QM*/MM: Hybrid Quantum Mechanical Molecular Mechanical Calculations Using a Multipolar Representation for the Electron Density in the Quantum Mechanical Region for the Electrostatic Embedding

Yuezhi Mao[1] and Yihan Shao[2, a)]

[1)] *Department of Chemistry, University of California, Berkeley, California 94720, United States*

[2)] *Department of Chemistry and Biochemistry, University of Oklahoma, Norman, Oklahoma 73019, United States*

(Dated: 16 December 2016)

Abstract. In this note, we are going to explore local expansions for the electrostatic embedding potential in hybrid quantum mechanical molecular mechanical calculations.

[a)] Author to whom correspondence should be addressed; Electronic mail: yihan.shao@ou.edu

I. INTRODUCTION

Hybrid quantum mechanical molecular mechanical (QM/MM) modeling has become an important tool for molecular and biomolecular simulations, where a core region of interest (called the *system*), such as reaction sites, binding sites, chromophores, solutes and their immediate solvent shells, requires an accurate quantum mechanical (QM) description. The rest of the system (called the *environment*) provides an electrostatic embedding potential to the electron motion in the QM system, and typically also interacts with the system through classical Lennard-Jones or similar terms to account for exchange repulsion and dispersion. In cases where the system and environment are connected through covalent bonds, several approaches using link-atoms, double link atoms, pseudobonds, generalized hybrid orbitals, frozen bond orbitals, can be employed to approximatelyly describe the interface bonds.

In this work, we will focus mainly on the electrostatic embedding potential from the environment, which would necessarily perturb the electron structure of the QM region. The treatment of such electrostatic embedding can be divided into two categories. In the first category, which is the dominant choice today, one represents the electrostatic embedding potential, $\phi(\vec{r})$, from the environment in the atomic orbital (AO) basis for the QM region,

$$V_{\mu\nu} = \int \chi_\mu(\vec{r})\phi(\vec{r})\chi_\nu(\vec{r})d\vec{r} \tag{1}$$

which becomes part of the one-electron effective Hamiltonian (i.e., the Fock matrix). The electron portion of the system-environment electrostatic energy is simply

$$E_V = Tr(PV) \tag{2}$$

where P is the one-particle density matrix in the AO basis. In other words, the electron density interacts explicitly with the electrostatic embedding potential from the environment,

$$E_V = \int \rho(\vec{r})\phi(\vec{r})d\vec{r} \tag{3}$$

In the second category, the electron density adopts a multipolar representation, via atomic charges, dipoles, and quadrupoles, etc (see, for example, Eqs. 22, 23, 24 below). The system-environment electrostatic interaction energy in Eqs. 2 or 3 can then be approximated by interacting local atomic multipoles with the local electrostatic embedding potential, its gradient (electrostatic field), and Hessian (quadrupolar field). Such a multipolar representation

of the QM electron density, called a QM* representation, was reported in the recent implementation of the QM/AMOEBA calculations within the ONETEP software package, where AMOEBA refers to an advanced induced-dipoles-type polarizable molecular mechanical force field. In that implementation, the QM* representation was only adopted for the QM $\mapsto$ MM polarization, namely for the computation of induced dipoles on polarizable atoms in the environment region due to the QM electron density. The MM $\mapsto$ QM polarization was carried out in the conventional way by computing the electrostatic potential of permanent multipoles and induced dipoles in the environment region and using this potential to perturb the QM electron density through Eqs. 1 through 3.

In this work, we will explore the possibility of using the multipolar (QM*) representation of the electron density in the MM $\mapsto$ QM polarization as well. In order to avoid any potential difficulties with computing analytical energy gradients, the simplest approach to obtain the local atomic charges, dipoles and quadrupoles will be adopted, which is based on Mulliken-type decomposition of the electron density. The QM*/MM calculations are attractive to us because of two factors: a) a huge reduction in the cost of computing one-electron integrals in Eq.1 and corresponding contributions to the analytical energy gradient; and (b) a potential use to help formulate new polarizable force fields models that mix responses to changes in the local electrostatic potential, field and quadrupolar field.

In principle, it might be desirable to have a rigorous computation of one-electron integrals in Eq.1. But it is arguably 100% necessary only when: (a) one has an accurate representation for the electron density in the MM region that accounts for the delocalized nature of valence electrons; (b) one has a well-calibrated combination of QM method and basis set for the system-at-hand; and (c) one accounts for other system-environment nonbonded (exchange repulsion and dispersion) interactions in an accurate and not-so-arbitrary manner.

II. THEORY

A. Local Taylor expansion of the MM electrostatic embedding potential

Suppose that there is an MM electrostatic embedding potential, $\phi(\vec{r})$, which might come from MM point charges (q_i at position $\vec{R}_i$)

$$\phi(\vec{r}) = \sum_i \frac{q_k}{|\vec{r} - \vec{R}_k|} \tag{4}$$

or MM multipoles and induced dipoles. A local Taylor expansion can be performed around each nucleus (A),

$$\phi(\vec{r}) = \phi(\vec{A}) + \vec{E} \cdot (\vec{r} - \vec{A}) + \frac{1}{2}(\vec{r} - \vec{A}) \cdot \vec{\vec{G}} \cdot (\vec{r} - \vec{A}) + \cdots \tag{5}$$

where $\vec{E}$ are the electrostatic field

$$\vec{E} = -\sum_k \frac{q_k(\vec{r} - \vec{R}_k)}{|\vec{r} - \vec{R}_k|^3} \tag{6}$$

computed at position $\vec{A}$, whereas $\vec{\vec{G}}$ is the field gradient (also called quadrupolar field)

$$\vec{\vec{G}} = \sum_k q_k \left[\frac{3(\vec{r} - \vec{R}_k) \otimes (\vec{r} - \vec{R}_k)}{|\vec{r} - \vec{R}_k|^5} - \frac{\vec{\vec{1}}}{|\vec{r} - \vec{R}_k|^3} \right] \tag{7}$$

B. Local multipole matrices

The atomic-basis overlap matrix is

$$S_{\mu\nu} = \int \chi_\mu(\vec{r}) \chi_\nu(\vec{r}) d\vec{r} \tag{8}$$

The dipole matrices are

$$D_{i,\mu\nu} = \int \chi_\mu(\vec{r}) (\mathbf{1}_i) \chi_\nu(\vec{r}) d\vec{r}, \quad i = x, y, z \tag{9}$$

where $\mathbf{1}_i$ means x, or y, or z. The quadrupolar matrices are

$$Q_{ij,\mu\nu} = \int \chi_\mu(\vec{r}) (\mathbf{1}_i)(\mathbf{1}_j) \chi_\nu(\vec{r}) d\vec{r}, \quad i = x, y, z \tag{10}$$

It is important to note that the dipolar and quadrupolar matrices are dependent on the position of the origin ($\vec{O}$).

