JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

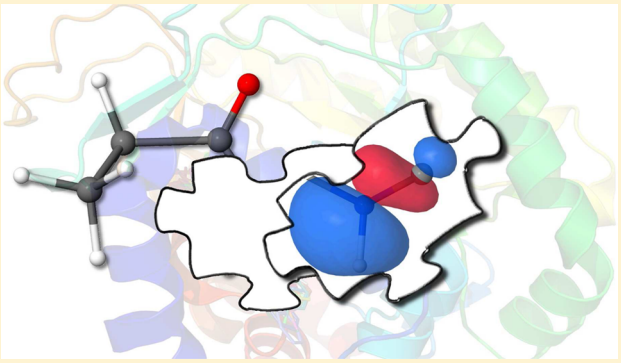
Exact and Optimal Quantum Mechanics/Molecular Mechanics Boundaries

Qiming Sun and Garnet Kin-Lic Chan*

Department of Chemistry, Princeton University, Princeton New Jersey 08544, United States

Supporting Information

ABSTRACT: Motivated by recent work in density matrix embedding theory, we define exact link orbitals that capture all quantum mechanical (QM) effects across arbitrary quantum mechanics/molecular mechanics (QM/MM) boundaries. Exact link orbitals are rigorously defined from the full QM solution, and their number is equal to the number of orbitals in the primary QM region. Truncating the exact set yields a smaller set of link orbitals optimal with respect to reproducing the primary region density matrix. We use the optimal link orbitals to obtain insight into the limits of QM/MM boundary treatments. We further analyze the popular general hybrid orbital (GHO) QM/MM boundary across a test suite of molecules. We find that GHOs are often good proxies for the most important optimal link orbital, although there is little detailed correlation between the detailed GHO composition and optimal link orbital valence weights. The optimal theory shows that anions and cations cannot be described by a single link orbital. However, expanding to include the second most important optimal link orbital in the boundary recovers an accurate description. The second optimal link orbital takes the chemically intuitive form of a donor or acceptor orbital for charge redistribution, suggesting that optimal link orbitals can be used as interpretative tools for electron transfer. We further find that two optimal link orbitals are also sufficient for boundaries that cut across double bonds. Finally, we suggest how to construct "approximately" optimal link orbitals for practical QM/MM calculations.



INTRODUCTION

In quantum mechanics/molecular mechanics (QM/MM) simulations, the optimal treatment of the boundary between the quantum mechanical (primary) and molecular mechanical (secondary) regions remains an open question. This is because, besides the more obvious electrostatic effects, artificially terminating the QM region leads to unsaturated valences at the boundary that can alter the primary QM region's electronic structure. Further, abrupt termination of the QM region creates artificial confinement of the wave function. This boundary problem is a perennial issue of interest that has been discussed and considered by many authors.[1−6]

Following the review in ref 3, we group the two common approaches in the literature as (i) link atom approaches, which either include a simple extra hydrogen atom, or parametrized pseudoatom, to cap the QM boundary,[7−13] and (ii) link orbital approaches, such as the local self-consistent field (SCF) approach (which constructs a "strictly localized bond orbital" from a small model molecule[14−18]) and the general hybrid orbital approach (GHO) (that is used to terminate sp^3 carbon boundaries[19−22]). All these approaches have merits and demerits, but the GHO approach will be of particular interest here, as we will use it as a benchmark.

In this work, we define *exact and optimal* link orbitals for the QM/MM boundaries. The exact link orbitals provide an exact treatment of the boundary in the sense that the error of observables in the primary QM region is precisely zero. If a subset of these link orbitals is chosen, then the boundary is treated approximately, but is optimal in a mathematically well-defined least-squares sense. The construction derives from our earlier density matrix embedding theory (DMET),[23−25] originally introduced for QM/QM embedding in correlated systems; however, for mean-field theories as used in most QM/MM calculations, the full self-consistent DMET machinery is not needed and the formulation is simplified. The exact and optimal link orbitals are obtained, in principle, from a QM treatment of the full molecule (primary + secondary regions) and are thus, without approximation, not practical for a priori calculation. However, once constructed, they provide a rigorous way to assess the optimality of empirically proposed boundary treatments. Here, we use this optimality to analyze and gain insight into the quality of the GHO construction across a large molecular data set. Further, we use the optimal construction to study new classes of boundaries not treated by GHO, such as carbon−carbon double bonds. At the end of this work, we discuss how to construct approximations to the exact and

Received: June 12, 2014
Published: July 25, 2014

   dx.doi.org/10.1021/ct500512f | J. Chem. Theory Comput. 2014, 10, 3784−3790

optimal link orbitals, which could be used a priori in practical QM/MM calculations and other forms of embedding.

