# Supplemental Material: Molecules in Environments: Toward Systematic Quantum Embedding of Electrons and Drude Oscillators

Matej Ditte,[1,*] Matteo Barborini,[1,†] Leonardo Medrano Sandonas,[1] and Alexandre Tkatchenko[1,‡]

[1] *Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg*

Table I. Parametrization of the quantum Drude oscillators used in this work.

| | q | $\omega$ | $\mu$ |
|---|---|---|---|
| Ar [70] | 1.3314 | 0.7272 | 0.3020 |
| Kr [70] | 1.3741 | 0.6359 | 0.2796 |
| Xe [70] | 1.3570 | 0.5152 | 0.2541 |
| $H_2O$[a] [73] | 1.1973 | 0.6287 | 0.3656 |

[a] The full water QDO model consists also of two point charges centered respectively on the hydrogen atoms (with charge $q_H$ =0.605) and on the point $M$ near to the oxygen (with charge $q_M$ =-1.21).

## PARAMETRIZATION OF THE QDOS

The parametrization of the QDO models used in this work are reported in the Tab. I.

In order to avoid numerical instabilities due to the divergence of the potential energy near the cusps, one way to proceed is to introduce cusp functions in the Jastrow factor of the trial wave function, forcing the correct cancellation of the divergence of the potential through the kinetic component between QDOs and between QDOs and electrons. For now we have decided to proceed in a simpler manner following the work of Martyna and coworkers [73], in which the divergence of the Coulomb potentials is avoided through the introduction of a cut-off function that screens the short-range divergence of the potential. For both QDO-QDO and El-QDO interactions we have used a damping function of the form [73]:

$$V\left(r_{ij}\right) = \frac{q_i q_j}{r_{ij}} \mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2}\sigma_{ij}}\right), \tag{1}$$

where $\sigma_{ij}$ is specified for all pairs using the combination rule $\sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2}$. The single particle damping parameters $\sigma$ were set to 0.1 for all particles except in the case of QDO model of water where for the center of the QDO it is fixed to $\sigma = 1.2$. [73]

The functional form of the short-range repulsion is shown in the Eq. 2, where we used $N = 2$ for both dimers and both QDO-QDO and El-QDO system. The fitted parameters can be found in the Tab. IV.

$$V_{rep}\left(R\right) = \sum_{i=1}^{N} a_i e^{-b_i R} \tag{2}$$

## PARAMETRIZATION OF THE EL-FF

We have used TIP3P model of water with $q_O$ = -0.834 and $q_H$ = 0.417 [123]. For the VdW part we use a Lennard-Jones type pairwise function, of the form:

$$E_{VdW}\left(R_{ij}\right) = \frac{A_{ij}}{R_{ij}^{12}} - \frac{C_{ij}}{R_{ij}^{6}}, \tag{3}$$

where the parmeters are defined through the combination rules $A_{ij} = \sqrt{A_i A_j}$ and $C_{ij} = \sqrt{C_i C_j}$, being $A_i = 4\epsilon_i \sigma_i^{12}$ and $C_i = 4\epsilon_i \sigma_i^{6}$.

These vdW parameters, originally optimized for benzine dimer interacting with water, were taken from the Ref. 124 and can be found in the Tab. II.

Table II. Parametrization of the pairwise vdW part of the El-FF used in this work.

| | $\epsilon$ [kcal/mol] | $\sigma$[Å] |
|---|---|---|
| $C_{C_6H_6}$ | 0.0700 | 3.5500 |
| $H_{C_6H_6}$ | 0.0300 | 2.4200 |
| $O_{H_2O}$ | 0.1521 | 3.1507 |
| $H_{H_2O}$ | 0.0460 | 0.4000 |

## HAMILTONIAN OF TWO DIPOLE COUPLED QDOS AND ITS EXACT SOLUTION

The Hamiltonian of two QDOs with equal parametrization $\{q, \omega, \mu\}$ in the dipole approximation has the form

$$\hat{H}^{dip} = \sum_{i=1}^{2}\left[-\frac{1}{2\mu}\nabla^2_{\mathbf{r}_i^d} + \frac{1}{2}\mu\omega^2\left|\mathbf{r}_i^d - \mathbf{R}_i^O\right|^2\right] + \\ + V^{dip}\left(\mathbf{r}_1^d, \mathbf{r}_2^d\right), \tag{4}$$

with two-body dipole coupling

$$V^{dip}\left(\mathbf{r}_1^d, \mathbf{r}_2^d\right) = \frac{q^2}{R^5}\times \\ \times\left[R^2\left(\mathbf{r}_1^d\cdot\mathbf{r}_2^d\right) - 3\left(\mathbf{r}_1^d\cdot\mathbf{R}\right)\left(\mathbf{r}_2^d\cdot\mathbf{R}\right)\right], \tag{5}$$

where $\mathbf{R} = \mathbf{R}_1^O - \mathbf{R}_2^O$ and $R = |\mathbf{R}|$.

This Hamiltonian can be directly diagonalized and its ground state wave function has the form of the *Ansatz* $\Psi_d$ from the main text. Here it is important to note that, despite the same functional from of the wave function, when this *Ansatz* is used for QDOs interacting via

full Coulomb potential, the final energy contains multipolar contributions, beyond the solution of $\hat{H}^{dip}$. The reason is that the elements of the coupling matrix in $\Psi_d$ are independently optmized, allowing spatial symmetry breaking that goes behyond the Gaussian exact solution of the dipole Hamiltonian in eq. 4.

An exact ground state energy of the Hamiltonian $\hat{H}^{dip}$ as a function of the distance between two QDOs $R$ shifted by the sum of the energies of non-interacting fragments ($E_0$) is written as

$$E^{dip}(R) = E^{+}(R) + E^{-}(R) - E_0 \tag{6}$$

where

$$E^{\pm}(R) = \frac{\omega}{2}\left(2\sqrt{1 \pm \frac{q^2}{\mu\omega^2 R^3}} + \sqrt{1 \mp \frac{2q^2}{\mu\omega^2 R^3}}\right) \tag{7}$$

and $E_0 = 2 \times \frac{3}{2}\omega$ is the energy of two non-interacting QDOs.

