

# ONIOM-based QM:QM electronic embedding method using Löwdin atomic charges: Energies and analytic gradients

Nicholas J. Mayhall,<sup>1</sup> Krishnan Raghavachari,<sup>1,a)</sup> and Hrant P. Hratchian<sup>2</sup>

<sup>1</sup>Department of Chemistry, Indiana University, Bloomington, Indiana 47405, USA

<sup>2</sup>Gaussian, Inc., Wallingford, Connecticut 06492, USA

(Received 16 December 2009; accepted 19 January 2010; published online 15 March 2010)

In this work, we report a new quantum mechanical:quantum mechanical (QM:QM) method which provides explicit electronic polarization of the high-level region by using the Löwdin atomic charges from the low-level region. This provides an embedding potential which naturally evolves with changes in nuclear geometry. However, this coupling of the high-level and low-level regions introduces complications in the energy gradient evaluation. Following previous work, we derive and implement efficient gradients where a single set of self-consistent field response equations is solved. We provide results for the calculation of deprotonation energies of a hydroxylated spherosiloxane cluster ( $\text{Si}_8\text{O}_{12}\text{H}_7\text{OH}$ ) and the dissociation energy of a water molecule from a  $[\text{ZnIm}_3(\text{H}_2\text{O})]^{2+}$  complex. We find that the Löwdin charge embedding model provides results which are not only an improvement over mechanical embedding (no electronic embedding) but which are also resistant to large overpolarization effects which occur more often with Mulliken charge embedding. Finally, a scaled-Löwdin charge embedding method is also presented which provides a method for fine tuning the extent of electronic polarization. © 2010 American Institute of Physics. [doi:10.1063/1.3315417]

## I. INTRODUCTION

The extent to which one can balance the relevance of a model chemical system with the accuracy of the theoretical method often determines whether a particular problem may be treated computationally. While sophisticated quantum chemical calculations are now capable of routinely providing accurate reaction energies, the poor scaling of *ab initio* methods with system size restricts applications to molecular systems of small to modest sizes.<sup>1-3</sup> To address this issue of computational complexity, hybrid energy methods<sup>4-6</sup> divide the molecule up into different regions (e.g., region I and region II) which can be treated with different computational methods. A particularly useful hybrid energy method called ONIOM (our own N-layer integrated molecular orbital molecular mechanics), which was developed by Morokuma and co-workers,<sup>7-13</sup> uses the following energy definition:

$$E_{\text{ONIOM}} = E_{\text{RL}} - E_{\text{ML}} + E_{\text{MH}}, \quad (1)$$

where the subscripts RL, ML, and MH denote the real system (region I+region II)-low level energy, the model system (region I)-low level energy, and the model system (region I)-high level energy, respectively.

If no electronic coupling terms are added to  $E_{\text{ML}}$  or  $E_{\text{MH}}$ , then this is referred to as *mechanical embedding* (ME),<sup>14</sup> and an obvious improvement is to explicitly include the electronic charge distribution of region II in the model system calculations. This is referred to as *electronic embedding* (EE).<sup>14</sup> If the ONIOM-EE method being used is of the type quantum mechanics:molecular mechanics with electronic embedding (QM:MM-EE), then molecular mechanics meth-

ods with fixed charges or more sophisticated models which employ polarizable force fields can be used to define the embedding potential for the QM region.<sup>15-17</sup> Gao and Xie<sup>18-20</sup> have also developed a QM:MM-EE method which obtains MM point charges from a QM wave function. In this approach, atomic charges are obtained from a Mulliken population analysis of a Hartree-Fock (HF) wave function (with the resulting charges scaled by a factor of 2.2).

Recently, we have developed a QM:QM-EE method within the ONIOM framework using Mulliken (or asymmetric Mulliken) atomic charges.<sup>21,22</sup> We will refer to this method as QM:QM-EE(Mull). In this approach, the region II Mulliken atomic point charges<sup>23</sup> from the RL calculation are included as one-electron potentials in the model-system low-level and high-level Hamiltonians. While energy evaluation is relatively straightforward to implement, analytic gradients for this class of models require care as the embedding point charges vary with the molecular geometry. Because the region I derivatives depend on region II atomic charge derivatives, the MH and ML gradient evaluations are coupled to the RL gradient evaluations. As a result the energy derivative expressions include RL density matrix derivatives. By judicious grouping of terms, the MH and ML gradient components of the total ONIOM gradient are no more complicated than in the ME case; however, the RL contribution to the ONIOM derivative takes on a post self-consistent-field (SCF) form, even when the RL model chemistry is a self-consistent field level of theory [e.g., density functional theory (DFT) or HF], and gives rise to a set of coupled-perturbed HF (CPHF) equations.<sup>24</sup> As in the case of post-

<sup>a)</sup>Electronic mail: kraghava@indiana.edu.

SCF derivative theories, application of the  $z$ -vector method of Handy and Schaefer<sup>25</sup> reduces the CPHF work to a single set of SCF response equations.