DEFINITION OF EXACT AND OPTIMAL LINK ORBITALS

We first define the exact link orbitals. *In a nutshell, exact link orbitals are the natural orbitals of the secondary region density matrix*. In more detail, we write the Hamiltonian of the full molecule in terms of components in the primary, primary−secondary, and secondary regions

$$H = H_p + H_{ps} + H_s \tag{1}$$

where p and s label the primary and secondary regions. We further restrict ourselves here to Hartree−Fock (HF) solutions of H; the Hartree−Fock ground-state of H will be denoted Φ. To cap the QM/MM boundary, we introduce link orbitals $\{|b\rangle\}$ ("b" for boundary) to describe the quantum mechanical effects of delocalization into the secondary region. The QM/MM Hamiltonian then becomes

$$H_{QM/MM} = H_p + H_{pb} + H_b + V_p \tag{2}$$

where V_p is an additional external potential, electrostatic in origin, arising from secondary region charge not contained in the link orbitals, as well as the nuclear charges not in the primary region. (V_p can be precisely obtained from a core density matrix γ_c, defined later). The HF ground-state of $H_{QM/MM}$ is denoted $\Phi_{QM/MM}$.

The exact link orbitals are eigenvectors of the secondary density matrix γ_s, defined as

$$[\gamma_s]_{ij} = \langle\Phi|a_i^\dagger a_j|\Phi\rangle, \quad i, j \in s$$
$$\gamma_s|\phi\rangle = \lambda|\phi\rangle \tag{3}$$

The eigenvectors of γ_s satisfy some important properties. The first is that the eigenvalues of γ_s fall into three categories (i) $\lambda = 0$, (ii) $\lambda = 2$, and (iii) $0 < \lambda < 2$. Category iii of eigenvalues and eigenvectors defines the *link* orbitals, while category ii defines the secondary core orbitals, which yield the electronic contribution to the external potential V_p as the Coulomb and exchange potentials from the core density matrix, $\gamma_c = \sum_i |\phi_i\rangle\langle\phi_i|$, $\lambda_i = 2$

$$[V_p]_{rs} = \sum_i 2\langle ri|si\rangle - \langle ri|is\rangle, \quad \lambda_i = 2 \tag{4}$$

The second important property is that the number of link orbitals is *exactly the same* as the number of orbitals in the primary region (denoted n_p), *regardless of the size of the secondary region*. Using the set of n_p link orbitals, $H_{QM/MM}$ is then *exact*, in the sense that

$$\langle\Phi_{QM/MM}|a_i^\dagger a_j|\Phi_{QM/MM}\rangle = \langle\Phi|a_i^\dagger a_j|\Phi\rangle \tag{5}$$

both for the primary density matrix ($i, j \in p$) *and* for the primary off-diagonal density matrix ($i \in p, j \in s$ or $i \in s, j \in p$). This is proved explicitly in refs 23−25; see also the references therein. Consequently, the expectation value of any one-particle operator, with any component on the primary region (including off-diagonal operators between the primary and secondary regions) is reproduced exactly in the ground-state of the QM/MM Hamiltonian.

The above construction produces exact link orbitals, with a boundary link region as large as the primary region. While this reduces complexity, since the secondary region is usually much larger than the primary region, we may nonetheless still desire a smaller set of link orbitals (for example, a single link orbital, if we only cut one bond). To do so, we truncate the set of link orbitals, choosing first those with eigenvalues λ furthest from 0 or 2 (or equivalently, the largest $|1 - \lambda|$). This truncated set of orbitals is *optimal* as follows. Define a projector onto the primary plus link region, $P = \sum_{i\in p,b} |\phi_i\rangle\langle\phi_i|$. Then, the first n_b link orbitals with the largest $|1 - \lambda|$ minimize the difference between the exact and projected density matrices

$$\sum_{i\in p,s,j\in p,s} |\langle\Phi P|a_i^\dagger a_j|P\Phi\rangle + [\gamma_c]_{ij} - \langle\Phi|a_i^\dagger a_j|\Phi\rangle|^2 \tag{6}$$

When $n_b = n_p$, the above error is 0.

