A Scalable Approach to Quantum Simulation via Projection-Based Embedding Supplementary Material

Alexis Ralli,[1,2] Michael Williams de la Bastida,[1,*] and Peter V. Coveney[1,3,4]

[1] *Centre for Computational Science, Department of Chemistry, University College London, WC1H 0AJ, United Kingdom*
[2] *Department of Physics and Astronomy, Tufts University, Medford, Massachusetts 02155, USA*
[3] *UCL Centre for Advanced Research Computing, Gower Street , London WC1E 6BT, United Kingdom*
[4] *Informatics Institute, University of Amsterdam, Amsterdam, 1098 XH, Netherlands* [†]

(Dated: September 22, 2023)

I. LOCALISATION

In this section, we describe how to generate a set of localized molecular orbitals (LMOs) from the cannoncial molecular orbitals (CMOs) of a global DFT calculation. This step is required in order to define each subsystem according to a certain metric discussed in further detail here.

We employ two methods to generate a set of localised molecular orbitals: the intrinsic bond orbitals (IBO) [1] and subsystem Projected AO DEcomposition (SPADE) [2]. Other methods can be used, such as Pipek-Mezey (PM) [3], Foster-Boys (FB) [4], Edmiston-Ruedenberg (ER)[5] and fourth moment (FM) localization [6].

A. SPADE

SPADE is designed to localise electrons to the general region of an active site and environment [7]. This avoids the need for a threshold used to define the active and environment systems, as needed by the PM, FB, ER, FM and IBO approaches. The next section goes into more details about this.

We first restrict the $\mathbf{C}$ matrix to only the occupied molecular orbitals $\{\psi_{occ}\} \subset \{\psi\}$. As our aim is to localise electrons to one region; virtual orbitals need not be localised. The remaining MOs are made mutually orthogonal by rotation with the matrix defining the overlap of atomic orbitals $\mathbf{S}$[7]:

$$\bar{\mathbf{C}} = \mathbf{S}^{0.5}\mathbf{C}_{occ}. \tag{1}$$

We wish to find the relative contribution of the active region AOs to the occupied MOs, so we now restrict the matrix to only contributions from AOs associated to atoms in the active region $\bar{C}_A$. By making a singular value decomposition of this matrix[7]:

$$\bar{\mathbf{C}}_A = \mathbf{U}_A\mathbf{\Sigma}_A\mathbf{V}_A^{\star}, \tag{2}$$

we can determine a basis which localises the electrons of the active region. The column vectors of $\mathbf{V}_A$ are identical to the eigenvectors of $\bar{\mathbf{C}}_{occ}^{A\dagger}\bar{\mathbf{C}}_{occ}^{A}$ [7]. A new set of orbitals are then obtained through a rotation of the original MOs $\mathbf{C}_{occ}$ as:

$$\mathbf{C}_{occ}^{SPADE} = \bar{\mathbf{C}}_{occ}\mathbf{V}_A, \tag{3}$$

where $\mathbf{C}_{occ}^{SPADE}$ are the SPADE orbitals. The singular values $\{\sigma\}$ of $\mathbf{\Sigma}_A$ allow appropriate partitioning of the orbital subspaces into active and environment subsystem [7]. The largest difference between successive singular values gives the partition of the SPADE MOs. Formally we write this as:

$$m_{occ}^{A} = \max_i\{\sigma_i - \sigma_{i+1}\}, \tag{4}$$

here m_{occ}^{A} is the index i where the largest difference occurs. The SPADE MOs (columns of $\mathbf{C}_{occ}^{SPADE}$) up to this index are the active MOs and the remaining orbitals (other columns of $\mathbf{C}_{occ}^{SPADE}$) are the environment orbitals.

B. Intrinsic Bonding Orbitals

Intrinsic Bonding Orbitals (IBOs) are constructed as a linear combination of intrinsic atomic orbitals (IAOs), where the the number of atoms the orbital charge is spread over is minimized.