## HAMILTONIAN OF INTERACTING NUCLEI, ELECTRONS, QDOS AND POINT CHARGES

The total Hamiltonian of interacting system containing $N_n$ atomic nuclei, $N_e$ electrons, $N_d$ QDOs and $N_p$ point charges, defined by positions, charges and indices of the corresponding QDO ($\{\mathbf{R}_i^p, Q_i, p_i\}_{i=1}^{N_p}$), has the form

$$\hat{H}^{tot} = \hat{H}^{e} + \hat{H}^{d} + \hat{V}_{int}^{d-e}, \tag{8}$$

where $\hat{H}^e$ is the standard electronic Hamiltonian describing the interaction between the electrons and the atomic nuclei, $\hat{H}^d$ is the drudonic Hamiltonian containing also the interactions between the QDOs and point charges

$$\hat{H}^{d} = \sum_{i=1}^{N_d} \hat{h}_i^d\left(\mathbf{r}_i^d\right) + \sum_{i=1}^{N_d}\sum_{j>i}^{N_d} \frac{q_i q_j}{\left|\mathbf{R}_i^O - \mathbf{R}_j^O\right|} + \\ + \sum_{i=1}^{N_d}\sum_{j>i}^{N_d} \frac{q_i q_j}{\left|\mathbf{r}_i^d - \mathbf{r}_j^d\right|} - \sum_{i=1}^{N_d}\sum_{\substack{j=1 \\ p_j \neq i}}^{N_p}\left(\frac{q_i Q_j}{\left|\mathbf{r}_i^d - \mathbf{R}_j^p\right|} - \frac{q_i Q_j}{\left|\mathbf{R}_i^O - \mathbf{R}_j^p\right|}\right), \tag{9}$$

with

$$\hat{h}_i^d\left(\mathbf{r}_i^d\right) = -\frac{1}{2\mu_i}\nabla_{\mathbf{r}_i^d}^2 + \frac{1}{2}\mu_i\omega_i^2\left|\mathbf{r}_i^d - \mathbf{R}_i^O\right|^2 - \\ - \sum_{j\neq i}^{N_d}\frac{q_i q_j}{\left|\mathbf{r}_i^d - \mathbf{R}_j^O\right|}. \tag{10}$$

The $\hat{V}_{int}^{d-e}$ term contains extra interactions between the electrons/nuclei and the point charges

$$\hat{V}_{int}^{d-e} = \sum_{i=1}^{N_e}\sum_{j=1}^{N_d}\left(\frac{q_j}{\left|\mathbf{r}_i^e - \mathbf{r}_j^d\right|} - \frac{q_j}{\left|\mathbf{r}_i^e - \mathbf{R}_j^O\right|}\right) + \\ + \sum_{i=1}^{N_n}\sum_{j=1}^{N_d}\left(\frac{Z_i q_j}{\left|\mathbf{R}_i^n - \mathbf{R}_j^O\right|} - \frac{Z_i q_j}{\left|\mathbf{R}_i^n - \mathbf{r}_j^d\right|}\right) - \\ - \sum_{i=1}^{N_e}\sum_{j=1}^{N_p}\frac{Q_j}{\left|\mathbf{r}_i^e - \mathbf{R}_j^p\right|} + \sum_{i=1}^{N_n}\sum_{j=1}^{N_p}\frac{Z_i Q_j}{\left|\mathbf{R}_i^n - \mathbf{R}_j^p\right|}. \tag{11}$$

## MAPPING OF THE WATER MOLECULES ONTO QDOS

The QDO model of water, introduced in Ref. 73, is mapped onto the particular geometries of the molecules in the cages. Two point charges are at the positions of the hydrogen atoms. The third point charge is at the M point, which is placed 0.2667 Å from the oxygen atom in the HOH plane with $\sphericalangle$HOM = $\frac{1}{2}\sphericalangle$HOH.

## DETAILS ON THE QUANTUM MONTE CARLO CALCULATIONS

### Variational Monte Carlo

We have generalized the variational Monte Carlo (VMC) algorithm [63, 65] to integrate a mixed system of drudons and electrons. In our approach, the two sets of particles are diffused particle-by-particle in random order always starting from all the electrons, followed by all the drudons, according to the Metropolis-Hastings algorithm [85, 86]. Each particle's trial move is proposed according to a 3-dimensional Gaussian transition probability centered on the initial particle's position and with standard deviation $dt_e$ for the electrons and of $dt_d$ for the drudons. The two standard deviations are optimized separately for the two particle types by automatically converging the acceptance probability of the moves to the value of 50%, which is the classical rule of thumb that has the purpose of lowering the correlation between configurations. This procedure is repeated until $\mathcal{N}$ configurations are sampled.

Within this VMC scheme, it is also possible to optimize the trial wave function through energy (or variance) minimization. In this work, the set of parameters is optimized through the Stochastic Reconfiguration procedure described in Refs. 109, 110 with the use of the Correlated sampling technique [87] in order to increase sensitivity within the range of the rather small energy differences that are usually involved in dispersion interactions.

### Trial wave function and variational parameters

As discussed in the main text the total variational wave function of the mixed system of electrons and drudons is factorized in three parts:

$$\Psi_{tot} = \Psi_e(\bar{\mathbf{r}}^e)\Psi_d(\bar{\mathbf{r}}^d)\mathcal{J}_{e-d}(\bar{\mathbf{r}}^e, \bar{\mathbf{r}}^d), \tag{12}$$

that are respectively the pure electronic $\Psi_e(\bar{\mathbf{r}}^e)$, pure drudonic $\Psi_d(\bar{\mathbf{r}}^d)$ and the interaction $\mathcal{J}_{d-e}(\bar{\mathbf{r}}^e, \bar{\mathbf{r}}^d)$ parts.

In this work the pure electronic wave function

$$\Psi_e(\bar{\mathbf{r}}^e) = \det[\mathbf{S}(\bar{\mathbf{r}}^e)]\mathcal{J}_{e-e}(\bar{\mathbf{r}}^e) \tag{13}$$

is constructed as a Slater determinant $\det[\mathbf{S}(\bar{\mathbf{r}}^e)]$ times a Jastrow factor $\mathcal{J}_{e-e}(\bar{\mathbf{r}}^e)$ that describes the many-body interactions between electron and nuclei. The molecular orbitals that define the elements of the Slater matrix $\mathbf{S}(\bar{\mathbf{r}}^e)$ are written as linear combinations

$$\varphi_k(\mathbf{r}) = \sum_{q=1}^{Q} c_q^k \phi_q(\mathbf{r}) \tag{14}$$

of $Q$ contracted Gaussian type orbitals $\phi_q(\mathbf{r})$ centered only on the nuclei of the electronic system and not on the oscillators' centers. The form of the electronic Jastrow factor is similar to the one described by Marchi *et al.* in Ref. 88, 89, and is written as the exponential of a sum of two terms

$$\mathcal{J}_{e-e}(\bar{\mathbf{r}}^e) = e^{\mathcal{J}(\bar{\mathbf{r}})} = e^{\mathcal{J}_2(\bar{\mathbf{r}})+\mathcal{J}_{3/4}(\bar{\mathbf{r}})}. \tag{15}$$

that are respectively a pure homogeneous two-body $\mathcal{J}_2(\bar{\mathbf{r}})$ term and a three/four-body inhomogeneous term $\mathcal{J}_{3/4}(\bar{\mathbf{r}})$.