Note that for $n_b < n_p$, the ground-state (energy-optimized) density matrix of $H_{QM/MM}$ is no longer precisely the same as the projected density matrix, i.e.

$$\langle\Phi_{QM/MM}|a_i^\dagger a_j|\Phi_{QM/MM}\rangle \neq \langle\Phi P|a_i^\dagger a_j|P\Phi\rangle \tag{7}$$

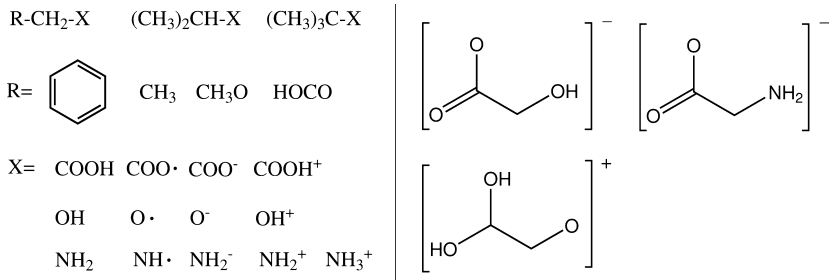
Thus, the truncated set of optimal link orbitals do not lead to an energy-optimized ground-state density matrix that is strictly optimal in the sense of eq 6. This discrepancy is analogous to the difference between natural orbitals in configuration interaction (which are obtained from a density matrix criterion) and the orbitals obtained from energy optimization in multiconfigurational self-consistent field. In practice, however, we expect the optimal link orbitals to be near optimal with respect to the energy-optimized density matrix as well. We examine this point further below.

To summarize, we define the exact link orbitals as the set of eigenvectors of the secondary density matrix whose natural eigenvalues satisfy $0 < \lambda < 2$. There are (at most) n_p such eigenvectors, where n_p is the number of orbitals in the primary region. We can truncate the set of exact link orbitals to obtain a smaller set of optimal link orbitals (e.g., one per cut bond) that are rigorously optimal in the sense of density matrix projection, and are expected to be near optimal in terms of energy minimization. We now use these optimal link orbitals to numerically assess the widely used GHO construction, and to determine the limits of a truncated link orbital boundary description.

COMPARISON WITH GENERAL HYBRID ORBITALS

The GHO method provides a single link orbital for every division between the primary and secondary regions at an sp^3 carbon. (Other kinds of boundaries are not usually considered). In the standard GHO approach, a set of 4 sp^3 hybrids are constructed using mixing coefficients that follow the detailed tetrahedral geometry at the carbon boundary center.[20] The hybrid orbital that points toward the primary region is used as the "link" orbital to define H_{pb}, H_b in eq 2, while the remaining 3 auxiliary orbitals are kept frozen, and contribute to V_p.

We first discuss some differences between GHOs and optimal link orbitals from a theoretical perspective. There are two sources. First, GHOs attempt to reproduce the most important secondary orbital(s) that couple to the primary region. These should be the link orbital(s) with the largest eigenvalues λ, however, since the GHO is empirically constructed, they will differ in practice. In fact, GHOs may not even lie in the exact link orbital space, and may overlap with the core or empty secondary orbitals defined in eq 3. (This has important consequences for their electronegativity as discussed later). Second, GHOs often define the external potential V_p



(a) Fragments for molecules in the test set. (b) Special structures.

Figure 1. QM boundary test set. The complete test set is formed by attaching the various R and X groups to the protomolecules shown at the top. The special structures denote molecules with geometries strongly distorted from their neutral counterparts.

from the Coulomb and exchange contributions of the GHO auxiliary orbitals, augmented with the C 1s core. This, too may differ from the exact V_p defined from the exact core secondary orbitals.

To further assess GHO methods in the context of the optimal link orbitals, we turn to numerical calculations. We construct a molecular test suite comprising different functional groups attached to a central sp^3 carbon atom, to simulate different molecular environments. The functional groups are $-OCH_3$, $-CH_3$, $-C_6H_5$, $-OH$, $-NH_2$, $-COOH$ (Figure 1a). In addition, we consider the ionized and electron attached ions, and protonated forms, obtaining in total 78 species spanning closed shell, charged closed shell, open shell, and charged open shell molecules. Denoting the four groups around the central carbon as R, A, B, and X, we obtain four classes of QM/MM boundaries from cutting the bond between the carbon and the functional groups. In total (excluding duplicates) this yields 221 QM/MM boundaries for our test suite.