Taking the canonical molecular orbitals defined as:

$$|\psi_i\rangle = \sum_j^K \mathbf{C}_{ji}\,|\phi_j\rangle\,, \tag{5}$$

where $|\phi_j\rangle \in B_1$ are basis functions from a basis set B_1. A key problem with this form is each MO $|\psi_i\rangle$ is hard to interpret, as each atomic orbital's (AO) basis functions $|\phi_j\rangle$ cannot be associated with a given atom [1]. Normally MOs are highly delocalized and each $|\phi_j\rangle$ will contribute where it is needed most. Thus, we would like to expand the MOs over another minimal basis B_2 of free-atom atomic orbitals for each atom, which we write as $\{\left|\tilde{\phi}\right\rangle\} \in B_2$. This would make the wavefunction easy to

* michael.williams.20@ucl.ac.uk
† p.v.coveney@ucl.ac.uk

interpret, but would be inaccurate and even incorrect, as free-atom AOs contain no polarization due to the molecular environment. The IAO method aims to combine the best properties of these.

First, a molecular SCF wavefunction $|\Phi\rangle$ (a single Slater determinant) is calculated that defines a set of molecular orbitals (MOs) $|\psi_i\rangle$. Then a set of polarized atomic orbitals $\{\left|\phi^{(IAO)}\right\rangle\} \notin B_2$ are formed that can express $|\Phi\rangle$s in the occupied MOs $|\psi_i\rangle$. To do this, projectors onto the occupied and virtual MOs are defined: $P = \sum_i |\psi_i\rangle \langle\psi_i|$ and $Q = I - P$ respectively [8]. The atomic orbital (AO) projectors onto the bases B_1 and B_2 are also defined as [1]:

$$\mathbf{P}_{12} = \sum_{|\phi\rangle_i, |\phi\rangle_j \in B_1} |\phi\rangle_i \, \mathbf{S}^{-1}_{ij} \, \langle\phi|_j \,, \tag{6a}$$

$$\mathbf{P}_{21} = \sum_{|\phi\rangle_k, |\phi\rangle_l \in B_2} |\phi\rangle_k \, \mathbf{S}^{-1}_{kl} \, \langle\phi|_l \,. \tag{6b}$$

Here $\mathbf{S}^{-1}_{ij}$ and $\mathbf{S}^{-1}_{kl}$ are the inverse overlap matrices in the bases B_1 and B_2. The set of depolarized MOs is then given by [1, 8]:

$$\{\left|\tilde{\psi}_i\right\rangle\} = \{\mathbf{P}_{12}\mathbf{P}_{21} \, |\psi_i\rangle\}, \tag{7}$$

These depolarized molecular orbitals are then used to define the projectors: $\tilde{\mathbf{P}} = \sum_i \left|\tilde{\psi}_i\right\rangle \left\langle\tilde{\psi}_i\right|$ and $\tilde{\mathbf{Q}} = \mathbf{I} - \tilde{\mathbf{P}}$. The IAOs are finally given by:

$$\left|\phi^{(IAO)}\right\rangle = \left(\mathbf{P}\tilde{\mathbf{P}} + \mathbf{Q}\tilde{\mathbf{Q}}\right)\mathbf{P}_{12} \left|\tilde{\phi}\right\rangle. \tag{8}$$

The IAO basis is then orthonormalized [1, 8]. Overall, each IAO is obtained by a simple set of matrix operations. The utility of IAOs stems from the fact that they are directly associated with atoms and can be used to define atomic properties like partial charges and are basis set independent, unlike the Pipek-Mezey approach.

Often we want to know about molecular bonding rather than atomic properties. Knizia showed that by combining IAOs with a localization in the spirit of Pipek-Mezey, one can obtain intrinsic bonding orbitals (IBOs). As discussed in the main text, a Slater determinant $|\Phi\rangle$ is invariant to unitary rotations $\left|\psi^{LMO}_i\right\rangle = |\psi_i\rangle \, \mathbf{U}$, we can thus define the IBOs by maximizing [1]:

$$L_{IBO} = \sum_{A}^{N_{atoms}} \sum_{i}^{N_{occ}} [n_A(i)]^4 \tag{9}$$

with respect to $\mathbf{U}$. Here $n_A(i)$ is the number of electrons from ψ^{LMO}_i located on the IAOs $\{\left|\phi^{(IAO)}\right\rangle\}$ of atom A. Explicitly, we write this as [1]:

$$n_A(i) = 2 \sum_{\left|\phi^{(IAO)}\right\rangle \in A} \langle\psi^{LMO}_i \left|\phi^{(IAO)}\right\rangle \left\langle\phi^{(IAO)}\right| \psi^{LMO}_i\rangle. \tag{10}$$

This construction minimizes the number of atoms upon which an orbital is centred [1].