Local overlap matrices (S_{AA} and S_{AB}) are defined as diagonal and off-diagonal blocks of the molecular overlap matrix

$$S_{AA} = \{S_{\mu\nu},\ \mu \in A,\ \nu \in A\} \tag{11}$$

$$S_{AB} = \{S_{\mu\nu},\ \mu \in A,\ \nu \in B\} \tag{12}$$

Local dipole matrices are defined as

$$D^{(A)}_{i,AA} = \{D_{i,\mu\nu}(\vec{O'} = \vec{A}),\ \mu \in A,\ \nu \in A\} \tag{13}$$

$$D^{(A)}_{i,AB} = \{D_{i,\mu\nu}(\vec{O'} = \vec{A}),\ \mu \in A,\ \nu \in B\} \tag{14}$$

where the dipole matrix elements on the right-hand-side correspond to integration values with $\vec{A}$ being the origin. Similarly, one can define the local quadrupolar matrices:

$$Q^{(A)}_{ij,AA} = \{Q_{ij,\ \mu\nu}(\vec{O'} = \vec{A}),\ \mu \in A,\ \nu \in A\} \tag{15}$$

$$Q^{(A)}_{ij,AB} = \{Q_{ij,\ \mu\nu}(\vec{O'} = \vec{A}),\ \mu \in A,\ \nu \in B\} \tag{16}$$

Of course, given the molecular multipolar matrices in Eqs. 8, 9 and 10, one can easily obtain the local dipolar and quadrupolar matrices:

$$D_{i,\ \mu\nu}(\vec{O'} = \vec{A}) = D_{i,\ \mu\nu} - (A_i - O_i)S_{\mu\nu} \tag{17}$$

$$Q_{ij,\ \mu\nu}(\vec{O'} = \vec{A}) = Q_{ij,\ \mu\nu} - (A_i - O_i)D_{j,\mu\nu} - (A_j - O_j)D_{i,\mu\nu} + (A_i - O_i) * (A_j - O_j)S_{\mu\nu} \tag{18}$$

C. QM/MM electrostatic embedding potential matrix

The AO representation of the MM electrostatic embedding potential was defined in Eq.1. Using the Taylor expansion in Eq. 5, its diagonal blocks can be approximated as

$$\tilde{V}_{AA} = \phi(\vec{A})S_{AA} + \vec{E}(\vec{A}) \cdot \vec{M}^{(A)}_{AA} + \frac{1}{2}\vec{\vec{G}}(\vec{A}) \odot \vec{\vec{Q}}^{(A)}_{AA} \tag{19}$$

For the computation of the off-diagonal blocks, local Taylor expansion is performed at both end atoms, and the average is used

$$\begin{aligned}\tilde{V}_{AB} = &\frac{1}{2}\phi(\vec{A})S_{AB} + \frac{1}{2}\vec{E}(\vec{A}) \cdot \vec{M}^{(A)}_{AB} + \frac{1}{4}\vec{\vec{G}}(\vec{A}) \odot \vec{\vec{Q}}^{(A)}_{AB} \\ &+ \frac{1}{2}\phi(\vec{B})S_{AB} + \frac{1}{2}\vec{E}(\vec{B}) \cdot \vec{M}^{(B)}_{AB} + \frac{1}{4}\vec{\vec{G}}(\vec{B}) \odot \vec{\vec{Q}}^{(B)}_{AB}\end{aligned} \tag{20}$$

(For the off-diagonal blocks, a single local Taylor expansion at the midpoint between two atoms was also tried, but it led to much less accurate results.)

When the QM electron density interacts with the approximate embedding potential defined in Eqs. 19 and 20, the electron portion of the electrostatic embedding energy as defined in Eq. 2 can be equivalently written as

$$E_{\tilde{V}} = \sum_A \left[\phi(\vec{A})C^{(A)} + \vec{E}(\vec{A}) \cdot \vec{\mu}^{(A)} + \frac{1}{2}\vec{\vec{G}}(\vec{A}) \odot \vec{\vec{\Theta}}^{(A)}\right] \tag{21}$$

Here, $C^{(A)}$, $\vec{\mu}^{(A)}$ and $\vec{\vec{\Theta}}^{(A)}$ are atomic charges, dipoles, and quadrupoles

$$C^{(A)} = \int \rho_A(\vec{r}) d\vec{r} \tag{22}$$

$$\vec{\mu}^{(A)} = \int (\vec{r} - \vec{A}) \rho_A(\vec{r}) d\vec{r} \tag{23}$$

$$\vec{\vec{\Theta}}^{(A)} = \int \left[(\vec{r} - \vec{A}) \otimes (\vec{r} - \vec{A}) \right] \rho_A(\vec{r}) d\vec{r} \tag{24}$$

computed using the Mulliken decomposition of the electron density:

$$\rho_A(\vec{r}) = \sum_{\mu,\nu \in A} P^{\mu\nu} \chi_\mu(\vec{r}) \chi_\nu(\vec{r}) + \frac{1}{2} \sum_{\mu \in A, \nu \notin A} P^{\mu\nu} \chi_\mu(\vec{r}) \chi_\nu(\vec{r}) \tag{25}$$

In other words, with the adoption of Eqs. 19 and 20, *the QM electron density will be represented as a QM* multipolar representation for its interaction with the MM electrostatic environment.*

III. QM*/MM RESULTS

A. Computational details

Calculations are performed on the water dimer molecule

```
O  -1.486845   0.125051   0.000000
H  -1.861405  -0.757772   0.000000
H  -0.540931  -0.032798  -0.000000
O   1.380795  -0.119842   0.000000
H   1.792649   0.307636  -0.753200
H   1.792649   0.307636   0.753200
```

with the first water (hydrogen bond donor) treated with B3LYP/6-31+G* or B3LYP/6-311++G**, and the second water described with the TIP3P model in QM/MM calculations.

Four models will be used:

- **Coulomb embedding.** The second water molecule is treated as an isolated molecule using the same level theory. Its density is used to provide an electrostatic potential for the first water molecule. The Hartree-Fock exchange effect between the two water molecules is NOT accounted for.

- **Standard QM/MM.** The embedding potential matrix elements are computed according to Eq. 1. It is referred to "MM w/o local expansion" or "no LE" later.