Molecular geometries were optimized at the HF/cc-pVDZ level using the ORCA program package.[26] The cation $CH_3OCH_2NH_2^+$ dissociated during geometry optimization, thus we studied this system at its neutral geometry. The structures of $COOHCH_2NH^-$, $COOHCH_2O^-$, and $COOHCH_2OH^+$ were significantly distorted from their neutral counterparts, as noted in Figure 1b. The single-point calculations both for the full QM system (to define the optimal link orbitals) as well as for the QM/MM calculations (primary region plus GHO or optimal link orbitals) all employed the STO-6G basis set. (Note that our use of a minimal basis is dictated by the fact that GHO calculations typically use a minimal basis; however, the optimal link orbitals can be defined in any basis. Our calculations in larger bases show that our conclusions about the relative merits of optimal link orbitals and GHOs are not significantly affected by basis. Calculations of adiabatic ionization energies (analogous to Table 1) in a cc-pVDZ basis are given in the Supporting Information). Orthogonal basis functions are required for both the GHO and optimal link orbitals. We carried out a two-step orthogonalization. First, we computed atomic orbitals using occupancy-averaged atomic HF calculations. Next, we projected out the atomic core orbitals, leaving behind the (perpendicular) valence subspace. Löwdin orthogonalization was then carried out within the core and valence subspaces separately.

We first compare the external potential V_p typically used with GHOs, constructed from the auxiliary GHO orbitals and the carbon 1s orbitals, with the exact V_p, obtained from the secondary core orbitals. We use the GHO to cap the primary region in both cases, and carry out a Hartree−Fock energy minimization for the corresponding $H_{QM/MM}$. (Note that the original GHO work replaced the carbon 1s core by an effective core potential (ECP).[19]) In the top part of Figure 2 we compare the exact and GHO V_p's. These differ by about 0.01 au or 1%−5%, as measured by the r.m.s. V_p, when the QM primary region is not a hydrogen atom; the deviation increases to about 0.05 au when the QM primary region is a hydrogen atom, since the entire primary region is then on the boundary and feels the effect of the secondary region most strongly. In the lower part of Figure 2 we compare the deviation of the primary region density matrix (from exact) in a GHO energy optimized calculation, using the standard GHO V_p, and using the exact core V_p. We see that the GHO density matrix error is significantly reduced when using the exact V_p as opposed to the

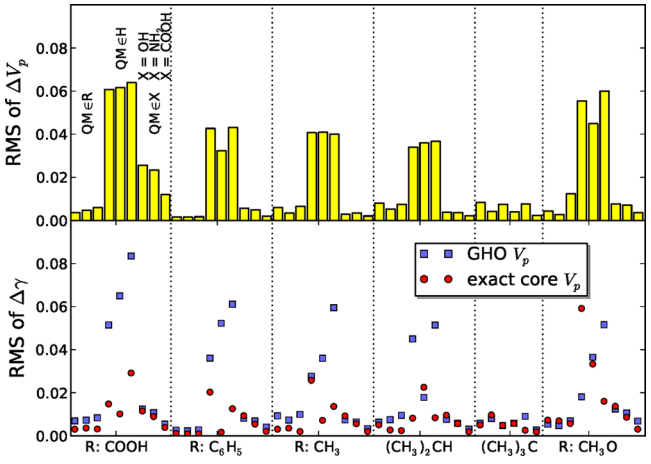


Figure 2. Difference between the (self-consistent) density matrices in a GHO calculation (lower panel), arising from external potentials V_p computed from (i) the auxiliary GHO orbitals and (ii) the exact core secondary orbitals. The top panel shows the deviation of the GHO V_p from the exact core V_p. Overall, this shows that the exact core V_p is much better than the V_p obtained from the GHO auxiliary orbitals, and this is used in the subsequent GHO calculations in this work.