In order to assign these localized molecular orbitals ($\mathbf{C}^{LMO}$) to the active and environment subsystems we calculate the percentage of the i_{th} LMO over the active atoms as:

$$p^{act}_i(\mathbf{C}) = \frac{\sum_{\nu \in act} \mathbf{C}^2_{i\nu}}{\sum_{\nu=1}^{K} \mathbf{C}^2_{i\nu}}, \tag{11}$$

where $\nu \in act$ are the atomic orbital indices for the atoms defined in the active region. The denominator includes all the AOs of the i-th MO. This is the approach given in Equation 10 in the work of Koch *et al.*[9]. Any $p^{act}_i > 95\%$ we associate to the active subsystem.

Our code on GitHub uses this metric for the other localization strategies suppored by PySCF. The code generates a localized $\mathbf{C}$ matrix, which can then be used in conjunction with Equation 11.

II. EMBEDDED SELF-CONSISTENT FIELD METHODOLOGY

In this section we summarise how the embedded self-consistent field calculations are performed for the μ-shift and Huzinaga methods. We consider restricted Hartree-Fock (RHF) calculations; extensions to the unrestricted case follow straightforwardly.

A. μ-shift RHF

To perform the μ-shifted RHF calculation, only the core Hamiltonian is modified (Equation 16b in the main text), by adding $\mathbf{V}_{emb} + \mathbf{P}^{env}_{proj}$ to it. The $\mathbf{g}(\gamma^{act} + \gamma^{env})$ term is obtained from the global DFT calculation. The $\mathbf{g}(\gamma^{act})$ potential energy matrix is calculated in the same way, except the density matrix is set to be that of the active system only. The μ-shift projector can then be defined using Equation 17 in the main text.

The standard RHF algorithm can then be run, where the only difference is that the standard core Hamiltonian has been modified to be $\mathbf{h}_{emb}$.

B. Huzinaga RHF

The Huzinaga RHF calculation is slightly more involved than the μ-shifted method. At each self-consistent field loop, when the new Fock matrix is defined, the Huzinaga projection operator $\mathbf{P}^{env}_{huz}$ must be built according to

Equation 20 in the main text. We reiterate this step uses the current Fock matrix in the self consistent field loop. The embedded Fock matrix is then constructed according to equation 15 of the main text, where the potential energy matrix $\mathbf{V}_{emb}^{proj}$ (Equation 16a in the main text) is defined in the same manor as the μ-shift RHF method. Given this embedded Fock matrix, $\mathbf{F}_{emb}^{act}\mathbf{C} = \mathbf{SC}\epsilon$ can be solved via standard SCF approaches. This generates a set of new MO coefficients $\mathbf{C}$ that are used to construct the new Fock matrix. This process is repeated until the energy converges as usual in an SCF calculation.

III. ACTIVE ATOM SELECTION

As the number of active atoms is configurable with this method, we demonstrate the effect of altering this parameter using cyclopentane. Figure 1 shows the change in calculated ground state energy, qubit count and number of terms in the Jordan-Wigner encoded qubit Hamiltonian.

IV. MOLECULAR GROUND STATE ENERGY

Our method provides flexibility to select a localisation method. In addition to the result displayed in Figure 2, which were calculated using the SPADE projection method, we present results for the same molecules using the Intrinsic Bonding Orbitals localisation method in Figure 2. Numerical values for these results are given in Table I for reference values, and Tables II and III for our calculated results.

Molecule	ϵ_{DFT}	Q	$\lvert H\rvert$
N-methylmethanamine	0.5733	44	338971
acetaldehyde	0.569	38	182702
acetonitrile	0.485	36	136067
ethanamine	0.573	44	329283
ethanol	0.609	42	283020
fluoroethane	0.637	40	217385
formamide	0.619	36	138235

TABLE I: Full-system reference values for embedding calculations of small molecules, as shown in Figures 2 and 2. ϵ_{DFT} gives the difference between full-system RKS DFT, using the B3LYP functional, and CCSD(T). Q and $|H|$ give respectively the number of qubits and terms in the Jordan-Wigner encoded qubit Hamiltonian of the full system. All energies reported in Hartree (Ha).

V. STRONG CORRELATION

We provide the numerical details of our strongly correlated H_2O study in this section, where SPADE localization has been used. These results form Figure 3 in the main text. We also include results for IBO localization. The raw data for all these results is provided in ancillary files with this text.