- **LE-1**. The approximate matrix elements are computed according to Eqs. 19 and 20, but with the second-order term involving $\vec{\vec{G}}$ removed.

- **LE-2**. The approximate matrix elements are computed up to the second order according to Eqs. 19 and 20.

B. Embedding potential matrix with B3LYP/6-31G* and B3LYP/6-311++G**

FIG. 1: Embedding potential matrix from B3LYP/6-31G* calculations. The matrix is 23-by-23 in size. LE-2 is shown to well reproduces the standard QM/MM matrix elements.

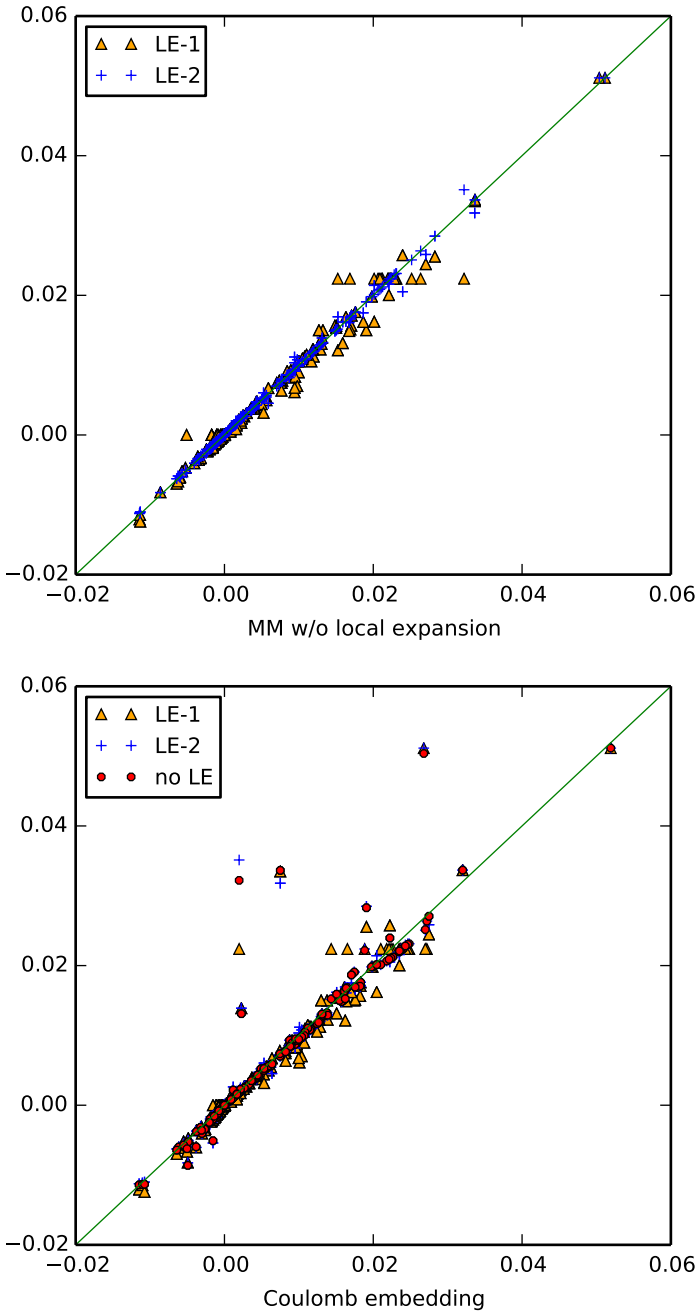


FIG. 2: Embedding potential matrix from B3LYP/6-311++G** calculations. The matrix is 36-by-36 in size. LE-2, in general, well reproduces the standard QM/MM matrix elements. But for the (33,33)-th element, LE-2 value (0.051) deviates significantly from the standard QM/MM value (0.026). The Coulomb embedding value is -0.016. The corresponding basis function is a very diffuse s function (exponent = 0.036) on the hydrogen atom (hydrogen bond donor).

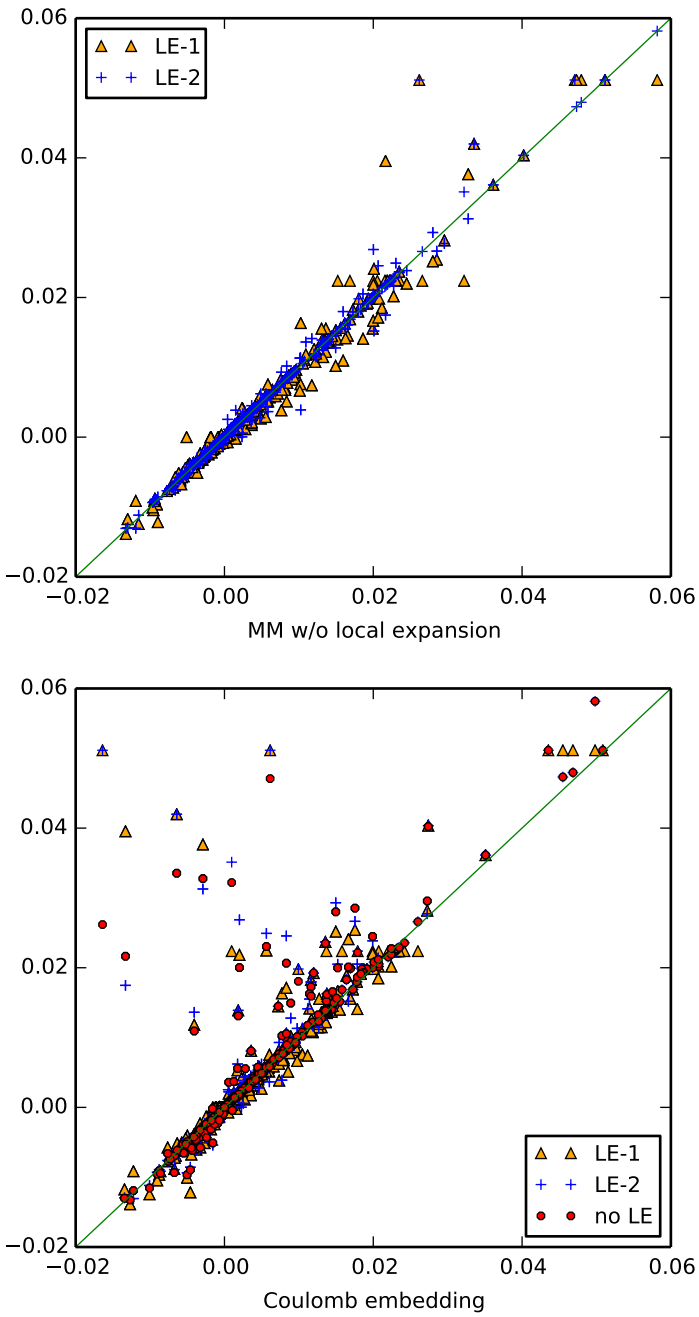


FIG. 3: QM/MM permanent electrostatics and polarization energies (in kcal/mol) for the water molecule (described with B3LYP/6-31G*) in 1000 different solvent configurations (each with 1687 TIP3P water molecules). QM*/MM values with the LE-1 and LE-2 approximations are shown against standard QM/MM results.