Table 1. Adiabatic Ionization Energy (eV)

	GHO	1 orb	2 orb	ref
$(CH_3)_2CH[COOH]$	5.20	5.84	5.84	5.95
$(CH_3)_2CH[NH_2]$	4.48	5.16	5.52	5.63
$(CH_3)_2CH[OH]$	6.00	7.16	7.78	8.05
$CH_3OC_2[COOH]$	7.60	7.21	6.92	7.06
$CH_3OC_2[NH_2]$	6.02	6.59	7.16	7.18
$CH_3OC_2[OH]$	8.48	7.73	6.37	6.15

dx.doi.org/10.1021/ct500512f | J. Chem. Theory Comput. 2014, 10, 3784−3790

standard GHO V_p. Consequently, in our subsequent calculations, to present GHOs in a better light *we use the exact V_p with both the GHO and optimal link orbital constructions.*

We next analyze the quality of the GHO as a link orbital, versus the optimal link orbital, to capture the boundary quantum effects. Since in all cases here we only have one sp^3 boundary and thus only one GHO, we compare against only a single optimal link orbital. We first use the projected density matrix error criterion. Following eq 6 this is $\langle \Phi P | a_i^\dagger a_j | P\Phi \rangle$ where $P = \sum_{k\in p,b} |\phi_k\rangle\langle\phi_k|$ and b denotes the GHO orbital, or the optimal link orbital, respectively. Note that the projected primary region density matrix ($i, j \in$ p) will be exact in both cases (since Φ is exact and the projector does not affect the primary region). Thus, in Figure 3 we compare the error in the primary off-diagonal density matrix ($i \in$ p, $j \in$ s). As seen from the figure and as proved above, the optimal link orbital always gives the least error in the projected density matrix, and in the case of the single-hydrogen QM primary region, a single optimal boundary orbital produces an exact embedding (because the primary region has only one orbital, $n_p = 1$, and thus there is only one link orbital). The performance of the GHO is roughly two times worse than the optimal boundary orbital across the different species.

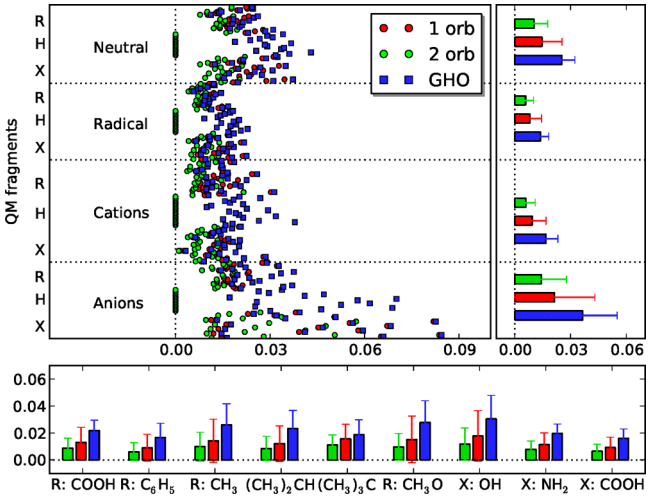


Figure 3. RMS error of the projected density matrix in the primary off-diagonal region γ_{ij}, $i \in$ p, $j \in$ s. The reference is the full system SCF density matrix. All QM region calculations are based on the exact core external potential V_p. The 1 and 2 optimal orbital boundaries are labeled as "1 orb" and "2 orb". The three letters R, H, and X stands for three kinds of QM fragments. In each kind of QM fragment, from top to bottom, there are six groups of points which correspond to R:COOH, R:C_6H_5, R:CH_3, $(CH_3)_2CH$, $(CH_3)_3C$, and R:CH_3O. In each group, there are three or four points associated with X = OH, X = NH_2, and X = COOH or their derived species as listed in Figure 1a. In total, there are 221 nonduplicated points. We see that, for this metric, the GHO error is always larger than the 1 optimal link orbital as can be mathematically proven.