The QM:QM-EE(Mull) method has recently seen success in application to work related to dye-sensitized solar cells.<sup>26</sup> In that study, EE was needed for sufficient modeling of eosin-loaded ZnO films. While this suggests promise for the Mulliken electronic embedding, we have recently applied the QM:QM-EE(Mull) method to a set of challenging test cases, and it was observed that certain molecules may call for a larger charge separation than is obtained using Mulliken charges, which splits the shared density equally between atoms. To address this, a variant of the QM:QM-EE(Mull) was developed which partitioned the shared density asymmetrically between atoms. This has been termed generalized asymmetric Mulliken embedding (GAME).<sup>22</sup> Besides the equal partitioning of density between atoms, Mulliken charges exhibit many undesirable features, which may make them less attractive for use in defining an embedding potential. Because the population analysis is done in a nonorthogonal basis, populations can become larger than two or even negative.<sup>27</sup> By using an orthogonal basis, the Löwdin population analysis<sup>28</sup> corrects both of these aspects, making for a more theoretically sound set of atomic point charges.

In addition to the theoretical motivations, there are a number of practical examples which suggest an EE model using Löwdin charges might be useful. In work by Wu and Van Voorhis,<sup>29</sup> Löwdin charges have been very useful in defining a regional charge to set up a constraint matrix for studying charge transfer chemistry with constrained-DFT methods. Additionally, we have evaluated the Löwdin and Mulliken charges for a variety of systems, and have observed that the Löwdin atomic charges are almost always smaller in magnitude than the corresponding atomic charges obtained from a Mulliken population analysis. This would suggest the polarization effects would tend to be smaller when using Löwdin charges as opposed to Mulliken charges, and this may be useful for treating systems which may be overpolarized with the QM:QM-EE(Mull) method. In this paper we report the development and implementation of a QM:QM-EE method which polarizes the model system by a field of Löwdin atomic charges rather than Mulliken atomic point charges.

## II. METHODS

We start with the following definitions for Mulliken,  $q_A^{\text{Mulliken}}$ , and Löwdin,  $q_A^{\text{Löwdin}}$ , atomic charges

$$q_A^{\text{Mulliken}} = z_A - \sum_{\alpha \in A} \sum_{\beta} P_{\alpha\beta} S_{\beta\alpha}, \quad (2)$$

$$q_A^{\text{Löwdin}} = z_A - \sum_{\alpha \in A} \sum_{\beta\gamma} (S^{1/2})_{\alpha\beta} P_{\beta\gamma} (S^{1/2})_{\gamma\alpha}, \quad (3)$$

where  $\mathbf{P}$ ,  $\mathbf{S}$ , and  $\mathbf{z}$  are the density matrix, overlap matrix, and nuclear charges. The only aspect of the QM:QM-EE(Mull) model which changes upon going to Löwdin embedding is the form of  $q_A$ . Therefore, the expression for the *effective*

*real-system gradient* which is given in the Mulliken embedding paper is still<sup>21</sup>

$$\tilde{E}_{\text{RL}}^x = E_{\text{RL}}^x + \tilde{E}_{\text{emb}}^x = E_{\text{RL}}^x - \sum_A \phi_A^\Delta q_A^x, \quad (4)$$

where  $\phi_A^\Delta$  has previously been defined in Eq. (17) in the original Mulliken embedding paper.<sup>21</sup> To find the gradient of the Löwdin atomic charges defined in Eq. (3) we must differentiate  $\mathbf{S}^{1/2}$ .<sup>30,31</sup> Using the eigenvectors,  $\mathbf{U}$ , and eigenvalues,  $\Lambda$ , of  $\mathbf{S}$ , and differentiating the identity  $\mathbf{S} = \mathbf{S}^{1/2} \mathbf{S}^{1/2}$  yields

$$\frac{d\mathbf{S}}{d\mathbf{x}} = \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}} \mathbf{S}^{1/2} + \mathbf{S}^{1/2} \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}}. \quad (5)$$

Using the eigenvectors of  $\mathbf{S}$  to substitute for  $\mathbf{S}^{1/2}$ ,

$$\frac{d\mathbf{S}}{d\mathbf{x}} = \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}} \mathbf{U} \Lambda^{1/2} \mathbf{U}^T + \mathbf{U} \Lambda^{1/2} \mathbf{U}^T \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}}. \quad (6)$$

Left multiplying by  $\mathbf{U}^T$  and right multiplying by  $\mathbf{U}$ ,

$$\mathbf{U}^T \frac{d\mathbf{S}}{d\mathbf{x}} \mathbf{U} = \mathbf{U}^T \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}} \mathbf{U} \Lambda^{1/2} \mathbf{U}^T \mathbf{U} + \mathbf{U}^T \mathbf{U} \Lambda^{1/2} \mathbf{U}^T \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}} \mathbf{U}. \quad (7)$$