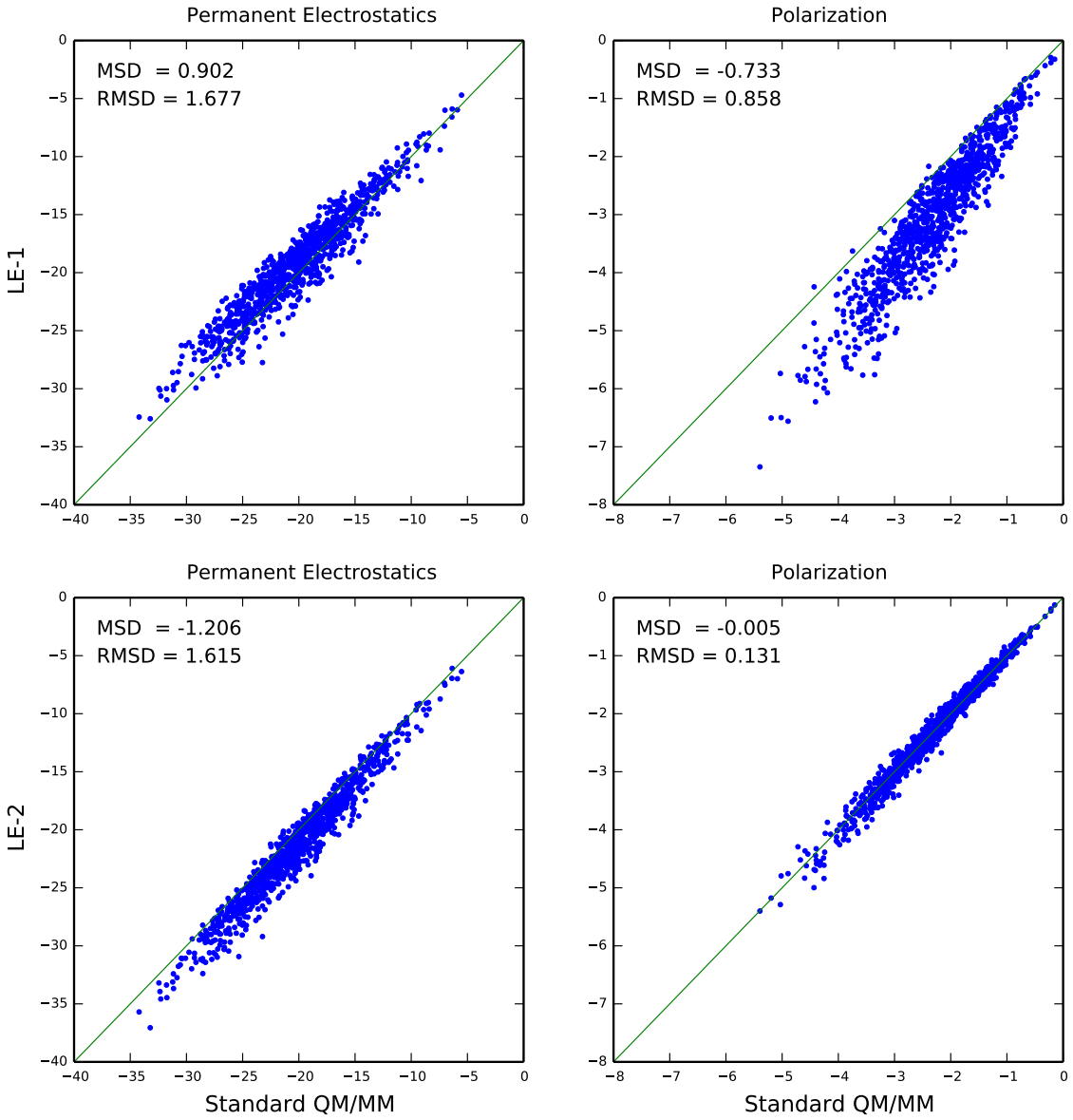


FIG. 4: QM/MM permanent electrostatics and polarization energies (in kcal/mol) for the water molecule (described with B3LYP/6-311++G**) in 1000 different solvent configurations (each with 1687 TIP3P water molecules). QM*/MM values with the LE-1 and LE-2 approximations are shown against standard QM/MM results.