We now examine the errors incurred using GHOs, and the optimal link orbital, for the *energy-optimized density matrix*, $\langle \Phi_{QM/MM} | a_i^\dagger a_j | \Phi_{QM/MM} \rangle$ (the ground-state density matrix of $H_{QM/MM}$). As discussed, the optimal link orbital is only optimal with respect to the projected density matrix error, not the energy-optimized density mtarix. These errors are shown for the various species in Figure 4. We find that although the optimal link orbital is not strictly optimal for this quantity, it

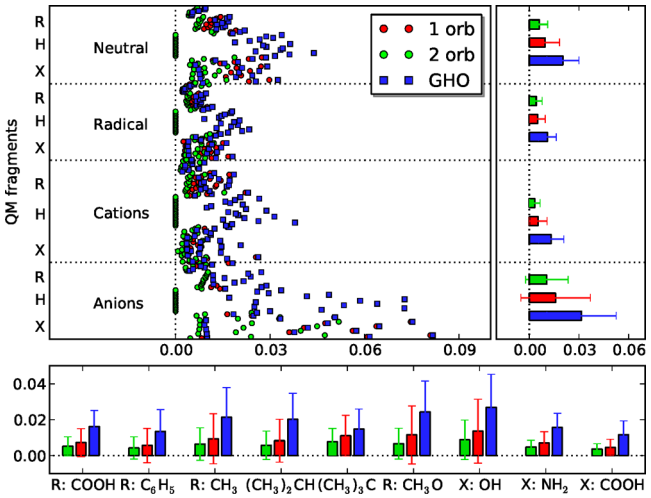


Figure 4. RMS error of the energy-optimized density matrix (in terms of $\Phi_{QM/MM}$) in the primary region and its off-diagonal region γ_{ij}, $i \in$ p, $j \in$ p,s. The labeling is as in Figure 3. Although the optimal link orbital is not strictly optimal with respect to this metric, we still see that the GHO is always worse than a single optimal link orbital.

nonetheless gives a very faithful representation of the energy optimized density matrices, as we expect. The GHO errors are roughly two times larger than those using the optimal link orbital, similar to what we see in the projected density matrix errors.

We next compare the GHO and optimal link orbitals with respect to errors in the total primary region charge. The primary region charge error ΔQ_p ($Q_p = -\sum_{i\in p}\gamma_{ii}$) criterion is quite different from the density matrix error criterion, as it allows for cancellation of positive and negative errors in the density matrix. It thus represents an error in an observable completely unrelated to the error criterion for which the link orbitals are optimal. Q_p has the further advantage that we can reason about its errors in a simple way. As we truncate orbitals from the exact set of link orbitals, the discarded link orbitals can be electron donating, or electron accepting, with respect to the primary region. If we truncate an important electron donating link orbital we expect ΔQ_p to be positive, and vice versa for an electron accepting link orbital. The relative error between the GHO and the optimal link orbital can be understood in terms of their relative electron accepting/donating character, or electronegativity.

Figure 5 shows that on average, the GHO is worse at reproducing the primary region charge than the optimal link orbital, as expected. In the neutral and radical species, we find that larger errors are obtained with the primary link orbital when the primary or secondary region contains the functional groups R = CH_3O and X = OH. When we examine the link orbitals that are omitted, they are involved in redistributing the lone-pair electrons of the oxygen atom that lies at the boundary.

Similarly, cations and anions give much larger charge errors ΔQ_p. Charge that normally would be redistributed across multiple orbitals in the secondary region is now accepted by only a single GHO or optimal link orbital. The errors can be quite large: in $CH_3OCH_2O^-$ with O^- as the QM primary region, there is an overestimate of 0.25 electrons for the QM region because the electron-rich O^- is unable to sufficiently redistribute its electrons across the boundary.

The GHO charge error is systematically shifted to the positive of the optimal link orbital error, indicating that the

dx.doi.org/10.1021/ct500512f | J. Chem. Theory Comput. 2014, 10, 3784−3790

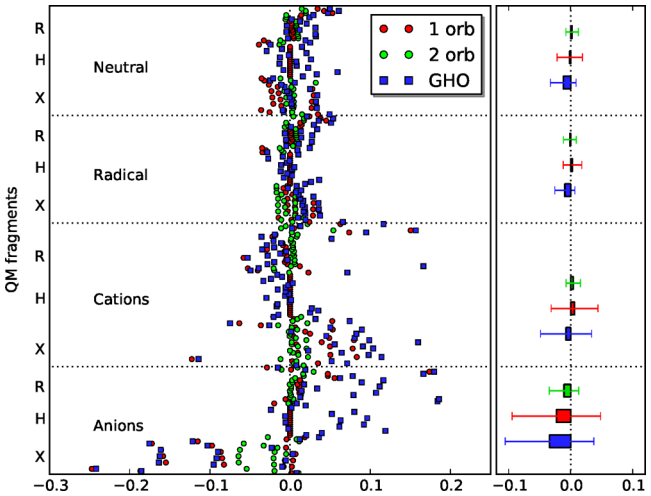


Figure 5. Errors, mean, and STD deviations of fragment Mulliken charges. The reference is the Mulliken charge of the full system SCF. The labeling is as in Figure 3. The side panel (but not main panel) GHO results include an additional orthogonalization of the GHO against the core orbitals, to remove the spurious electronegativity due to neglect of the Pauli potential from the core. See main text.