Since  $\mathbf{U}^T \mathbf{U} = \mathbf{1}$ ,

$$\frac{\widetilde{d\mathbf{S}}}{d\mathbf{x}} = \frac{\widetilde{d\mathbf{S}^{1/2}}}{d\mathbf{x}} \Lambda^{1/2} + \Lambda^{1/2} \frac{\widetilde{d\mathbf{S}^{1/2}}}{d\mathbf{x}}, \quad (8)$$

where we have represented the derivatives of  $\mathbf{S}$  and  $\mathbf{S}^{1/2}$  in the overlap matrix eigenspace, denoted by a tilde. Rearranging we obtain the expression

$$\left( \frac{d\mathbf{S}^{1/2}}{d\mathbf{x}} \right)_{pq} = \sum_{ijkl} u_{pi} u_{kj} \left( \frac{d\mathbf{S}}{d\mathbf{x}} \right)_{kl} u_{lj} (\sqrt{\lambda_j} + \sqrt{\lambda_i})^{-1} u_{qj}, \quad (9)$$

where  $u_{pq}$  and  $\lambda_i$  are elements of the  $\mathbf{U}$  and  $\Lambda$  matrices. To obtain the expression for  $q_A^x$ , we define a vector  $\hat{n}$  such that  $\hat{n}_\mu = 1$ , if  $\mu \in A$ , otherwise it is zero. This lets us write  $q_A$  in terms of a trace,  $q_A = z_A - \text{tr}(\hat{n} \mathbf{S}^{1/2} \mathbf{P} \mathbf{S}^{1/2})$ . Differentiating this yields

$$\begin{aligned} q_A^x &= -\text{tr}(\hat{n} \mathbf{S}^{(1/2)x} \mathbf{P} \mathbf{S}^{1/2}) - \text{tr}(\hat{n} \mathbf{S}^{1/2} \mathbf{P} \mathbf{S}^{(1/2)x}) - \text{tr}(\hat{n} \mathbf{S}^{1/2} \mathbf{P} \mathbf{S}^{(1/2)x}) \\ &= -\text{tr}([\mathbf{P} \mathbf{S}^{1/2} \hat{n} + \hat{n} \mathbf{S}^{1/2} \mathbf{P}] \mathbf{S}^{(1/2)x}) - \text{tr}(\mathbf{S}^{1/2} \hat{n} \mathbf{S}^{1/2} \mathbf{P}^x) \\ &= -\text{tr}(\mathbf{M} \mathbf{S}^{(1/2)x}) - \text{tr}(\mathbf{S}^{1/2} \hat{n} \mathbf{S}^{1/2} \mathbf{P}^x), \end{aligned} \quad (10)$$

where

$$M_\mu = \sum_\lambda P_{\mu\lambda} (S^{1/2})_\lambda \hat{n} + \sum_\lambda \hat{n}_\mu (S^{1/2})_{\mu\lambda} P_\lambda. \quad (11)$$

Using Eq. (9) we have two terms which are multiplied with  $\mathbf{S}^x$  and  $\mathbf{P}^x$ ,

$$\begin{aligned} q_A^x &= -\sum_\mu M_\mu \sum_{pq} \sum_{\alpha\beta} \frac{u_{p\alpha} u_{q\beta}}{\sqrt{\lambda_q} + \sqrt{\lambda_p}} S_{\alpha\beta}^x \\ &\quad - \sum_{\alpha\beta} \sum_{\gamma \in A} (S^{1/2})_{\beta\gamma} (S^{1/2})_{\gamma\alpha} P_{\alpha\beta}^x. \end{aligned} \quad (12)$$

Defining atomic orbital (AO) density matrix derivative elements in terms of their backtransformed molecular orbital (MO) representations, we obtain an expression analogous to that obtained in Eq. (19) of Ref. 21. Again, we recast the real

system energy gradient expression in the post-SCF gradient form<sup>32</sup> as

$$\begin{aligned} \tilde{E}_{\text{RL}}^x = & \sum_{\alpha\beta\gamma} \text{emb}_{\alpha\beta\gamma} (\alpha\beta\gamma)^x + \sum_{\alpha\beta} P_{\alpha\beta}^{\text{emb}} H_{\alpha\beta}^x + \sum_{\alpha\beta} W_{\alpha\beta}^{\text{emb}} S_{\alpha\beta}^x \\ & + V_{\text{nuc}}^x, \end{aligned} \quad (13)$$

where the effective embedding density and energy-weighted density matrices are a sum of the corresponding HF matrix and embedding correction (i.e.,  $\mathbf{P}^{\text{emb}} = \mathbf{P}^{\text{HF}} + \mathbf{P}^{\Delta}$  and  $\mathbf{W}^{\text{emb}} = \mathbf{W}^{\text{HF}} + \mathbf{W}^{\Delta}$ ).  $\text{emb}_{\alpha\beta\gamma}$  is the embedding effective two-particle density matrix and is written as a sum of products of the HF density matrix and the embedding correction. Using the new definitions for the Lagrangian and the energy-weighted density matrix, the general post-SCF form of Eq. (13) can still be used. From Eq. (12), we can obtain the corrections to the gradient expressions that arise from the Löwdin atomic charge embedding. The new Lagrangian takes a form very similar to that found in the Mulliken atomic charge embedding gradients, and is given as