The homogeneous two-body Jastrow describes the paring of electronic coordinates

$$\mathcal{J}_2(\bar{\mathbf{r}}) = \sum_{j>i=1}^{N_e} f_{ee}(r_{ij}), \tag{16}$$

where the $f_{ee}$ function is defined as the sum of a cusp function and a linear combination of Gaussian functions

$$f_{ee}(r_{ij}) = \begin{cases} -\frac{1}{4b_0^p(1+b_0^p r_{ij})} + \sum_{n=1}^{N} b_n^p e^{\zeta_n^p r_{ij}^2} & \text{undist.} \\ -\frac{1}{2b_0^a(1+b_0^a r_{ij})} + \sum_{n=1}^{N} b_n^a e^{\zeta_n^a r_{ij}^2} & \text{dist.} \end{cases}. \tag{17}$$

The $\mathcal{J}_{3/4}$ term in the Jastrow factor is derived from the construction of the geminal function of electron couples $\sum_{q,p=1}^{\mathcal{Q}} \gamma_{qp}\chi_q(\mathbf{r}_i)\chi_p(\mathbf{r}_j)$, as introduced by Sandro Sorella [88]

$$\mathcal{J}_{3/4} = \sum_{j>i=1}^{N_e} \sum_{q,p=1}^{\mathcal{Q}} \gamma_{qp}\chi_q(\mathbf{r}_i)\chi_p(\mathbf{r}_j). \tag{18}$$

The $\gamma_{qp}$ parameters define the coupling of non-normalized atomic orbitals $\chi_q(\mathbf{r})$ that can be centered on the same atom (defining three-body terms) or on different atoms (four body terms). These terms are necessary to recover the dynamical correlation between electronic pairs, suppressing nonphysical charge fluctuations [90].

Here we do not recall the pure drudonic wave function and the electron-drudon Jastrow correlation function that are described in the main text. We just want to specify that in our optimizations the wave function's parameters that are optimized are namely: all the Jastrow parameters for the electronic part of the wave function, while for the drudonic wave function and the coupling Jastrow between the electronic and drudonic system we optimize all the coefficients of the matrices $\mathbf{A}$ and $\mathbf{B}$.

### Diffusion Monte Carlo

In order to generalize the diffusion Monte Carlo [91, 92] (DMC) algorithm to integrate both particle types at the same time, the main change consists on a modification of the Langevin dynamics of the particles, to include also the different masses.

Thus, during the drift/diffusion process we update the particles' positions with a particle-by-particle scheme such that from time step $m$ to $m+1$ we will have

$$\mathbf{r}_i^{(m+1)} = \mathbf{r}_i^{(m)} + \frac{\delta\tau}{\mu_i}\tilde{\mathbf{v}}_i(\bar{\mathbf{r}}^{(m)}) + \sqrt{\frac{\delta\tau}{\mu_i}}\boldsymbol{\eta}, \tag{19}$$

where $\boldsymbol{\eta}$ is a 3-dimensional vector of random variables extracted with a Gaussian distribution with zero mean value and unitary variance, and $\tilde{\mathbf{v}}_i(\bar{\mathbf{r}}^{(m)})$ is the drift velocity rescaled according to the procedure introduced by Umrigar *et al.* [93] to avoid divergences near the nodal surface.

For the systems described at the electronic level, we substitute the core electrons with the ccECP pseudopotentials [94–97] , which are integrated with the Determinant Locality approximation (DLA) [98] in which the non-local operator is projected only on the Slater determinant part of the many-body wave function and not on the Jastrow factor. For the electronic systems and the mixed system, we also introduce an energy cut-off using Zen's correction [99] with a parameter set to $\alpha = 0.2$. For pure QDO systems, no cut-off has been used since the systems don't include nodes.

## COMPUTATIONAL DETAILS

To construct the wave functions of the electronic systems we employ ccECP effective core potentials [94–97] with the corresponding (aug)-cc-pVDZ Gaussian basis sets for the Ar and water dimers and cc-pVTZ for everything else. The molecular orbitals are obtained from DFT calculations using the PBE0 functional [122] in GAMESS

Table III. Number of walkers and the total number of steps per walker used for the DMC calculations. The settings for the El-FF approach are identical to those used for the El-QDO one.

| | # of walkers | # of steps |
|---|---|---|
| QDO-QDO | | |
| $Ar_2$ | 25600 | 100000 |
| $(H_2O)_2$ | 25600 | 200000 |
| El-QDO | | |
| $Ar_2$ | 153600 | 200000 |
| $(H_2O)_2$ | 102400 | 40000 |
| Orthobenzyne | | |
| Mol. in vacuum | 13440 | 800000 |
| Mol. in 4 waters cage | 13440 | 550000 |
| Mol. in 4 waters QDO cage | 15360 | 800000 |
| Mol. in 30 waters QDO cage | 13440 | 800000 |
| QDO cages | 10080 | 320000 |
| Benzene dimer (50 waters cage) | | |
| Monomers | 15360 | 320000 |
| Dimer | 15360 | 320000 |
| QDO cage | 15360 | 320000 |
| Monomers in QDO cage | 15360 | 320000 |
| Dimer in QDO cage | 15360 | 320000 |

(2016 R1) [100] and Orca 5.0 [101] codes. The dynamical Jastrow factor is built from 3s2p1d uncontracted Gaussian-type orbitals (GTOs) for all the heavy atoms and from 2s1p GTOs for the Hydrogen atoms.

The DMC calculations have been carried out with a fixed time step of $\delta\tau = 0.005$ a.u. for all systems, which was chosen after careful convergence tests. The statistics used for all DMC calculations in this paper can be found in the Tab. III. The reference PBE0+TS and PBE0+MBD calculations on the large water cluster have been done with the FHI-aims package [102] using the tight basis set.

## WATER CAGES

The geometry of the T-shaped benzene dimer in the 50W cage is taken from Ref. 121 and the cage has been expanded by 1.5 Å from the center, with preserved orientations. A picture of the final cage is shown in the Fig. 1.

4W and 30W cages used for the orthobenzyne were obtained by scanning over a large number of randomly generated geometries with oxygens randomly placed on a sphere of radius 6 Å centered at the center of the orthobenzyne. The minimal distance between oxygens was set to 3 Å. The OH distances and HOH angles were fixed to 0.958 Å and 104.4°respectively and the orientation of the HOH plane was generated randomly. The final cages were the ones maximizing the change of the singlet-triplet gap compared to the gap in vacuum. Pictures of the final cages with 4 and 30 water are shown in Fig. 2 and Fig. 3.