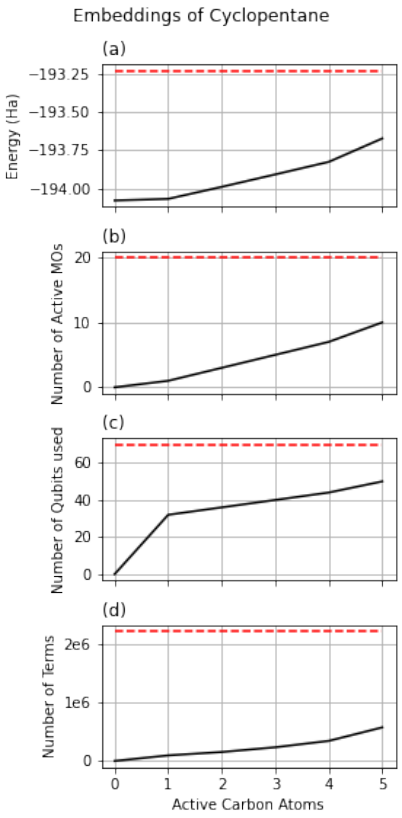

FIG. 1: Results of cyclopentane embedding, with increasing numbers of carbon atoms assigned to the active region. For each, IBO localisation was used; results for the μ-shift and Huzinaga projector are overlapping and indistinguishable. Different methods do not give noticeably distinct results. (a) Ground state energy. (b) Number of occupied molecular orbitals assigned to the active region. (c) Number of qubits in output Hamiltonian. (d) Number of terms in the Jordan-Wigner encoded qubit Hamiltonian.

For the H_2O projection based embedding calculations, at different molecular geometries, we considered two different active regions. One had the atoms in the fixed OH bond set active and the other had the atoms in the changing OH bond set active. The structure for H_2O with an OH bond length of 0.4 $\AA$ is given in Table IV. The other geometries can be generated from this structure. Tables V and VI summarise the numerical results for the different active systems where SPADE localization has been used. Tables VII and VIII give the numerical results for the different active systems when IBO localization was used and Figure 3 provides a graphical summary of these results.

For the FCI-in-DFT results given in Figure 3, we note that the error in the embedded calculation is actually

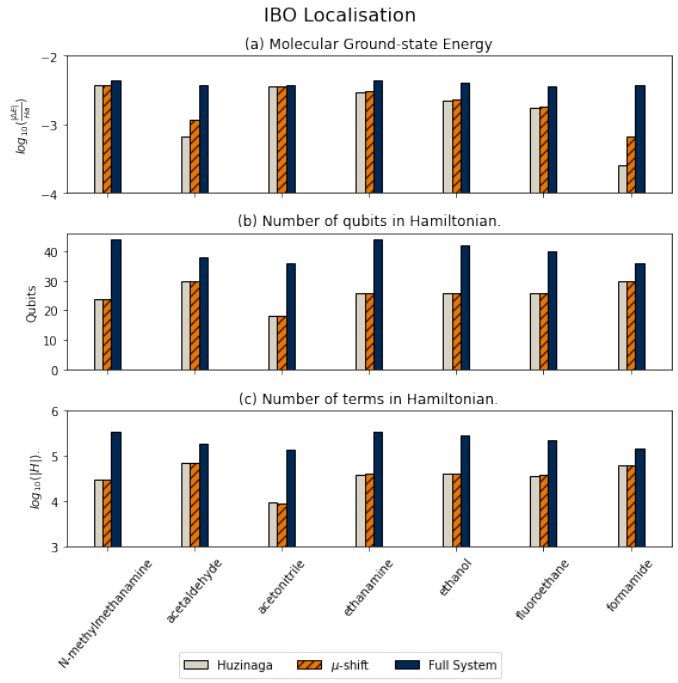

FIG. 2: Results for embedding of small molecules using the IBO localisation method. (a) Ground state energies error for small molecules, with full-system DFT energy error as reference, μ-shift CCSD-in-DFT embedding energy in orange and Huzinaga CCSD-in-DFT embedding in grey. All error values are calculated with respect to whole system CCSD(T) energies. (b) The number of qubits required to describe the embedded FCI-in-DFT Hamiltonians, with reference showing the number required for the full system FCI Hamiltonian. (c) The number of terms in the Jordan-Wigner encoded FCI-in-DFT qubit Hamiltonian for each molecule. Again the reference gives the number needed for the full system FCI Hamiltonian.

higher than the global DFT calculation at an OH bond length of 4.0 Å. There is also a significant difference in the number of qubits and Hamiltonian terms at OH bond lengths of 0.4, 0.6 and 4.0 Å. The variation, compared with the SPADE results (Figure 3 in the main text), is mainly due to the localization method giving different numbers of active MOs. We repeated the IBO calculation using a different active region threshold (minimum setting of 90%) and obtained similar results, as the number of active MOs then matched that of the SPADE calculation. Figure 4 summarises this result.