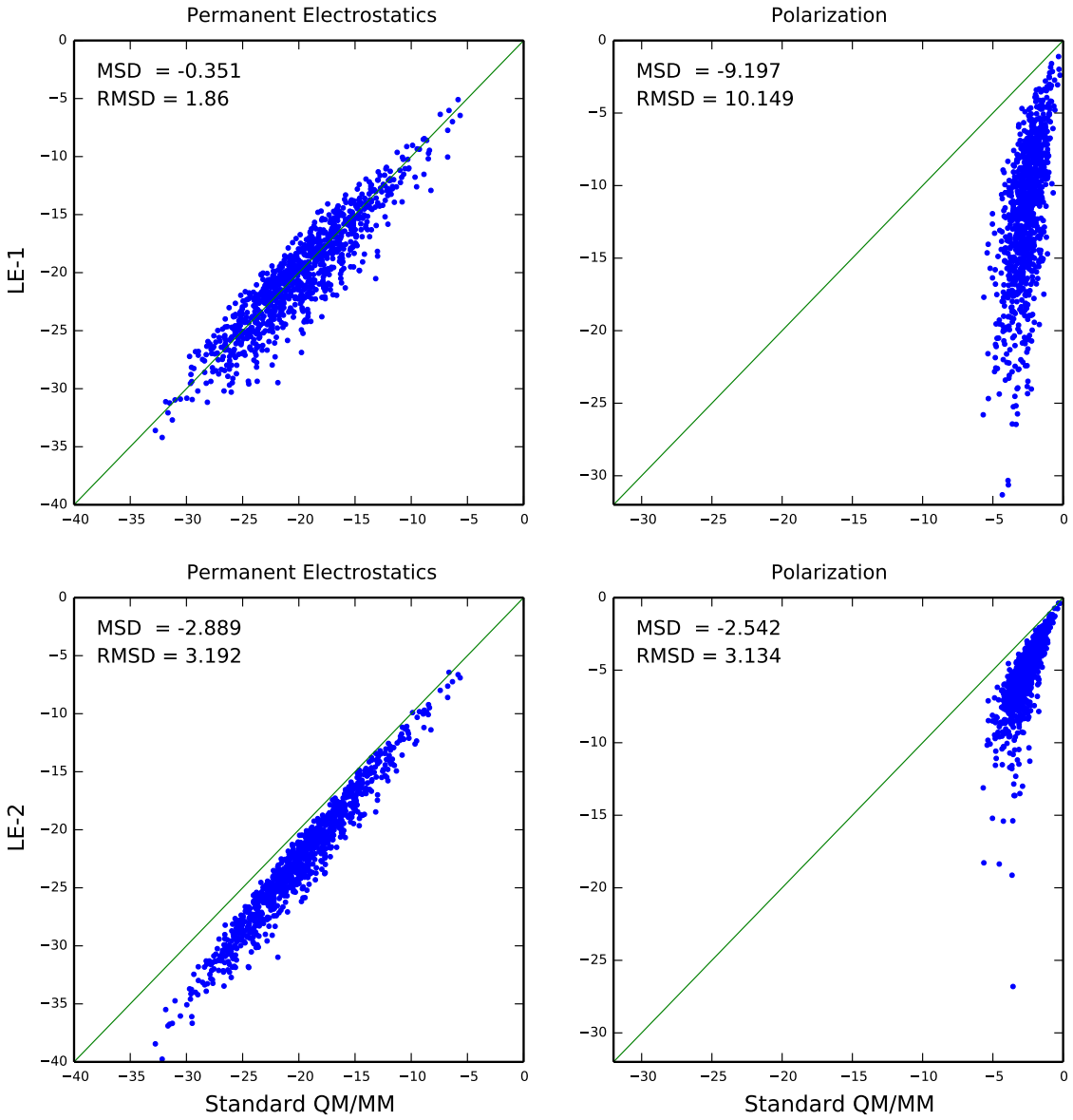


FIG. 5: QM/MM permanent electrostatics and polarization energies (in kcal/mol) for the methanol molecule (described with B3LYP/6-31G*) in 1000 different solvent configurations (each with 1687 TIP3P water molecules). QM*/MM values with the LE-1 and LE-2 approximations are shown against standard QM/MM results.

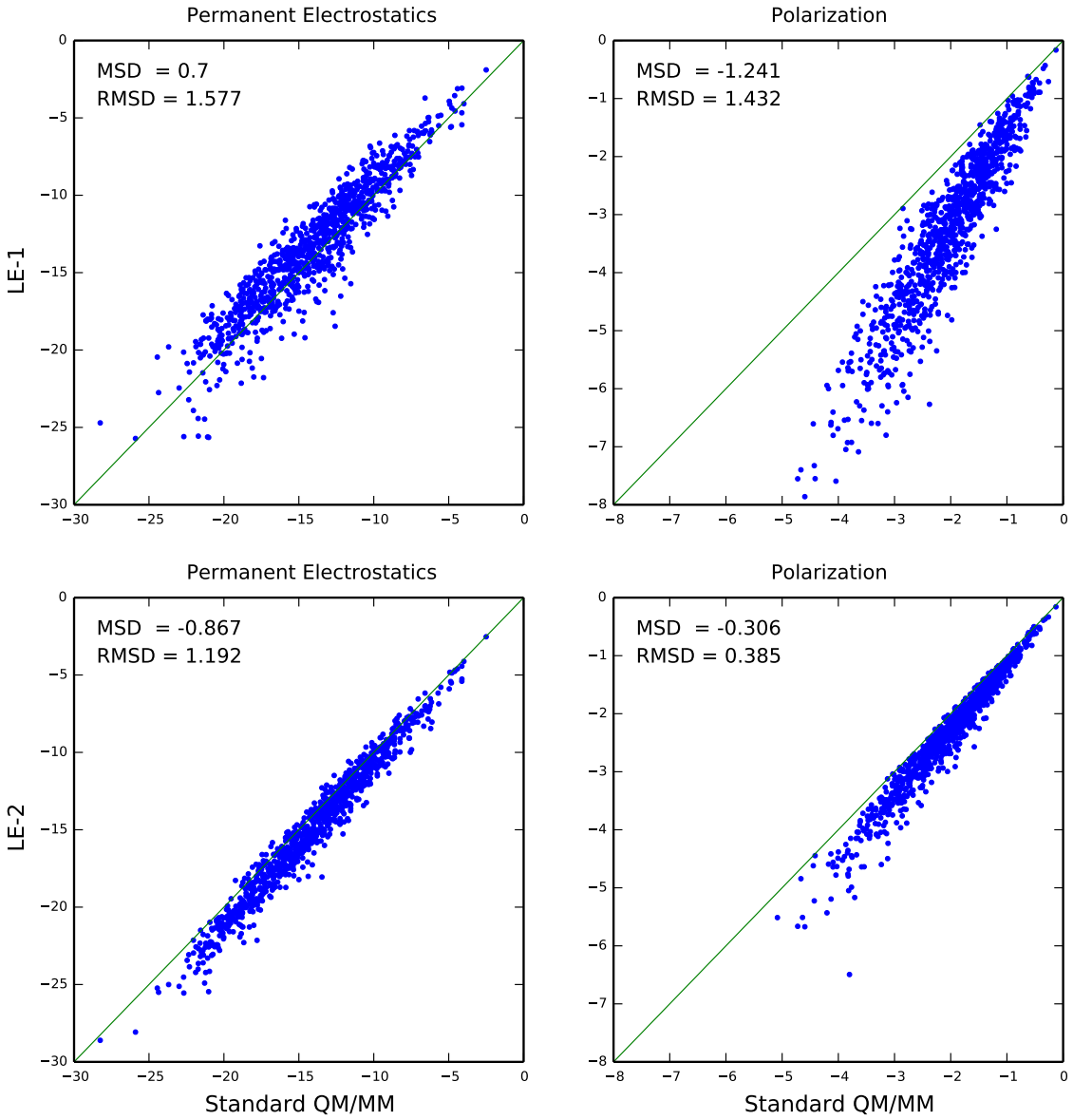


FIG. 6: QM/MM permanent electrostatics and polarization energies (in kcal/mol) for the acetamide molecule (described with B3LYP/6-31G*) in 1000 different solvent configurations (each with 1687 TIP3P water molecules). QM*/MM values with the LE-1 and LE-2 approximations are shown against standard QM/MM results.