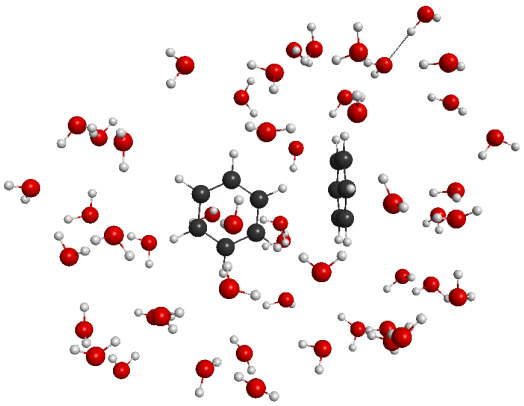

Figure 1. T-shaped benzene dimer in a cage of 50 water molecules [121] with the 1.5 Å shift. The minimal distance between the cage and the benzene dimer is 3.4 Å. The same cage has been used for both monomers.

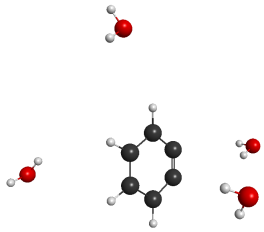

Figure 2. Orthobenzyne (singlet) [120] in an artificial cage composed of 4 water molecules. The minimal distance between the cage and the orthobenzyne is 3.61 Å. The same cage has been used for the triplet state.

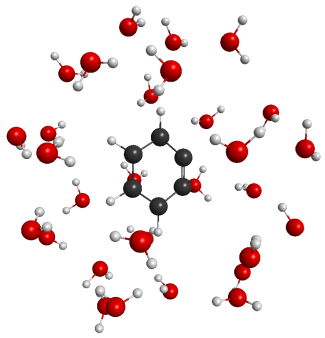

Figure 3. Orthobenzyne (singlet) [120] in an artificial cage composed of 30 water molecules. The minimal distance between the cage and the orthobenzyne is 2.99 Å. The same cage has been used for the triplet state.

## ADDITIONAL RESULTS

### Dispersion and polarization contributions in the El-QDO and QDO-QDO approaches.

In order to better understand the results, we present in Figs 4, 5, 6 and 7, the dissociation curves of $Ar_2$, $Kr_2$, $Xe_2$ and water dimers, uncorrected for the short-range re-

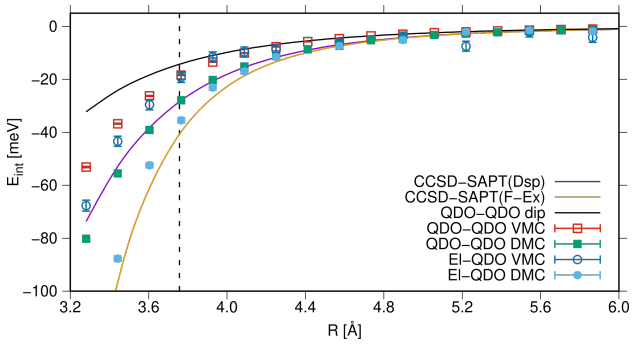


Figure 4. Interaction energies as a function of the atom-atom distance of the argon dimer obtained using the QDO-QDO model and the electrons-QDO (El-QDO) embedding approach with VMC and DMC. The results are compared to the exact solution of dipole coupled QDOs (QDO-QDO dip), pure dispersion interaction (Dsp), and with the full interaction energy minus the exchange contribution (F-Ex) obtained from the Coupled Cluster based symmetry adapted perturbation theory (SAPT-CCSD) [116]. The straight vertical line represents the equilibrium geometry of the dimer.

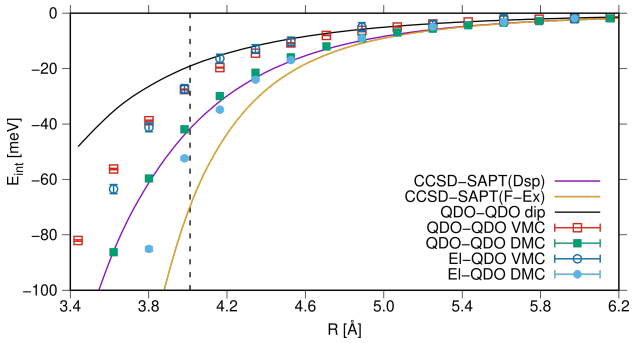


Figure 5. Interaction energies as a function of the atom-atom distance of the Krypton dimer obtained using the QDO-QDO model and the electrons-QDO (El-QDO) embedding approach with VMC and DMC. The results are compared to the exact solution of dipole coupled QDOs (QDO-QDO dip), pure dispersion interaction (Dsp), and with the full interaction energy minus the exchange contribution (F-Ex) obtained from the Coupled Cluster based symmetry adapted perturbation theory (SAPT-CCSD) [116]. The straight vertical line represents the equilibrium geometry of the dimer.

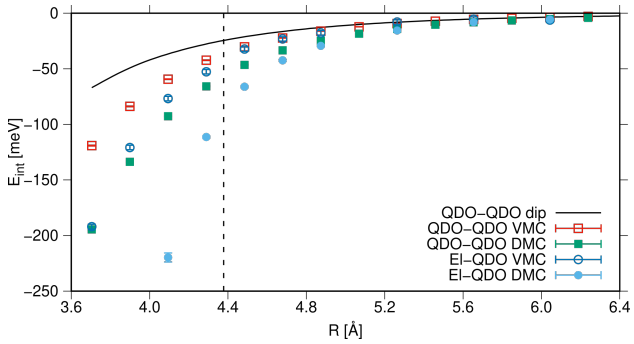


Figure 6. Interaction energies as a function of the atom-atom distance of the Xenon dimer obtained using the QDO-QDO model and the electrons-QDO (El-QDO) embedding approach with VMC and DMC. The results are compared to the exact solution of dipole-coupled QDOs (QDO-QDO dip).

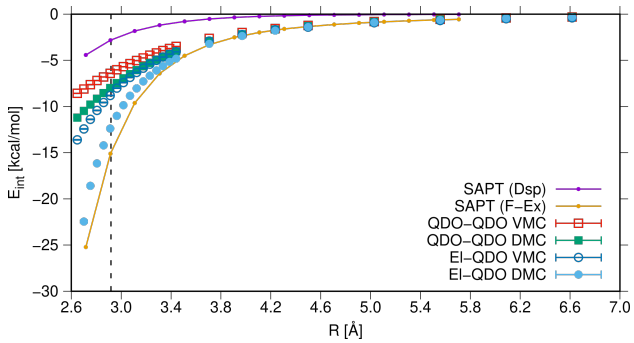


Figure 7. Interaction energies as a function of the oxygen-oxygen distance of the water dimer in the equilibrium geometry [114] obtained using the QDO-QDO model and the electrons-QDO (El-QDO) embedding approach (with QDO approximating the donor) with VMC and DMC. The results are compared to the pure dispersion interaction (Dsp) and with the full interaction energy minus the exchange contribution (F-Ex) obtained from the DFT-based symmetry adapted perturbation theory (SAPT-DFT) [117]. The straight vertical line represents the equilibrium geometry of the dimer.