[1] G. Knizia, Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts, Journal of Chemical Theory and Computation **9**, 4834 (2013).

[2] 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.

[3] 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).

[4] J. Foster and S. Boys, Canonical configurational interaction procedure, Reviews of Modern Physics **32**, 300 (1960).

[5] C. Edmiston and K. Ruedenberg, Localized atomic and molecular orbitals, Reviews of Modern Physics **35**, 457 (1963).

[6] I.-M. Høyvik, B. Jansik, and P. Jørgensen, Orbital localization using fourth central moment minimization, Journal of Chemical Physics **137**, 224114 (2012).

[7] 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).

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

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

Molecule	ϵ_{huz}	ϵ_{μ}	Q_{huz}	Q_{μ}	$\lvert H^{huz}_{emb}\rvert$	$\lvert H^{\mu}_{emb}\rvert$
N-methylmethanamine	0.491	0.494	24	24	29701	29737
acetaldehyde	0.099	0.178	30	30	70118	69554
acetonitrile	0.462	0.462	18	18	9232	9016
ethanamine	0.394	0.399	26	26	38716	39924
ethanol	0.340	0.350	26	26	40948	41060
fluoroethane	0.299	0.312	26	26	35020	38180
formamide	0.043	0.109	30	30	62434	61914

TABLE II: Numerical values of the calculations shown in Figure 2. For each calculation the energy difference between CCSD(T)-in-DFT embedding and full system CCSD(T), ϵ, is given. The number of qubits Q and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. Results for IBO localisation are shown, with results for SPADE in Table III reference values using the full system given in Table I. All energies reported in Hartree (Ha).

Molecule	ϵ_{huz}	ϵ_{μ}	Q_{huz}	Q_{μ}	$\lvert H^{huz}_{emb}\rvert$	$\lvert H^{\mu}_{emb}\rvert$
N-methylmethanamine	0.135	0.169	36	36	152223	152415
acetaldehyde	0.098	0.176	30	30	70474	69982
acetonitrile	0.403	0.409	20	20	13439	13111
ethanamine	0.136	0.169	36	36	145067	149819
ethanol	0.132	0.169	34	34	120928	121036
fluoroethane	0.136	0.170	32	32	85781	89377
formamide	0.045	0.110	30	30	62366	61958

TABLE III: Numerical values of the calculations shown in Figure 2. For each calculation the energy difference between CCSD(T)-in-DFT embedding and full system CCSD(T), ϵ, is given. The number of qubits Q and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. Results for SPADE localisation are shown, with results for IBO in Table II reference values using the full system given in Table I. All energies reported in Hartree (Ha).

atom	x	y	z
H	0.3751747	0.0000000	0.1387225
O	0.0000000	0.0000000	0.0000000
H	-0.7493682	0.0000000	0.2770822

TABLE IV: Cartesian coordinates of atoms in H_2O for the structure with an OH bond length of 0.4 $\AA$ defined from the first H and O atoms in this Table. The other structures (different OH bond lengths) were generated from this file by changing the position of the first H atom. Note the H-O-H angle remained fixed.

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{\text{FCI-in-DFT}}^{huz}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$\lvert H_{full}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{huz}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{\mu}\rvert$	Q_{FCI}^{global}	$Q_{\text{FCI-in-DFT}}^{huz}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.988008	-72.988009	2110	1079	1079	14	12	12	4
0.600000	-74.499220	-74.773419	-74.508545	-74.508545	2110	1079	1079	14	12	12	4
0.798954	-74.851089	-75.122053	-74.864002	-74.864002	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.918225	-74.918226	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.890912	-74.890913	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.840739	-74.840739	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.816902	-74.816902	2110	1079	1079	14	12	12	4
3.000000	-74.771687	-74.890577	-74.820223	-74.820224	2530	1327	1327	14	12	12	4
4.000000	-74.771719	-74.876129	-74.820261	-74.820262	3346	1819	1819	14	12	12	4
5.000000	-74.771718	-74.872116	-74.820410	-74.820411	3054	1799	1799	14	12	12	4

TABLE V: Numerical values of the calculations shown in Figure 3 for the case when the changing OH bond is set as the active region and SPADE localization has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in Hartree (Ha).