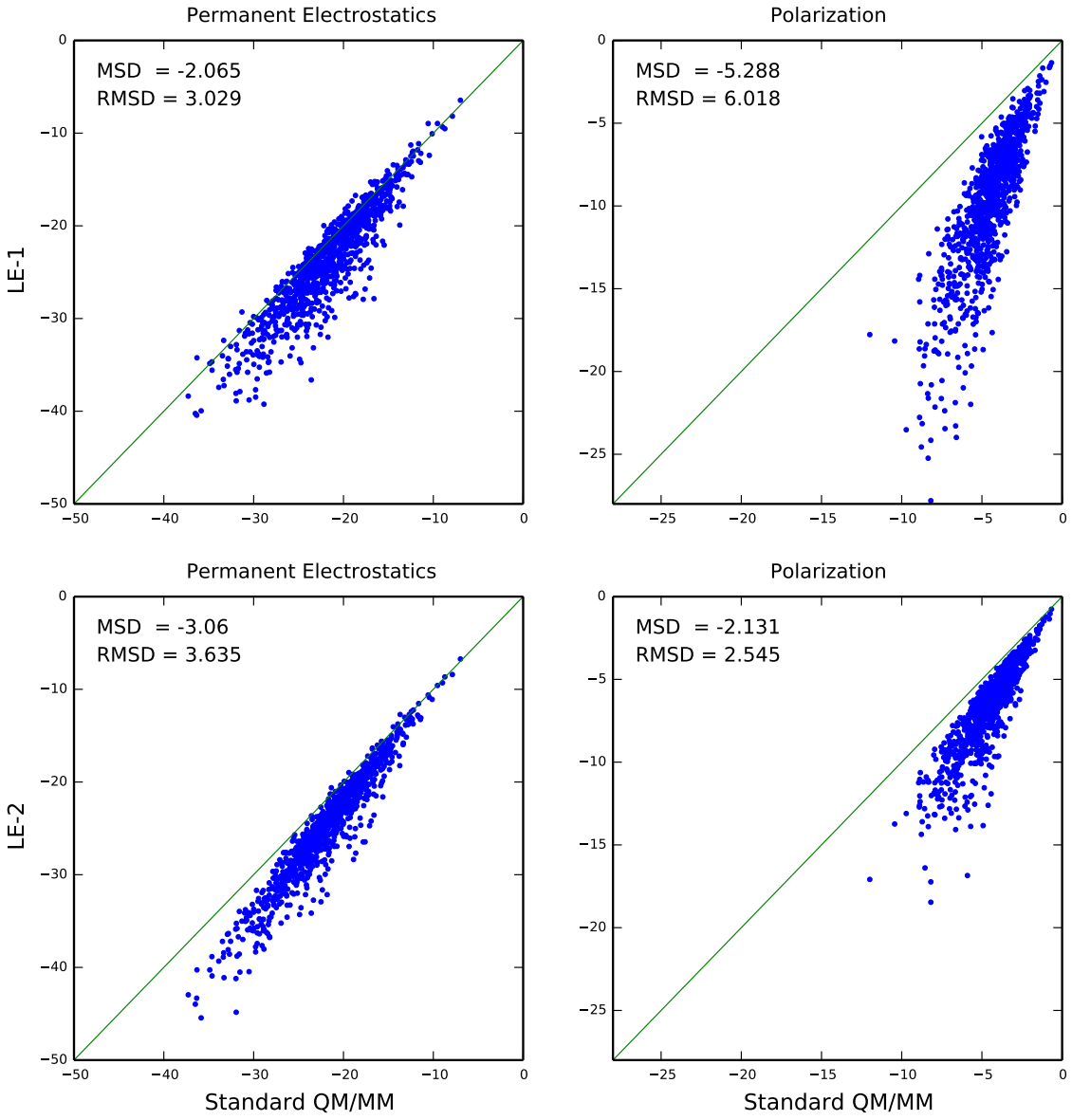


FIG. 7: QM/MM permanent electrostatics and polarization energies (in kcal/mol) for the methanethiol molecule (described with B3LYP/6-31G*) in 1000 different solvent configurations (each with 1687 TIP3P water molecules). QM*/MM values with the LE-1 and LE-2 approximations are shown against standard QM/MM results.