$$L_{ai} = \sum_{\alpha\beta} \sum_{\gamma} \phi_{\gamma}^{\Delta}(S^{1/2})_{\beta\gamma}(S^{1/2})_{\gamma\alpha}(C_{\alpha i}C_{\beta a} + C_{\alpha a}C_{\beta i}), \quad (14)$$

where we have defined a vector  $\phi_{\alpha}^{\Delta}$  where the  $\alpha$  index runs over RL basis functions. When basis function  $\alpha$  is center on an atomic center A residing in region II,  $\phi_{\alpha}^{\Delta} = \phi_A^{\Delta}$ , otherwise  $\phi_{\alpha}^{\Delta} = 0$ . The only difference between the Mulliken and Löwdin embedding is that, here,  $\phi_{\alpha}^{\Delta}$  shares an index with the inside of the product of  $\mathbf{S}^{1/2}\mathbf{S}^{1/2}$ , thus forming a truncated overlap matrix. To obtain the correction to the energy-weighted density matrix, we first define an intermediate

$$\tilde{M}_{\mu} = \sum_A \phi_A^{\Delta} M_{\mu} = \sum_{\lambda} P_{\mu\lambda}(S^{1/2})_{\lambda} \phi^{\Delta} + \sum_{\lambda} \phi_{\mu}^{\Delta}(S^{1/2})_{\mu\lambda} P_{\lambda}. \quad (15)$$

Therefore,

$$W_{\alpha\beta}^{\Delta} = \sum_{\mu} \sum_{pq} \frac{\tilde{M}_{\mu} u_p u_{ap} u_{\beta q} u_{\mu q}}{\sqrt{\lambda_q} + \sqrt{\lambda_p}} + \sum_{pq} C_{ap} C_{\beta q} \bar{W}_{pq}^{\Delta}, \quad (16)$$

where  $\bar{W}_{pq}^{\Delta}$  can be broken into the occupied-occupied (oo), occupied-virtual (ov), and virtual-virtual (vv) parts

$$\bar{W}_{ij}^{\Delta} = - \sum_{\alpha\beta} \sum_{\gamma} \phi_{\gamma}^{\Delta}(S^{1/2})_{\beta\gamma}(S^{1/2})_{\gamma\alpha} C_{\alpha i} C_{\beta j} - \sum_{ak} P_{ak}^{\Delta} (aj \quad ki), \quad (17)$$

and

$$\bar{W}_{ai}^{\Delta} = - P_{ai}^{\Delta} \quad i, \quad (18)$$

$$\bar{W}_{ab}^{\Delta} = 0. \quad (19)$$

These expressions can now be used to fill the appropriate intermediates of a generalized CPHF code to solve for the ov block of the density correction using the  $z$ -vector method of Handy and Schaefer.<sup>25</sup>

TABLE I. List of combinations of high and low levels of theory used throughout this paper.

Entry	High level:low level
1	MP2/6-31G(d):HF/6-31G(d)
2	B3LYP/6-31G(d):HF/6-31G(d)
3	B3LYP/6-31G(d,p):BLYP/6-31G
4	MP2/6-31G(d):BLYP/6-31G
5	B3LYP/6-311++G(d,p):HF/3-21G
6	MP2/6-31G(d):HF/3-21G

### III. ASSESSMENT OF THE QM:QM-EE(LÖWD) METHOD

We have carried out a preliminary set of calculations on two challenging chemical systems to gauge the performance of the ONIOM based, QM:QM-EE(Löwd) method. In the first test case, we explore the ability of our new EE method to reproduce the high level results for the deprotonation energies of a hydroxylated spherosiloxane cluster ( $\text{Si}_8\text{O}_{12}\text{H}_7\text{OH}$ ). In the second test case, we evaluate the QM:QM-EE(Löwd) method by calculating the dissociation energy of a water molecule from a  $[\text{ZnIm}_3(\text{H}_2\text{O})]^{2+}$  complex. Both of these chemical systems were previously used to test the performance of the QM:QM-ME, QM:QM-EE(Mull), and QM:QM-EE(GAME) methods.<sup>22</sup> To measure the variations in performance when pairing different levels of theory, we have evaluated six high-level:low-level combinations for both test cases. The combinations chosen for both studies are listed in Table I. While other metrics may be investigated in the future, here, we determine the performance of an ONIOM method by its ability to reproduce the results (e.g., bond energies, geometries) obtained from the real system-high level of theory (RH) calculations. Therefore, the deviations reported for the ONIOM methods are calculated as  $\Delta E^{\text{ONIOM}} - \Delta E_{\text{RH}}$ , where  $\Delta E_{\text{RH}}$  is the reaction energy obtained from the real system-high level of theory (RH) calculations. Gradients and energies have been implemented in the development version of the Gaussian program suite.<sup>33</sup> All assessment calculations have been performed with this software.