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{\text{FCI-in-DFT}}^{huz}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$\lvert H_{full}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{huz}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{\mu}\rvert$	Q_{FCI}^{global}	$Q_{\text{FCI-in-DFT}}^{huz}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.887822	-72.887827	2110	1079	1079	14	12	12	4
0.600000	-74.499220	-74.773419	-74.473689	-74.473691	2110	1079	1079	14	12	12	4
0.798954	-74.851089	-75.122053	-74.864002	-74.864002	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.936101	-74.936101	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.914336	-74.914336	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.847328	-74.847328	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.755506	-74.755506	2110	1079	1079	14	12	12	4
3.000000	-74.771692	-74.890579	-74.683921	-74.683921	2406	1203	1327	14	12	12	4
4.000000	-74.771719	-74.876127	-74.669427	-74.669427	2598	1083	1119	14	12	12	4
5.000000	-74.771718	-74.872381	-74.665626	-74.665626	3046	1759	1799	14	12	12	4

TABLE VI: Numerical values of the calculations shown in Figure 3 for the case when the fixed OH bond is set as the active region and SPADE localization has been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in Hartree (Ha).

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{\text{FCI-in-DFT}}^{huz}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$\lvert H_{full}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{huz}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{\mu}\rvert$	Q_{FCI}^{global}	$Q_{\text{FCI-in-DFT}}^{huz}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.928542	-72.923051	2110	492	492	14	10	10	3
0.600000	-74.499220	-74.773419	-74.484545	-74.484551	2110	492	492	14	10	10	3
0.798954	-74.851089	-75.122053	-74.852614	-74.852616	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.899540	-74.899543	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.868096	-74.868099	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.819776	-74.819778	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.798952	-74.798954	2110	1079	1079	14	12	12	4
3.000000	-74.771692	-74.890578	-74.804924	-74.804927	2238	1379	1383	14	12	12	4
4.000000	-74.771719	-74.873986	-74.470749	-74.470821	2342	876	876	14	10	10	3
5.000000	-74.771718	-74.872112	-74.805202	-74.805204	3210	1779	1775	14	12	12	4

TABLE VII: Numerical values of the calculations when the changing OH bond is set as the active region and IBO localization (95% threshold)is been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in Hartree (Ha).

OH length	E_{FCI}^{global}	E_{DFT}^{global}	$E_{\text{FCI-in-DFT}}^{huz}$	$E_{\text{FCI-in-DFT}}^{\mu}$	$\lvert H_{full}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{huz}\rvert$	$\lvert H_{\text{FCI-in-DFT}}^{\mu}\rvert$	Q_{FCI}^{global}	$Q_{\text{FCI-in-DFT}}^{huz}$	$Q_{\text{FCI-in-DFT}}^{\mu}$	# active MOs
0.400000	-72.981056	-73.259936	-72.928542	-72.923051	2110	492	492	14	10	10	3
0.600000	-74.499220	-74.773419	-74.484545	-74.484551	2110	492	492	14	10	10	3
0.798954	-74.851089	-75.122053	-74.852614	-74.852616	1086	1079	1079	14	12	12	4
1.000000	-74.900658	-75.170068	-74.899540	-74.899543	2110	1079	1079	14	12	12	4
1.200000	-74.867498	-75.134418	-74.868096	-74.868099	2110	1079	1079	14	12	12	4
1.500000	-74.807539	-75.057384	-74.819776	-74.819778	2110	1079	1079	14	12	12	4
2.000000	-74.776263	-74.962535	-74.798952	-74.798954	2110	1079	1079	14	12	12	4
3.000000	-74.771692	-74.890578	-74.804924	-74.804927	2238	1379	1383	14	12	12	4
4.000000	-74.771719	-74.873986	-74.470749	-74.470821	2342	876	876	14	10	10	3
5.000000	-74.771718	-74.872112	-74.805202	-74.805204	3210	1779	1775	14	12	12	4

TABLE VIII: Numerical values of the calculations when the fixed OH bond is set as the active region and IBO localization (95% threshold) is been used. For each calculation the absolute energy is reported (each Hamiltonian was diagonalized to give the exact ground state). The number of qubits Q and number of terms in the output Jordan-Wigner encoded qubit Hamiltonian $|H|$ are given. All energies reported in Hartree (Ha).