pulsion potential. The QDO-QDO and El-QDO models at VMC and DMC levels are compared to the energy contributions obtained from the energy decomposition of the symmetry adapted perturbation theory (SAPT), namely: pure dispersion (Dsp) and the full interaction energy minus the exchange contribution (F-Ex); and to the exact solution of the dipole coupled QDOs in the case of noble gas dimers. For $Ar_2$ and $Kr_2$ we notice the ability of the QDO-QDO model at the DMC level to exactly reproduce the dispersion curve. Unfortunately for $Xe_2$ we couldn't find reference values in the literature. On the other hand, the El-QDO model includes some short-range polarization effects that are neglected in the case of pure QDOs. The differences between the exact solution of the dipole coupled QDOs (QDO-QDO dip) and VMC results with the full Coulomb potential show the presence of multipolar contributions in the case of the full Coulomb, which diminish for large separations between the QDOs/atoms. In the water dimer, the QDO-QDO model does not correspond to the pure dispersion from SAPT due to the presence of the point charges in the model. The difference between the QDO representing the donor and acceptor in the El-QDO model is shown in the Fig. 8. The two curves are not identical in the short-range region due to the different charge transfer effects that take place in the dimer, depending on which water molecule is represented by the QDO, if the donor one (the molecule that offers the H atom) or the acceptor one (the molecule that contains with the lone pair electrons on the oxygen atom

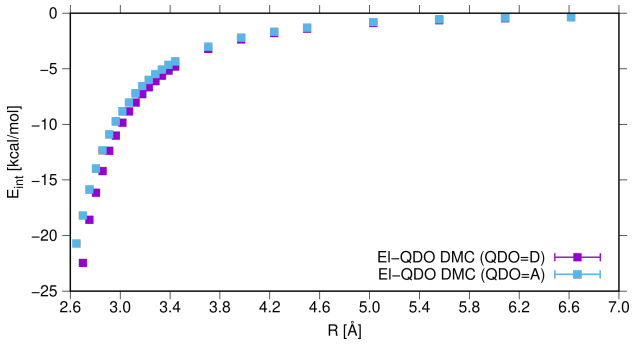


Figure 8. Comparison of El-QDO DMC interaction energies of water dimer in the equilibrium geometry without the short-range repulsion with QDO representing the donor (QDO=D) and acceptor (QDO=A).

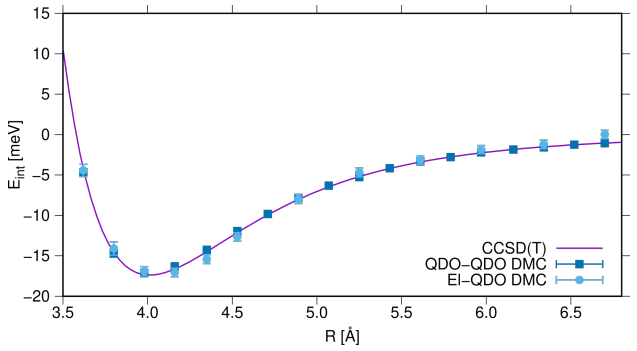


Figure 9. Binding energy curve for the Kr dimer obtained using the QDO-QDO model and electrons-QDO (El-QDO) embedding approach at the DMC level of theory, with the exponential fit of the short-range repulsion. The results are compared to the CCSD(T) [104] reference curve.

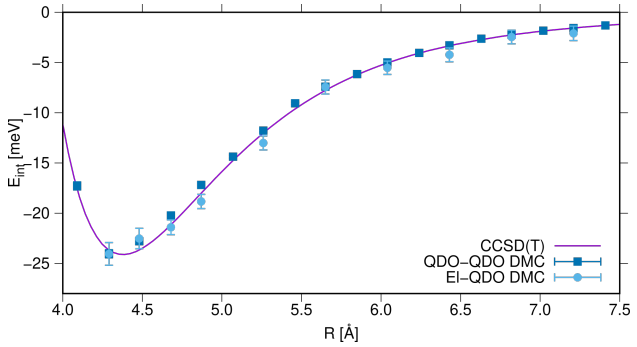


Figure 10. Binding energy curve for the Xe dimers obtained using the QDO-QDO model and electrons-QDO (El-QDO) embedding approach at the DMC level of theory, with the exponential fit of the short-range repulsion. The results are compared to the CCSD(T) [105] reference curve.

that contribute directly to the bond). Biswas et al. [103] pointed out that for standard QM/MM methods, the QM should describe the acceptor water molecule, in order to enable the charge transfer of the two lone pair electrons on the oxygen atom towards the Hydrogen of the donor molecule. In our El-QDO model a partial charge transfer is also possible in the case in which the acceptor molecule is replaced by the QDO, since the drudon is able to drift towards the donor. For this reason the differences in the energies of the two curves in Fig. 8 are relatively small at the equilibrium distance of the dimer, and the total binding curve with the fit of the short-range repulsion is indistinguishable from the one in the main text.

Clearly, the VMC results underestimate the binding energies in all cases due to the limits of the variational *Ansatz*. The general conclusion from both results is that QDO-QDO, El-QDO, and F-Ex curves overlap in the long-range region, and this demonstrates that our embedding approach is able to recover the important quantum effects at this scale.

## Modelling short-range repulsion QDO-QDO and El-QDO

In order to show the general applicability of the QDO model in Figs. 9 and 10 the QDO-QDO and El-QDO DMC dissociation curves are also shown respectively for the Kryton and Xenon dimers with the addition of the interpolated short-range repulsion (see Table IV for the parameters), and compared to the CCSD(T) references.

The comparison of the short-range repulsions of the $Ar_2$ and the water dimer are shown in Figs. 11 and 12. Here we compare our interpolated potentials obtained from the QDO-QDO and El-QDO models with the pure exchange contribution from SAPT (Ex pure), the sum of all SAPT terms containing exchange (Ex all in the figures includes pure exchange, pol-ex, and disp-ex). It is interesting to notice that for $Ar_2$ in which the interaction is purely dispersive, the repulsion of interpolated for the QDO-QDO interaction corresponds to the pure exchange contribution coming from SAPT. On the other hand for the El-QDO model the interpolated repulsion is more similar to the full exchange contributions from SAPT, meaning that in the El-QDO model the attractive potential energy curve is more compatible with the sum of terms interaction terms arising from electrostatics, polarization and dispersion (see also Fig. 4). The analysis for the water dimer is more complex due to the Hydrogen bond and to the presence of additional point charges in the QDO model, which provide additional electrostatic contributions. The repulsion obtained for the QDO-QDO model differs from the SAPT components, while the El-QDO interpolated repulsion on the other hand is similar to the pure exchange contribution from SAPT.