GHO is substantially more electronegative. This increase in electronegativity is an artifact because (unlike the optimal link orbital) the GHO orbital is not constrained to be orthogonal to the core orbitals in the secondary region that generate the external potential V_p, and we neglect in our calculations (as is standard in GHO treatments) the repulsive Pauli exclusion that would arise from this nonorthogonality. This exacerbates the GHO charge errors for the cations, but for a few anionic species, the GHO charge error is smaller (and of opposite sign) to the optimal link orbital error (despite the mathematical optimality of the latter with respect to the density matrix metric) due to cancellation of the truncation error and the neglect of the Pauli potential. If we correctly account for nonorthogonality by orthogonalizing the GHO against the secondary core orbitals (see side-panel of Figure 5), this removes the artificial electronegativity, and the average GHO charge error is then larger than that of the optimal link orbital in *all* species, including the anions. (The same figure where the side-panel does not include orthogonalization is presented in the Supporting Information).

We next briefly present some energetic data for ionization energies and proton affinities. Table 1 shows the adiabatic ionization energy while Table 2 shows the proton affinity. In the case of the ionization energies, the electrons mainly come from the 2p orbitals of the N and O atoms. Significantly better

Table 2. Proton Affinity (kcal/mol)[a]

	GHO	1 orb	2 orb	ref.
$(CH_3)_2CH[COO^-]$	504.2	494.3	494.3	491.2
$(CH_3)_2CH[NH_2]$	317.3	298.3	286.5	286.8
$(CH_3)_2CH[NH^-]$	603.3	583.6	560.4	549.2
$(CH_3)_2CH[O^-]$	640.7	618.2	582.1	543.5
$CH_3OCH_2[COO^-]$	497.4	490.4	484.2	487.7
$CH_3OCH_2[NH_2]$	313.7	293.9	279.2	276.5
$CH_3OCH_2[NH^-]$	620.6	603.2	552.2	538.0
$CH_3OCH_2[O^-]$	656.7	637.1	567.6	525.6

[a]The positive sign indicates a release of energy by the system.

results are obtained when using an optimal boundary orbital as compared to using a GHO; the error is often less than half that of using a GHO.

We finish our comparison by examining the physical resemblance between the first optimal link orbital and the sp^3 hybrid constructed in the GHO. As seen in Figure 6 for for $CH_3OCH_2O^-$ and $COOHCH_2COOH$, the first optimal link orbital typically has sp^3-like hybrid character. Figure 7 compares the valence components of the GHO and optimal link orbital in the neutral systems. Here the valence components are the orbital coefficients associated with the valence space of the boundary carbon atom. The optimal link orbitals are similar to the GHO active orbitals in the sense that the valence components of both methods are roughly sp^3 hybrids, more specifically, between $sp^{2.5}$ and $sp^{3.5}$ hybrids. Other than the general sp^3 nature, however, we surprisingly do not observe any obvious correspondence between the weights of the different valence components in the optimal link orbital, and weights obtained from the detailed tetrahedral geometry as used in the GHO method. This suggests there may be better empirical ways to choose the degree of sp^3 hybridization, not based on geometry, as in GHO methods.

CATIONS, ANIONS, AND DOUBLE BONDS

Since the error observed for the cations and anions is large even using the optimal link orbital, this means that such boundaries simply cannot be represented using a single link orbital. Additional orbitals are needed to represent charge redistribution effects to the secondary region. In the optimal link orbital construction we can do so by including the link orbital with the second largest eigenvalue $|1 - \lambda|$. Figures 4 and 5 show the effect of using two link orbitals: the improvement is remarkable. In the cations, the performance using two link orbitals is as good as in in the neutral systems. For anions, the standard deviation of the errors is reduced to about one-third of that in the single link orbital formulation. We similarly find in the energetic data in Tables 1 and 2 that the accuracy can be further improved when using two optimal link orbitals rather than a single one.