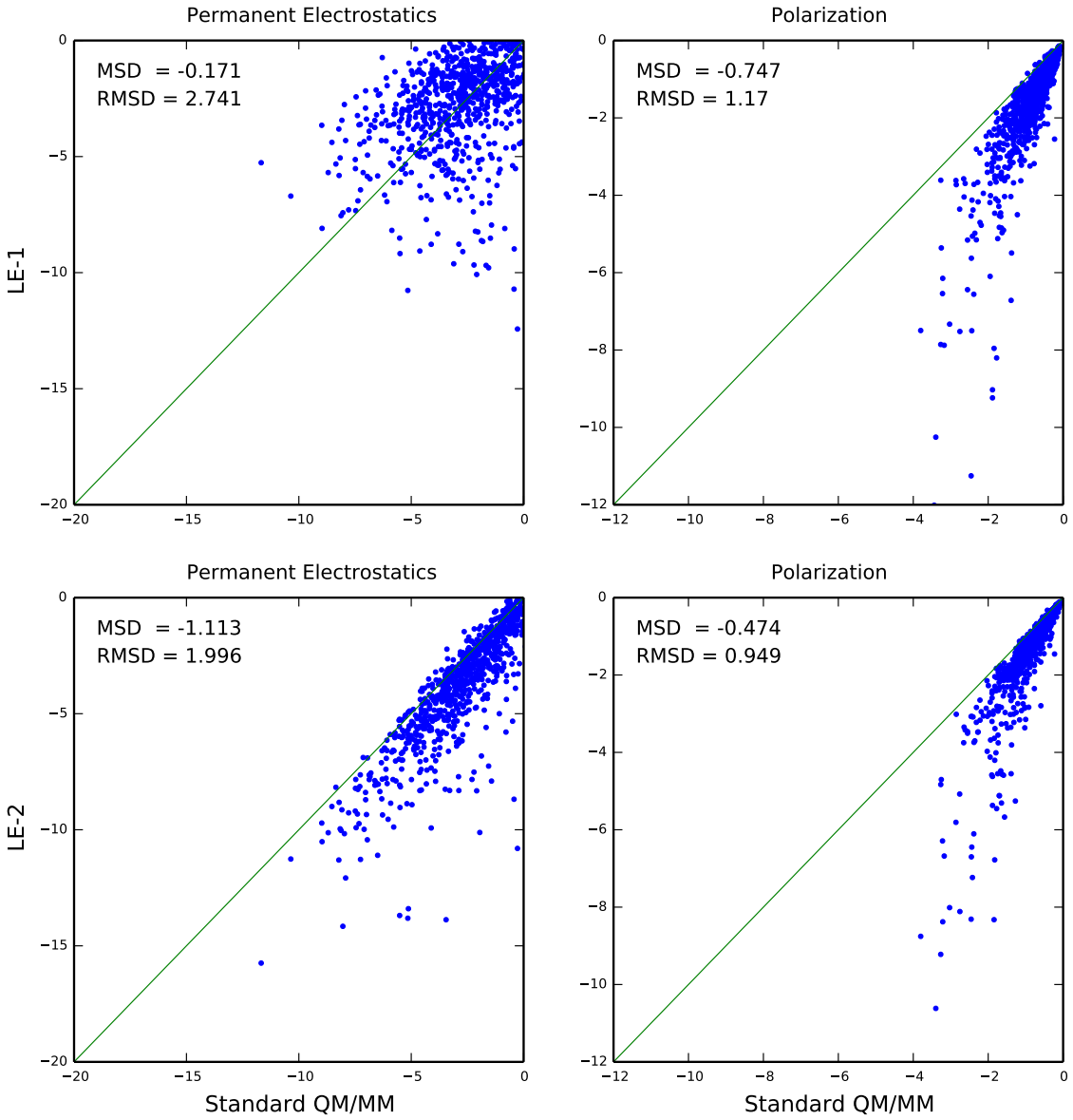
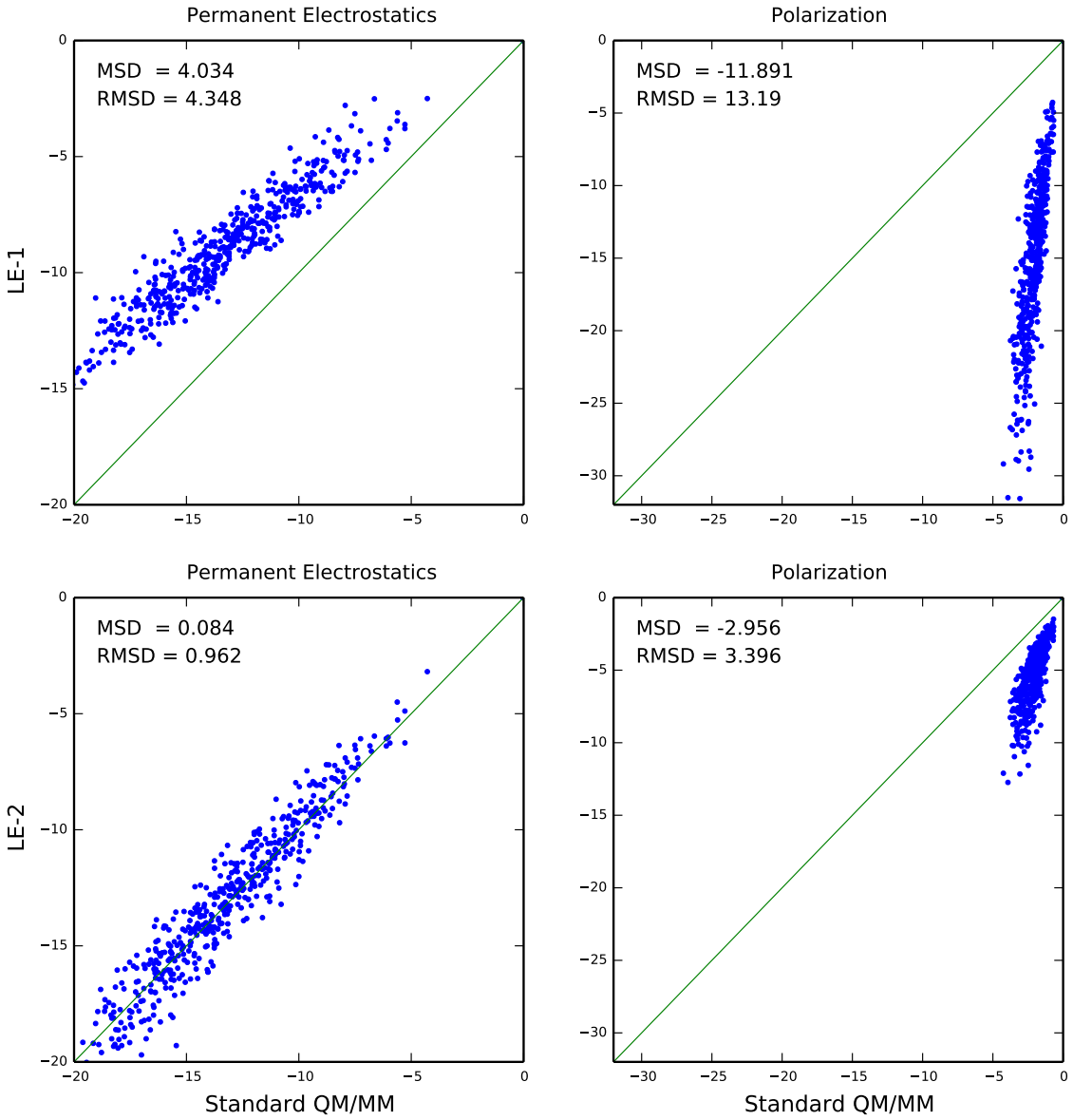


FIG. 8: QM/MM permanent electrostatics and polarization energies (in kcal/mol) for the aniline molecule (described with B3LYP/6-31G*) in 516 different solvent configurations (each with 1687 TIP3P water molecules). QM*/MM values with the LE-1 and LE-2 approximations are shown against standard QM/MM results.



IV. EFFECT OF ORTHONORMALIZATION ON THE DENSITY

For a two-fragment system (AB), let us use $C_A^{(0)}$ and $C_B^{(0)}$ to represent occupied orbitals for each isolated fragment. The overlap matrix between these two sets of fragments are:

$$S = \begin{pmatrix} I & S_{AB} \\ S_{BA} & I \end{pmatrix} \tag{26}$$

where

$$S_{AB} = C_A^{(0),T} S_{ao} C_B^{(0)} \tag{27}$$

which is non-zero for two fragments separated at a finite distance. S_{BA} is its transpose.