Table IV. Parameters of the short-range repulsive potentials in atomic units obtained by interpolating the DMC calculations of the QDO-QDO and El-QDO systems to match the reference potentials.

| | $a_1$ [mHa] | $a_2$ [mHa] | $b_1$ [$a_0^{-1}$] | $b_2$ [$a_0^{-1}$] |
|---|---|---|---|---|
| QDO-QDO | | | | |
| Ar | $4.030{\cdot}10^{5}$ | $5.925{\cdot}10^{-2}$ | 1.891 | 33.38 |
| Kr | $4.867\ {\cdot}10^{5}$ | $3.376\ {\cdot}10^{0}$ | 1.753 | 5.957 |
| Xe | $5.049\ {\cdot}10^{5}$ | $3.466\ {\cdot}10^{0}$ | 1.567 | 5.020 |
| $H_2O$ (C1) | $1.851\ {\cdot}10^{5}$ | $1.491\ {\cdot}10^{2}$ | 1.927 | 5.176 |
| $H_2O$ (C2) | $1.573\ {\cdot}10^{5}$ | $1.496\ {\cdot}10^{2}$ | 1.956 | 5.064 |
| El-QDO | | | | |
| Ar | $3.112\ {\cdot}10^{6}$ | $3.462\ {\cdot}10^{0}$ | 2.129 | 5.140 |
| Kr | $5.788\ {\cdot}10^{6}$ | $3.436\ {\cdot}10^{0}$ | 2.034 | 4.714 |
| Xe | $2.041\ {\cdot}10^{7}$ | $3.652\ {\cdot}10^{0}$ | 1.932 | 4.073 |
| $H_2O$ (C1, QDO=D) | $9.074\ {\cdot}10^{5}$ | $1.503\ {\cdot}10^{2}$ | 2.042 | 52.388 |
| $H_2O$ (C2, QDO=D) | $1.153\ {\cdot}10^{6}$ | $1.503\ {\cdot}10^{8}$ | 2.080 | 3.858 |
| $H_2O$ (C1, QDO=A) | $1.198\ {\cdot}10^{6}$ | $1.492\ {\cdot}10^{2}$ | 2.136 | 5.163 |
| $H_2O$ (C2, QDO=A) | $5.887\ {\cdot}10^{5}$ | $1.501\ {\cdot}10^{2}$ | 2.091 | 4.909 |

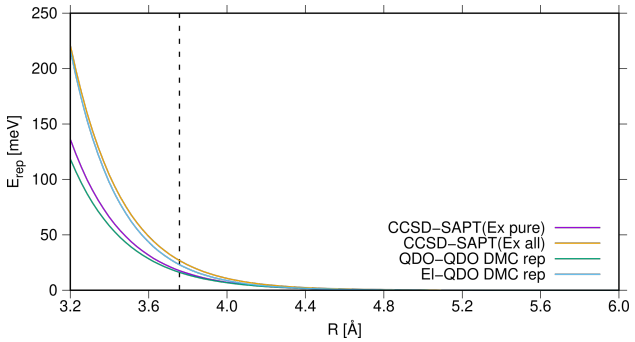


Figure 11. Short range repulsion contribution from the $E_{int}$ of $Ar_2$. We compare pure exchange from SAPT (Ex pure), and the sum of all mixed terms from SAPT containing exchange (Ex all), with the interpolated repulsion for QDO-QDO DMC and EL-QDO DMC.

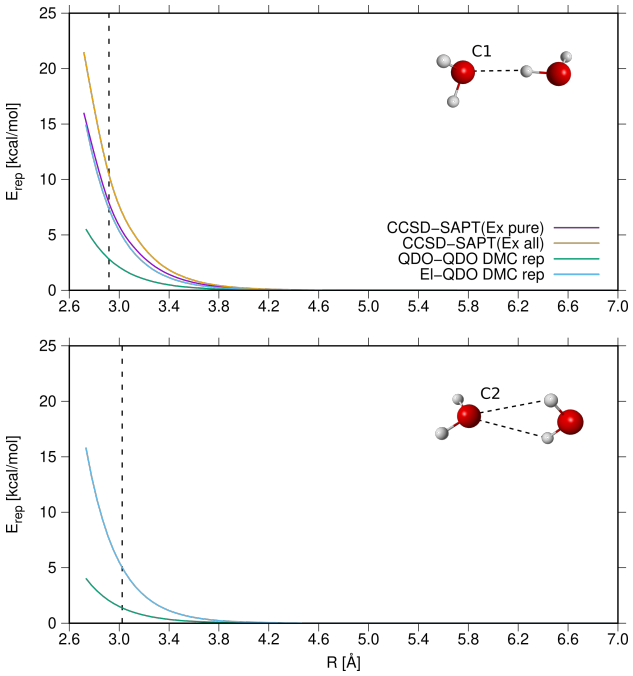


Figure 12. Short range repulsion contribution from the $E_{int}$ of water dimer. We compare pure exchange from SAPT (Ex pure), and the sum of all mixed terms from SAPT containing exchange (Ex all), with the interpolated repulsion for QDO-QDO DMC and EL-QDO DMC.

Table V. The total energies (in Hartree) of the benzene dimer in 50 water molecules used for the Tab. I of the main article.

| | PBE0 | PBE0+TS | PBE0+MBD | El-QDO DMC |
|---|---|---|---|---|
| Cage | -3823.95689 | -3823.99421 | -3824.00394 | -4.5656(1) |
| In Vacuum | | | | |
| M1 | -232.21019 | -232.21243 | -232.21701 | -37.6457(1) |
| M2 | -232.21016 | -232.21241 | -232.21698 | -37.6428(1) |
| D | -464.42066 | -464.42964 | -464.43843 | -75.2931(1) |
| In Water Cage | | | | |
| M1 | -4056.16938 | -4056.21324 | -4056.22697 | -42.2176(1) |
| M2 | -4056.17241 | -4056.21712 | -4056.23057 | -42.2186(1) |
| D | -4288.38541 | -4288.44122 | -4288.45825 | -79.8752(2) |

## Total energy comparisons

In tabs. V and VI we report the total energies of the various systems, used to compute the energy differences reported in the manuscript.

In addition, in Tab. VII we report the binding energies of the benzene dimer in vacuum and in the water cage of 50 water molecules. Furthermore, Tab. VIII contains the analysis of the El-FF DMC calculations, separating the contributions coming from the electrostatics of the TIP3P model [123] from the ones corresponding to the additional vdW potentials [124].

## Runtimes and computational efficiency

In Tabs. IX and X we compare the runtimes of the standard VMC and DMC calculations compared to the El-QDO method. Tab. XI shows the cost of the optimization of the wave functions and Tab. XII contains the root mean square deviation $\sigma$ of wave functions of orthobenzyne embedded in 30 QDOs after optimizing different sets of parameters. Fig. 13 shows the energy during the optimization for the same system.