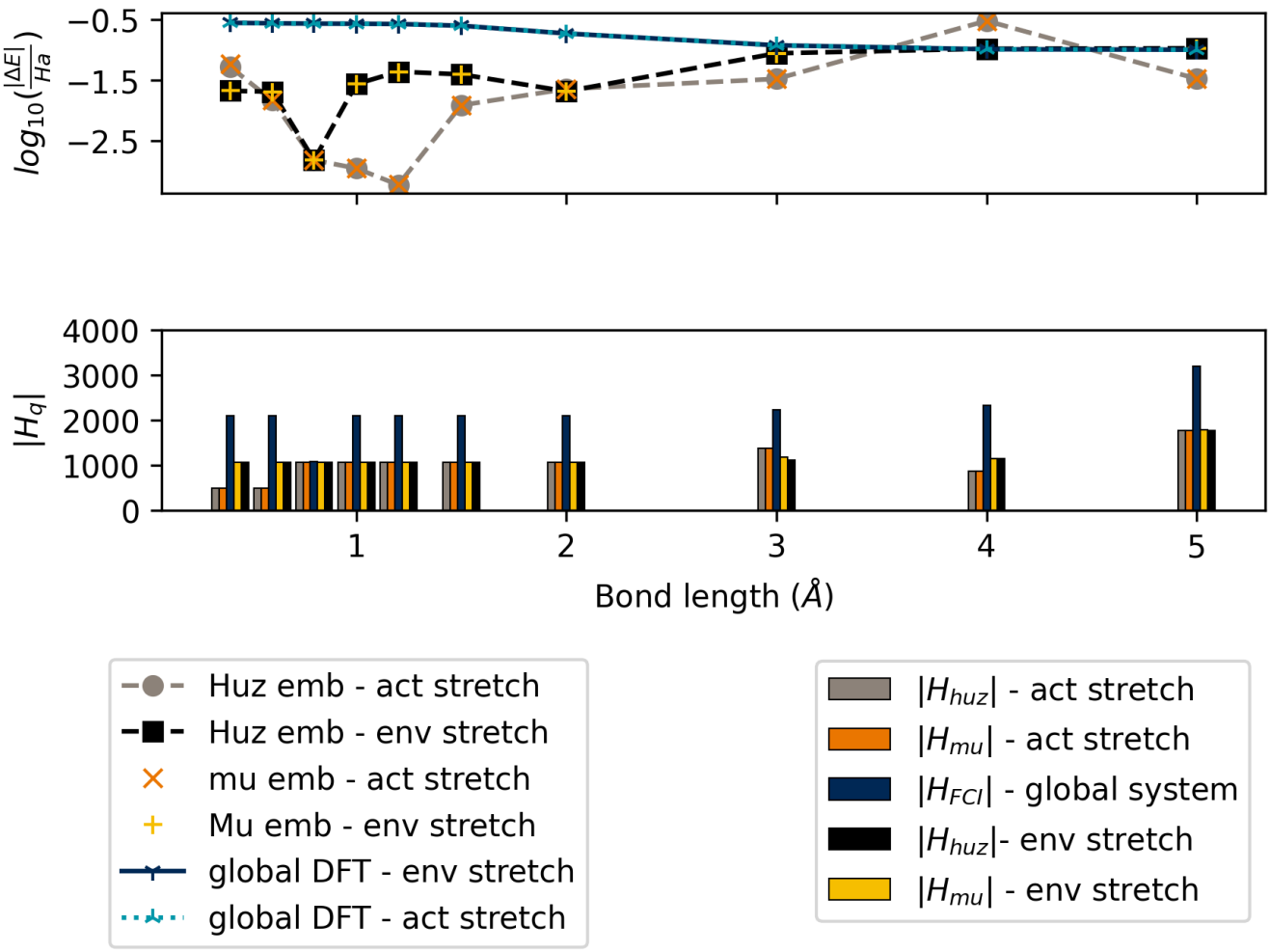

FIG. 3: Potential energy curve for H_2O, with changing OH bond length. *Active stretch* result has the changing OH bond as the active region and *environment stretch* result has the fixed OH bond selected as the active region. These results use IBO localization (95% threshold). For each data set the full problem is reduced from 14 to 12 qubits, with the number of active MOs being three or 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 μ-shifted embedded Hamiltonians while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are provided in Tables VII and VIII.