One can orthonormalize these orbitals "symmetrically" via

$$\begin{pmatrix} C'_A & C'_B \end{pmatrix} = \begin{pmatrix} C_A^{(0))} & 0 \\ 0 & C_B^{(0))} \end{pmatrix} \begin{pmatrix} I & S_{AB} \\ S_{BA} & I \end{pmatrix}^{-\frac{1}{2}} \tag{28}$$

where fragment MOs on both fragments are changed.

Alternatively, one can take a "non-symmetrical" scheme, where the fragment MOs on B are fixed. Only MOs on fragment A are changed,

$$\begin{pmatrix} C''_A & C_B^{(0)} \end{pmatrix} = \begin{pmatrix} C_A^{(0)} & 0 \\ 0 & C_B^{(0)} \end{pmatrix} \begin{pmatrix} (I - S_{AB}S_{BA})^{-\frac{1}{2}} & 0 \\ -S_{BA}(I - S_{AB}S_{BA})^{-\frac{1}{2}} & I \end{pmatrix} \tag{29}$$

Namely,

$$C''_A = \left[C_A^{(0)} - C_B^{(0)} S_{BA} \right] (I - S_{AB}S_{BA})^{-\frac{1}{2}} \tag{30}$$

which clearly satisfy the orthonormal conditions

$$\left(C''_A\right)^T S_{ao} C''_A = I \tag{31}$$

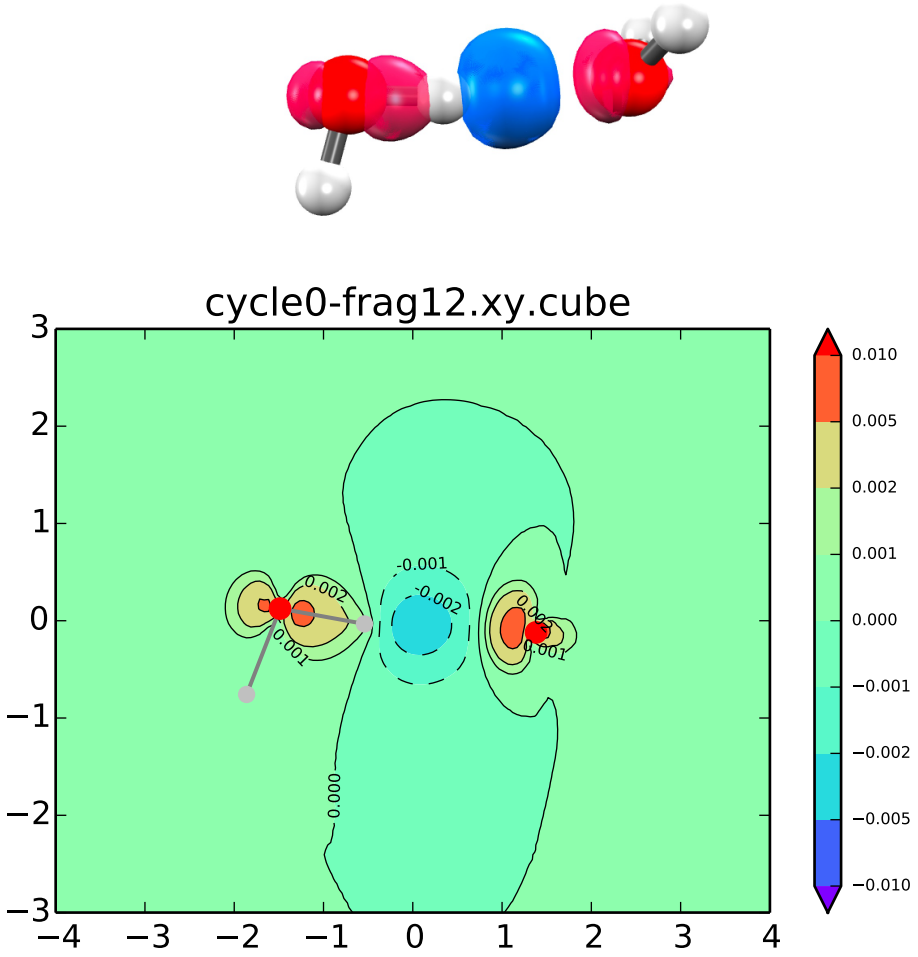
$$C_B^{(0)} S_{ao} C''_A = 0 \tag{32}$$

This "non-symmetrical" scheme would resemble QM/MM calculations where fragment A is treated quantum mechanically, and fragment B is described with a non-polarizable force field.

Eq. 30 poses a technical challenge to practical QM/MM calculations, because it requires basis functions on fragment B to be included in the QM calculation on fragment A. It will be worthwhile to find which basis functions on fragment B matter the most.

Eq. 30 seems to indicate that a small fraction of electron is transferred from (occupied orbitals of) fragment A to fragment B. This is actually not true — there is no net charge transfer between the two fragments. Without opening up the virtual space to add electrons, the total number of electrons in each fragment cannot change. To add to our confusion, though, the Mulliken charges within each fragment do seem to change.

FIG. 9: Density changes to water dimer from orthonormalizing fragment MOs.



B3LYP/6-31G* calculations were performed on water dimer, and the density change from orthonormalizing the fragment MOs are shown in Fig. 9. Clearly, a small amount of electron change is depleted from the hydrogen bond area (between the H and O atoms), and transferred to the two oxygen atoms. By diagonalizing the difference in the fragment A density matrix (due to MO changes in Eq. 30), the amount of charge is estimated to be

around 0.1 e.

A. Further Thoughts

The MO representation of the density matrix after the orthogonalization, to 2nd order, can be expressed

$$P = \begin{pmatrix} I + S_{AB}S_{BA} & -S_{AB} \\ -S_{BA} & I + S_{BA}S_{AB} \end{pmatrix} \tag{33}$$

which seems to work well for water dimer at the equilbrium bond length. To compute the Fock matrix using this density matrix, we need

$$P_{AA} \longrightarrow \sum_{i,j \in A} (\mu\nu||ij) \left[\delta_{ij} + \sum_{k \in B} S_{ik}S_{jk} \right] \tag{34}$$

$$P_{AB} \longrightarrow \sum_{i \in A, k \in B} (\mu\nu||ik) S_{ik} \tag{35}$$

$$P_{BB} \longrightarrow \sum_{k,l \in B} (\mu\nu||kl) \left[\delta_{kl} + \sum_{i \in A} S_{ik}S_{il} \right] \tag{36}$$