## Pseudopotential error

The 2.85 kcal/mol ECP error at UPBE0 level with aug-cc-pVTZ Gaussian basis set is shown in the Tab. XIII and the uncorrected DMC / El-QDO DMC S-T gaps in the Tab. XIV.

Table VI. The total energies (in Hartree) of the orthobenzyne calculations from the main article.

| | PBE0 | PBE0+TS | PBE0+MBD | DMC | El-QDO DMC |
|---|---|---|---|---|---|
| 4W Cage | -305.891122 | -305.891138 | -305.891976 | ... | -0.3302304(6) |
| 30W Cage | -2294.198539 | -2294.223299 | -2294.227611 | ... | -2.55168(2) |
| S (V) | -230.873629 | -230.875158 | -230.879244 | -36.30409(6) | -36.30409(6) |
| T (V) | -230.828648 | -230.830168 | -230.834204 | -36.24930(6) | -36.24930(6) |
| S (4W) | -536.766354 | -536.768679 | -536.773502 | -105.2887(1) | -36.63654(6) |
| T (4W) | -536.720994 | -536.723287 | -536.728066 | -105.2333(1) | -36.58132(5) |
| S (30W) | -2525.076666 | -2525.109265 | -2525.116585 | ... | -38.86565(9) |
| T (30W) | -2525.030960 | -2525.063680 | -2525.070911 | ... | -38.81012(9) |

Table VII. The binding energies (in kcal/mol) of the benzene dimer in vacuum ($dE_v$) and in the 50 water molecules ($dE_c$).

| | PBE0 | PBE0+TS | PBE0+MBD | El-QDO DMC |
|---|---|---|---|---|
| $dE_v$ | -0.2 | -3.0 | -2.8 | -2.9(1) |
| $dE_c$ | -0.3 | -3.2 | -2.9 | -2.8(2) |
| diff | -0.1 | -0.2 | -0.1 | 0.1(2) |

Table VIII. Decomposition of the QM/MM total energies and energy differences into electrons - point charges contribution (El-TIP3P [123]) and pairwise van der Waals contributions [124] (vdW). Full interactions are defined as El-FF.

| | El-TIP3P | vdW | El-FF |
|---|---|---|---|
| Total energies (in Hartree) | | | |
| M1 | -7.01122(6) | -0.00390 | -7.01512(6) |
| M2 | -7.01075(7) | -0.00424 | -7.01499(7) |
| D | -44.6632(2) | -0.00814 | -44.6713(2) |
| 50W cage | 30.63610 | ... | ... |
| S (4W) | -33.82662(5) | -0.00069 | -33.82731(5) |
| T (4W) | 33.77152(5) | -0.00067 | -33.77219(5) |
| S (30W) | -17.72151(6) | -0.00541 | -17.72692(6) |
| T (30W) | -17.66595(6) | -0.00546 | -17.67141(6) |
| 4W cage | 2.47887 | ... | ... |
| 30W cage | 18.58584 | ... | ... |
| Solvation energies (in kcal/mol) | | | |
| M1 | -1.04(6) | -2.45 | -3.48(6) |
| M2 | -2.52(6) | -2.66 | -5.18(6) |
| D | -3.8(1) | -5.11 | -8.9(1) |
| S (4W) | -0.87(5) | -0.43 | -1.30(5) |
| T (4W) | -0.69(5) | -0.42 | -1.11(5) |
| S (30W) | -2.04(5) | -3.40 | -5.44(5) |
| T (30W) | -1.56(5) | -3.43 | -4.99(5) |
| S-T gaps (in kcal/mol) | | | |
| S-T (4W) | 34.57(5) | 0.01 | 34.58(5) |
| S-T (30W) | 34.86(5) | -0.03 | 34.83(5) |

Table IX. Relative runtimes of El-QDO DMC method with respect to vacuum and electronic cage DMC. Test calculations have been done on 56 CPUs using 12 walkers per CPU and 20 bins with 100 steps per block. The time per block is an average over all 20 blocks. Relative runtime is calculated as the ratio of time per block times the square of the ratio of the root mean square deviation $\sigma$.

| | time per block [s] | $\sigma$ [Ha] | runtime relative to vacuum |
|---|---|---|---|
| orthobenzyne (singlet) | | | |
| El (V) | 6.36 | 0.80 | 1.00 |
| El (4W) | 25.10 | 1.40 | 11.99 |
| El-QDO (4W) | 6.45 | 0.80 | 1.02 |
| El-QDO (30W) | 6.78 | 0.87 | 1.25 |
| benzene (monomer$_1$) | | | |
| El (V) | 7.15 | 0.80 | 1.00 |
| El-QDO (50W) | 8.34 | 0.89 | 1.47 |
| benzene dimer | | | |
| El (V) | 30.95 | 1.31 | 1.00 |
| El-QDO (50W) | 32.90 | 1.57 | 1.27 |

Table X. Relative runtimes of El-QDO VMC method with respect to vacuum and electronic cage VMC. Test calculations have been done on 56 CPUs using 1 walker per CPU and 20 bins with 100 steps per block. The time per block is an average over all 20 blocks. Relative runtime is calculated as the ratio of time per block times the square of the ratio of the root mean square deviation $\sigma$.

| | time per block [s] | $\sigma$ [Ha] | runtime relative to vacuum |
|---|---|---|---|
| orthobenzyne (singlet) | | | |
| El (V) | 1.26 | 0.80 | 1.00 |
| El (4W) | 4.63 | 1.42 | 11.47 |
| El-QDO (4W) | 1.29 | 0.82 | 1.05 |
| El-QDO (30W) | 1.45 | 0.88 | 1.36 |

* matej.ditte@uni.lu
† matteo.barborini@gmail.com
‡ Corresponding author:alexandre.tkatchenko@uni.lu

[63] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. 73, **33** (2001).
[65] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, England, 2017).
[70] A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield, and G. J. Martyna, Phys. Rev. B **87**, 144103 (2013).
[73] F. S. Cipcigan, J. Crain, V. P. Sokhan, and G. J. Martyna, Rev. Mod. Phys. **91**, 025003 (2019)
[85] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

Table XI. Total runtimes of El-QDO optimization of orthobenzyne in the singlet state. Test calculations have been done on 56 CPUs using 1 walker per CPU, 200 bins with 2 steps per block, and 10 optimization steps without stochastic reconfiguration.

| | # of parameters | runtime [s] |
|---|---|---|
| El (V) | 376 | 56.00 |
| El (4W) | 692 | 198.21 |
| El-QDO (4W) | 490 | 58.04 |
| El-QDO (30W) | 4741 | 72.65 |