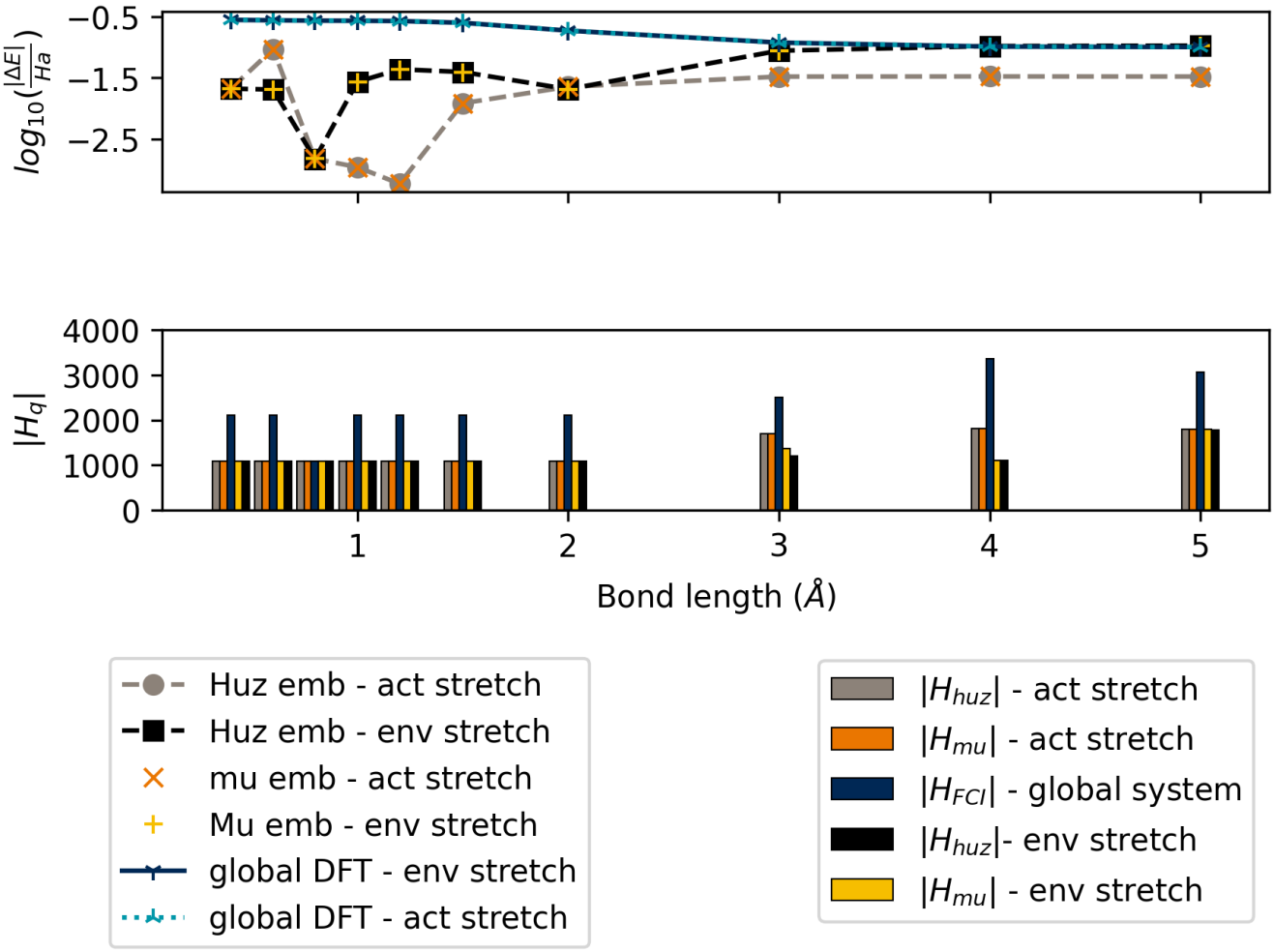

FIG. 4: Potential energy curve for H_2O, with changing OH bond length. *Active stretch* result has the changing OH bond as the active region and *environment stretch* result has the fixed OH bond selected as the active region. results use IBO localization (see raw data for threshold values - lowest setting was 90%). 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 μ-shifted embedded Hamiltonians while the grey and black results are for the Huzinaga embedded Hamiltonians. Numerical details are provided in the supplied raw data.