#### A. Deprotonation energy: Siloxane cube

In the first example, we provide results for the deprotonation energy of a hydroxylated spherosiloxane cube which has been successfully used as a cluster model for the isolated hydroxyl of silica.<sup>34</sup> When treating the siloxane cube with an ONIOM method, there are a variety of ways one can partition the system into regions I and II. We consider the two different model systems shown in Fig. 1. In both of these figures, the region I subsystem (model system) is depicted by colored atoms, while the region II subsystem (real system-model system) is depicted with white atoms. In both of these systems, covalent bonds severed by the region I-II boundary, are capped with hydrogen atoms to retain a closed shell model system. The H link atom which replaces an oxygen atom is placed along the cut bond and at a distance of 0.9 times the cut Si-O bond distance. Except for the atoms di-

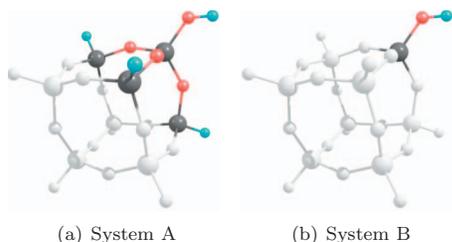
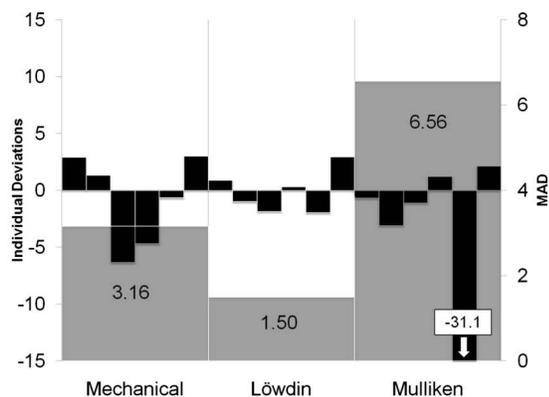


FIG. 1. The two model system-real system divisions used in this study. The real system contains both the colored and white atoms, while the model system is comprised of only the colored atoms.

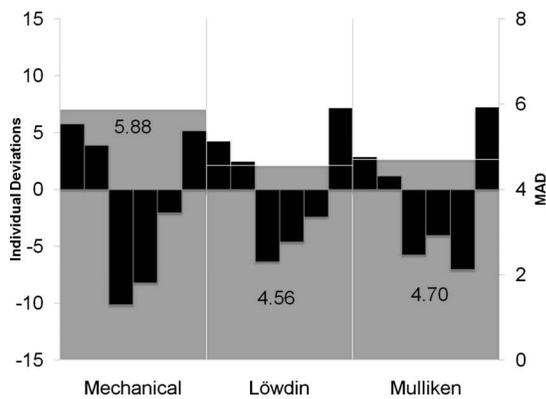
rectly replaced by H link atoms (O atoms), all region II atomic charges are included in the embedding potentials.

The results for the deprotonation energies are given in Fig. 2. Deviations are calculated as deprotonation energy of the QM:QM model minus the deprotonation energy of the real system-high level calculation. Individual deviations for each of the 6 high-low combinations (thin columns) are shown in the same order, left to right, as given in Table I and are plotted against the left axis. The total mean absolute deviation for each ONIOM model, QM:QM-ME, QM:QM-EE(Löwd), and QM:QM-EE(Mull), are shown as wide superimposed columns plotted against the right vertical axis.

Taking both systems into account (A and B) the QM:QM-EE(Löwd) method performs the best with an over-



(a) SiO<sub>2</sub> - System A



(b) SiO<sub>2</sub> - System B

FIG. 2. Siloxane cube: Comparison of the performance for the mechanical, Löwdin, and Mulliken charge embedding schemes. Individual model deviations (thin, black columns) and MAD values (wide, black columns) are in units of kcal/mol.

all MAD of 3.03 kcal/mol, followed by QM:QM-ME with a mean absolute deviation (MAD) of 4.52 kcal/mol. Surprisingly, the QM:QM-EE(Mull) method had the largest overall MAD of 5.63 kcal/mol. The large MAD for the Mulliken embedding comes primarily from a single high:low combination which fails drastically. High:low combination five (B3LYP/6-311++G(*d,p*):HF/3-21G) has the largest basis set dissimilarity of all the high:low combinations considered. If high:low combination five is removed from 2(a), the MADs of the Mechanical, Löwdin, and Mulliken embedding methods are 3.7, 1.4, and 1.7 kcal/mol, respectively.

For each high:low combination and both system A and system B, the QM:QM-EE(Löwd) method has a deviation which is in between that of the QM:QM-ME and QM:QM-EE(Mull) methods (on average being closer to the more accurate of the two). This is a direct result of the previously mentioned observation that Löwdin atomic charges tend to be smaller in magnitude than the corresponding Mulliken atomic charges, *vide supra*.