Table XII. The root mean square deviation $\sigma$ [Ha] of orthobenzyne wave function in singlet state in vacuum and in a water cage composed of 30 molecules using El-QDO embedding after each of 5 blocks of optimization: vacuum - $\Psi_e$ optimized in vacuum, QDO - optimization of $\Psi_d$, El- optimization of $\Psi_e$, El-QDO - optimization of $\mathcal{J}_{e-d}$ and all - optimization of all parts together. The initial guess of $\Psi_d$ was set to non-interacting oscillators and the coupling matrix in $\mathcal{J}_{e-d}$ was filled with zeroes. Decrease of the energy during the optimization is shown in the Fig. 13.

| Optimization block | $\sigma$ [Ha] |
|---|---|
| vacuum | 0.80 |
| unopt QDO | 0.91 |
| QDO | 0.86 |
| El | 0.86 |
| El-QDO | 0.86 |
| all | 0.85 |

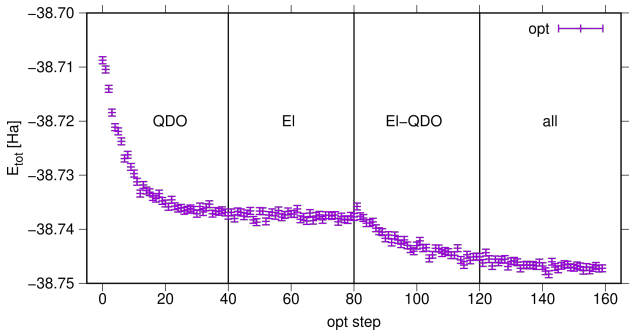


Figure 13. Energy optimization of orthobenzyne (S) in a cage of 30 water molecules using El-QDO embedding. The electronic part was preoptimized in vacuum followed by 4 blocks of optimization: QDO - optimization of $\Psi_d$, El- optimization of $\Psi_e$, El-QDO - optimization of $\mathcal{J}_{e-d}$ and all - optimization of all parts together. The initial guess of $\Psi_d$ was set to non-interacting oscillators and the coupling matrix in $\mathcal{J}_{e-d}$ was filled with zeroes. The decrease of the root mean square deviation $\sigma$ during the optimization is shown in the Tab. XII.

Table XIII. All electrons vs ECP calculations of orthobenzyne with aug-cc-pVTZ basis set at UPBE0 level.

| | singlet [Ha] | triplet [Ha] | gap [kcal/mol] |
|---|---|---|---|
| all electrons | | | |
| vacuum | -230.705646 | -230.660829 | 28.12 |
| 4w cage | -536.226991 | -536.181805 | 28.35 |
| ECP | | | |
| vacuum | -36.326242 | -36.285998 | 25.25 |
| 4w cage | -105.292890 | -105.252275 | 25.49 |

Table XIV. Uncorrected S-T excitation energies (in kcal/mol) of the orthobenzyne in vacuum (V), 4-water cage (4W) and 30-water cage (30W).

| | DMC | El-FF DMC | El-QDO DMC |
|---|---|---|---|
| S-T (V) | 34.38(5) | ... | ... |
| S-T (4W) | 34.8(1) | 34.58(5) | 34.65(5) |
| S-T (30W) | ... | 34.83(5) | 34.85(8) |


[86] W. K. Hastings, Biometrika **57**, 97 (1970).
[87] C. Filippi and C. J. Umrigar, Phys. Rev. B **61**, R16291 (2000).
[88] M. Casula, C. Attaccalite, and S. Sorella, J. Chem. Phys. **121**, 7110 (2004).
[89] M. Marchi, S. Azadi, C. Casula, and S. Sorella, J. Chem. Phys. **131**, 154116 (2009).
[90] F. Sterpone, L. Spanu, L. Ferraro, S. Sorella, and L. Guidoni, J. Chem. Theory Comput. **4**, 1428 (2008).
[91] I. Kosztin, B. Faber, and K. Schulten, Am. J. Phys. **64**, 633 (1996).
[92] L. Mitas, in *Quantum Monte Carlo Methods in Physics and Chemistry*, NATO ASI Series, Series C, Math. & Phys. Sciences, Vol. C-525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic Publishers, Boston, 1999).
[93] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).
[94] G. Wang, A. Annaberdiyev, C. A. Melton, M. C. Bennett, L. Shulenburger, and L. Mitas, J.Chem. Phys. **151**, 144110 (2019).
[95] M. C. Bennett, G. Wang, A. Annaberdiyev, C. A. Melton, L. Shulenburger, and L. Mitas, J. Chem. Phys. **149**, 104108 (2018).
[96] A. Annaberdiyev, G. Wang, C. A. Melton, M. C. Bennett, L. Shulenburger, and L. Mitas, J. Chem. Phys. **149**, 134108 (2018).
[97] M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, J. Chem. Phys. **147**, 224106 (2017).
[98] A. Zen, J. G. Brandenburg, A. Michaelides, and D. Alfè, J. Chem. Phys. **151**, 134105 (2019).
[99] A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, and D. Alfè, Phys. Rev. B **93**, 241118(R) (2016).
[100] G. M. J. Barca *et al.*, J. Chem. Phys. **152**, 154102 (2020).
[101] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, J. Chem. Phys. **152**, 224108 (2020).
[102] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. **180**, 2175 (2009).
[103] P. K. Biswas and V. Gogonea, J. Chem. Phys. **123**, 164114 (2005).
[104] J. M. Waldrop, B. Song, K. Patkowski, and X. Wang, J. Chem. Phys. **142**, 204307 (2015).

[105] R. Hellmann, B. Jäger, and E. Bich, J. Chem. Phys. **147**, 034304 (2017).
[109] S. Sorella, Phys. Rev. B **64**, 024512 (2001).
[110] S. Sorella, Phys. Rev. B **71**, 241103(R) (2005).
[114] M. P. Metz, K. Szalewicz, J. Sarka, R. Tóbiás, A. G. Császár, and E. Mátyus, Phys. Chem. Chem. Phys. **21**, 13504 (2019).
[116] L. Shirkov and V. Sladek, J. Chem. Phys. **147**, 174103 (2017).
[117] A. Altun, F. Neese, and G. Bistoni, Beilstein J. Org. Chem. **14**, 919 (2018).
[120] J. Shee, E. J. Arthur, S. Zhang, D. R. Reichman, and R. A. Friesner, J. Chem. Theory Comput. **15**, 4924 (2019).
[121] D. A. Sirianni, X. Zhu, D. F. Sitkoff, D. L. Cheney, and C. D. Sherrill, J. Chem. Phys. **156**, 194306 (2022).
[122] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).
[123] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983)
[124] W. L. Jorgensen and D. L. Severance, J. Am. Chem. Soc. **112**, 4768 (1990)