical Physics **156**, 224108 (2022).

[61] I. O. Sokolov, P. K. Barkoutsos, P. J. Ollitrault, D. Greenberg, J. Rice, M. Pistoia, and I. Tavernelli, Quantum orbital-optimized unitary coupled cluster methods in the strongly correlated regime: Can quantum algorithms outperform their classical equivalents?, The Journal of chemical physics **152**, 124107 (2020).

[62] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, Doubling the size of quantum simulators by entanglement forging, PRX Quantum **3**, 010309 (2022).

[63] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Insights into current limitations of density functional theory, Science **321**, 792 (2008).

[64] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Challenges for density functional theory, Chemical Reviews **112**, 289 (2012).

[65] M. Fuchs, Y.-M. Niquet, X. Gonze, and K. Burke, Describing static correlation in bond dissociation by kohn–sham density functional theory, The Journal of Chemical Physics **122** (2005).

[66] M. Svensson, S. Humbel, R. D. Froese, T. Matsubara, S. Sieber, and K. Morokuma, Oniom: a multilayered integrated mo+ mm method for geometry optimizations and single point energy predictions. a test for diels- alder reactions and pt (p (t-bu) 3) 2+ h2 oxidative addition, The Journal of Physical Chemistry **100**, 19357 (1996).

[67] W. M. Kirby and P. J. Love, Classical simulation of noncontextual pauli hamiltonians, Physical Review A **102**, 032418 (2020).

[68] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic hamiltonians, arXiv preprint arXiv:1701.08213 (2017).

[69] J. P. Stenger, D. Gunlycke, and C. S. Hellberg, Expanding variational quantum eigensolvers to larger systems by dividing the calculations between classical and quantum hardware, Physical Review A **105**, 022438 (2022).

[70] D. Claudino and N. J. Mayhall, Simple and efficient truncation of virtual spaces in embedded wave functions via concentric localization, Journal of Chemical Theory and Computation **15**, 6085 (2019).

[71] T. Giovannini and H. Koch, Energy-based molecular orbital localization in a specific spatial region, Journal of Chemical Theory and Computation **17**, 139 (2020).

[72] S. Barison, D. E. Galli, and M. Motta, Quantum simulations of molecular systems with intrinsic atomic orbitals, Physical Review A **106**, 022404 (2022).