For system A [Fig. 2(a)] the QM:QM-EE(Löwd) method clearly outperforms the other ONIOM methods. Comparing the deviations of the QM:QM-EE(Löwd) and QM:QM-ME methods, both have similar behaviors, though the polarization from the Löwdin charges appears to correct for some of the deficiencies in the ME models for high:low combinations one, three, and four. For high:low combination five, the QM:QM-EE(Mull) deprotonation energy is 31.1 kcal/mol smaller than the RH calculation. This is likely resulting from a preferential overpolarization<sup>37</sup> of the proton-detached (anionic) species. The closest point charge to the model system is obtained from the positive Si atomic charge. If Mulliken atomic charges are used, this embedding point charge oddly becomes more positive when going from the proton-attached state to the proton-detached state (1.982–2.016). If Löwdin atomic charges are used however, the opposite is observed (1.319–1.305). For this high:low pairing, the basis set in the high level of theory (6-311++G(*d,p*)) is apparently flexible enough to allow the model system electron density to interact too strongly with the embedding point charges which are considerably larger for Mulliken than Löwdin.<sup>38</sup>

For system B, there is less distinction between the two EE methods, With QM:QM-EE(Löwd) providing overall better results. It is clear that electronic embedding, using either the QM:QM-EE(Löwd) or QM:QM-EE(Mull) methods, improves the performance of the ONIOM approach.

## B. Water dissociation: [ZnIm<sub>3</sub>(H<sub>2</sub>O)]<sup>2+</sup>

In the second example, we consider the dissociation energy of water from a [ZnIm<sub>3</sub>(H<sub>2</sub>O)]<sup>2+</sup> complex shown in Fig. 3(a). In this example, region I consists of a Zn(II) ion and a bonded water molecule, while region II contains the three imidazole ligands. Because the Zn ion is better described as being “solvated” by the imidazole ligands rather than being covalently bound to them, link atoms were not needed to truncate any bonds. In previous work,<sup>22</sup> we have shown that QM:QM-ME can fail catastrophically when calculating the potential energy surface for the Zn(II)–OH<sub>2</sub> bond length. This was found to be a result of the Zn and H<sub>2</sub>O

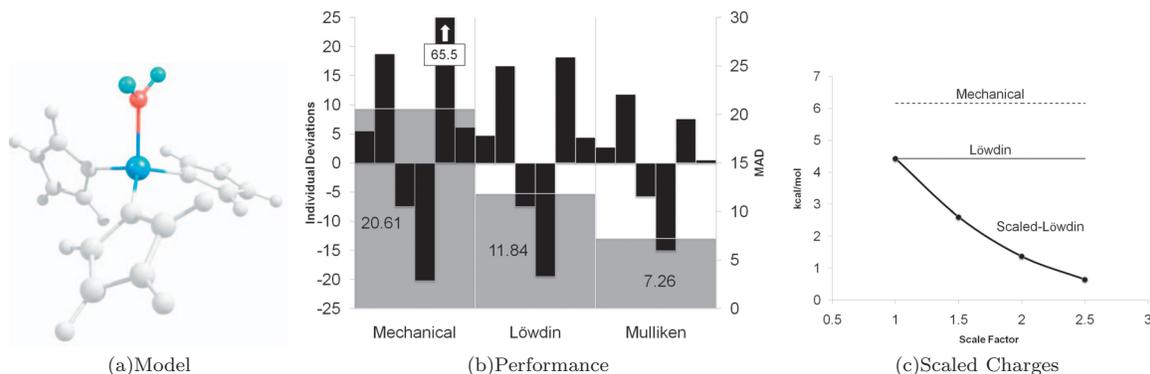


FIG. 3. [ZnIm<sub>3</sub>(H<sub>2</sub>O)]<sup>2+</sup>: (a) Illustration of the model used in the [ZnIm<sub>3</sub>(H<sub>2</sub>O)]<sup>2+</sup> test case. Region I atoms are shown in color. Region II atoms shown in white. (b) Comparison of the performance for the mechanical, Löwdin, and Mulliken charge embedding schemes. Individual model deviations (thin, black columns) and MAD values (wide, black columns) are in units of kcal/mol. (c) Performance of the scaled charge scheme for the sixth high:low combination. Deviations and MAD values are in units of kcal/mol.

dissociating into partially charged species, a known failure of DFT. Therefore, we have decided to calculate dissociation energies so that the H<sub>2</sub>O and Zn(II) are calculated separately, ensuring proper dissociation into Zn<sup>2+</sup> and H<sub>2</sub>O.

In Fig. 3(b), we show the results for the QM:QM-ME, QM:QM-EE(Löwd), and QM:QM-EE(Mull) methods, respectively. Again, the thin black columns represent the deviations for each of the six high:low pairings, given in the same order as in Table I, and the superimposed, thick gray columns represent the MAD for all six high:low pairings. With MADs of 20.6, 11.8, and 7.3 kcal/mol for QM:QM-ME, QM:QM-EE(Löwd), and QM:QM-EE(Mull), respectively, it is clear that the QM:QM-EE(Mull) method outperforms the other two methods. Noting the poor performance of the ME, electronic polarization effects appear to be crucial in describing the Zn–OH<sub>2</sub> bond strength with a QM:QM method. Overpolarization does not appear a problem with this system, as it was with the previous siloxane cube example. Therefore, the Löwdin charges, which are smaller in magnitude than the Mulliken charges, do not polarize the high level region sufficiently enough.

To address systems which require more extensive electronic embedding, such as the [ZnIm<sub>3</sub>(H<sub>2</sub>O)]<sup>2+</sup>, we also present a scaled charge method. The most straightforward approach would be to simply scale all real system atomic charges by a constant  $k$ . However, for charged species, scaling the charges by  $k$  will change the total charge. We therefore define a set of atomic charges,  $q'_A$ , which are obtained by first centering the unscaled charges (either Mulliken or Löwdin) at zero, then scaling by  $k$ , and finally unshifting to preserve the overall charge. We use the following definition for the scaled charges:

$$q'_A = k(q_A - s) + s, \quad (20)$$

where we have defined  $s$  to be the total, real system molecular charge divided by the number of real system atoms. By defining  $s$  as such, we effectively shift all the charges by removing the excess charge (positive or negative) evenly across the molecule. We note that due to the inclusion of the charge-retaining shift,  $s$ , scaling by  $k$  does not simply in-

crease the magnitude of the overall embedding potential, but is interpreted as increasing the polarity of the embedding potential. This is highlighted by considering the limiting case of  $k=0$ , in which all the region II atoms have exactly the same atomic charge,  $q_A^{k=0}=s$ , hence a completely unpolarized embedding potential. As this is obviously the case for neutral molecules ( $q_A^{k=0}=s=0$ ), our approach has the same conceptual interpretation for both charged and neutral systems.

Because the scale and shift factors are constant with respect to all variables, the scaled gradients are simply

$$q_A'^x = kq_A^x. \quad (21)$$

In Fig. 3(c), we take high:low combination six as an example case and plot the deviation of the QM:QM-EE(Löwd,  $k$ ) method versus scale factor  $k$ . Plotted with the mechanical and unscaled Löwdin embedding methods as a reference point, we demonstrate that the unscaled Löwdin atomic charge embedding does not polarize region I to a large enough extent, and that by scaling the charges (while maintaining the system's overall 2+ charge) one can greatly improve the results by increasing the magnitude of the region I embedding potential.

The scaled-Löwdin charge method, QM:QM-EE(Löwd,  $k$ ), provides a means to fine tune the amount of electronic polarization that occurs within the EE model. As the optimal value for  $k$  would clearly depend on both the model chemistries employed and the molecular systems being studied, we suggest calibrating the  $k$  parameter prior to use in applications. In future work, we aim to apply the QM:QM-EE(Löwd,  $k$ ) method on a variety of systems to establish a set of suggested values of  $k$ .

In this report, we have extended the applicability of the recently developed QM:QM-EE method<sup>21</sup> to Löwdin atomic charges. In addition to the energies, we have presented an efficient implementation of the QM:QM-EE(Löwd) [and QM:QM-EE(Löwd,  $k$ )] energy gradients, which involve direct coupling between the real and model system subcalcu-

lations. By evaluating the performance on two challenging molecular systems, we come to the following conclusions:

- (1) Because the Löwdin population analysis tends to produce atomic charges which are smaller in magnitude than the corresponding Mulliken charges, we have observed that, for every case considered, the QM:QM-EE(Löwd) method provides reaction energies which are in between those from the QM:QM-ME and QM:QM-EE(Mull) methods.
- (2) For the siloxane cube example, the QM:QM-EE(Löwd) method outperforms the other two methods, when compared to the high level of theory-real system calculation. The improvement over the Mulliken charge embedding model is primarily due to a few particularly large deviations for certain high:low combinations. It appears as though the QM:QM-EE(Löwd) method achieves a good balance between underestimating the region I polarization (which is seen with QM:QM-ME) and overestimating the polarization [which was found for a couple of cases using QM:QM-EE(Mull)].
- (3) For the  $[\text{ZnIm}_3(\text{H}_2\text{O})]^{2+}$  example, electronic polarization effects appear to be more significant than in the siloxane cube example, and the benefits of Löwdin embedding toward reducing overpolarization now result in underpolarization of the high level region.
- (4) By using the scaled-Löwdin charge method, QM:QM-EE(Löwd,  $k$ ), we have shown that by applying a scale factor which maintains the overall molecular charge for neutral and charged systems alike, one can fine tune the extent of electronic polarization and obtain a value for  $k$  which better compares to the high level of theory-real system calculation.

By providing reaction energies which are intermediate to the ME and Mulliken embedding methods, the Löwdin embedding approach is likely to be useful in helping achieve a balance between no explicit polarization and the overpolarization of the region I density. For systems which require more extensive polarization, the QM:QM-EE(Löwd, $k$ ) method can be used to manually control the embedding potential without changing the overall molecular charge of anions and cations.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant No. CHE-0911454) and Gaussian, Inc. The authors would like to acknowledge Professor Donald Truhlar for helpful correspondence and Dr. Priya Parandekar for insightful discussions.

- <sup>1</sup>L. A. Curtiss, P. C. Redfern, and K. Raghavachari, *J. Chem. Phys.* **126**, 084108 (2007).
- <sup>2</sup>N. J. DeYonker, T. R. Cundari, and A. K. Wilson, *J. Chem. Phys.* **124**, 114104 (2006).
- <sup>3</sup>A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, *J. Chem. Phys.* **120**, 4129 (2004).
- <sup>4</sup>M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.* **11**, 700 (1990).
- <sup>5</sup>F. Maseras and K. Morokuma, *J. Comput. Chem.* **16**, 1170 (1995).
- <sup>6</sup>U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **7**, 718 (1986).
- <sup>7</sup>S. Humbel, S. Sieber, and K. Morokuma, *J. Chem. Phys.* **105**, 1959 (1996).
- <sup>8</sup>M. Svensson, S. Humbel, R. D. J. Froese, T. Maturaba, S. Sieber, and K. Morokuma, *J. Phys. Chem.* **100**, 19357 (1996).
- <sup>9</sup>P. B. Karadakov and K. Morokuma, *Chem. Phys. Lett.* **317**, 589 (2000).
- <sup>10</sup>T. Vreven and K. Morokuma, *J. Comput. Chem.* **21**, 1419 (2000).
- <sup>11</sup>T. Vreven, B. Mennucci, C. da Silva, K. Morokuma, and J. Tomasi, *J. Chem. Phys.* **115**, 62 (2001).
- <sup>12</sup>T. Vreven and K. Morokuma, *Theor. Chem. Acc.* **109**, 125 (2003).
- <sup>13</sup>N. Rega, S. Iyengar, and G. A. Voth, *J. Phys. Chem. B* **108**, 4210 (2004).
- <sup>14</sup>D. Bakowies and W. Thiel, *J. Phys. Chem.* **100**, 10580 (1996).
- <sup>15</sup>J. Gao, *J. Comput. Chem.* **18**, 1061 (1997).
- <sup>16</sup>Y. Lin and J. Gao, *J. Chem. Theory Comput.* **3**, 1484 (2007).
- <sup>17</sup>M. J. Field, *Mol. Phys.* **91**, 835 (1997).
- <sup>18</sup>J. Gao, *J. Phys. Chem. B* **101**, 657 (1997).
- <sup>19</sup>J. Gao, *J. Chem. Phys.* **109**, 2346 (1998).
- <sup>20</sup>W. Xie and J. Gao, *J. Chem. Theory Comput.* **3**, 1890 (2007).
- <sup>21</sup>H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, *J. Chem. Phys.* **128**, 034107 (2008).
- <sup>22</sup>P. V. Parandekar, H. P. Hratchian, and K. Raghavachari, *J. Chem. Phys.* **129**, 145101 (2008).
- <sup>23</sup>R. S. Mulliken, *J. Chem. Phys.* **36**, 3428 (1962).
- <sup>24</sup>J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, *Int. J. Quantum Chem., Quantum Chem. Symp.* **13**, 225 (1979).
- <sup>25</sup>N. C. Handy and H. F. Schaefer III, *J. Chem. Phys.* **81**, 5031 (1984).
- <sup>26</sup>F. Labat, I. Ciofini, H. P. Hratchian, M. Frisch, K. Raghavachari, and C. Adamo, *J. Am. Chem. Soc.* **131**, 14290 (2009).
- <sup>27</sup>F. Jensen, *Introduction to Computational Chemistry*, 2nd ed. (Wiley, Hoboken, 2007).
- <sup>28</sup>P. O. Löwdin, *Adv. Quantum Chem.* **5**, 185 (1970).
- <sup>29</sup>Q. Wu and T. Van Voorhis, *J. Phys. Chem. A* **110**, 9212 (2006).
- <sup>30</sup>H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria, and M. J. Frisch, *J. Chem. Phys.* **114**, 9758 (2001).
- <sup>31</sup>P. Jørgensen and J. Simons, *J. Chem. Phys.* **79**, 334 (1983).
- <sup>32</sup>J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, *J. Phys. Chem.* **96**, 135 (1992).
- <sup>33</sup>M. J. Frisch *et al.*, GAUSSIAN Development Version, Revision G.01, 2007.
- <sup>34</sup>B. Civalleri, E. Garrone, and P. Ugliengo, *Chem. Phys. Lett.* **294**, 103 (1998).
- <sup>35</sup>H. Lin and D. G. Truhlar, *Theor. Chem. Acc.* **117**, 185 (2007).
- <sup>36</sup>H. Lin and D. G. Truhlar, *J. Phys. Chem. A* **109**, 3991 (2005).
- <sup>37</sup>As evidence of overpolarization, if we employ Eq. (20) and scale the real-system point charges ( $k=0.5$ ), the large deviation of 31.1 is reduced to 1.0 kcal/mol. We note that other methods may be of use here as well, particularly the redistributed-charge schemes of Lin and Truhlar (Refs. 35 and 36).
- <sup>38</sup>We note that the poor performance of QM:QM-EE(Mull) for high:low combination five is not only a problem with the basis set disparity. If B3LYP is replaced with Hartree-Fock (HF/6-311++G( $d,p$ ):HF/3-21G) all embedding methods provide accurate deprotonation energies.