The second optimal link orbitals for $CH_3OCH_2O^-$ and $COOHCH_2COOH$ are shown in Figure 6. The second optimal link orbital of $CH_3OCH_2O^-$ (with O^- as the primary fragment) is an antibonding orbital on the neighboring CH bond that acts as an electron acceptor for the O^- charge. In $COOHCH_2$–COOH, the second optimal link orbital is a σ antibonding orbital between O–H, which is a part of the hydrogen bond. We see that the link orbitals provide a chemically intuitive visualization of electron donation and transfer.

Using the optimal link orbitals we can also see how to represent QM/MM boundaries not commonly treated in other methods. (Recall that by using n_p link orbitals, the boundary is treated exactly regardless of its nature). We briefly examine the errors (as measured by primary region charge) obtained by using one and two optimal link orbitals when cutting across a double bond, an sp^2 boundary. In Table 3 we see that the errors are larger than for single bonds; cutting across the double bond of $CH_2CHCHCH_2$ and using a single optimal link orbital leads to almost a unit charge error in the primary region. However, using two optimal link orbitals once again brings significant improvement, and the errors fall below 0.02 au.

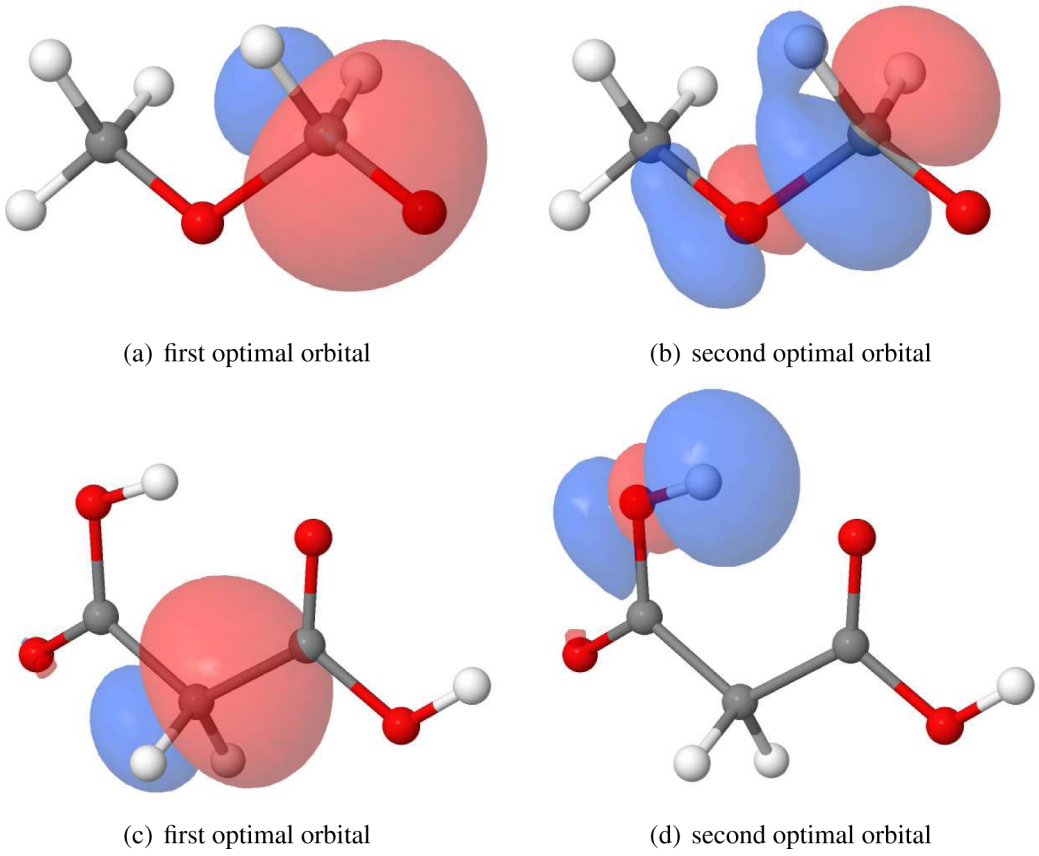


Figure 6. Optimal boundary orbitals of $CH_3OCH_2O^-$ (top: the QM region is O^-) and $COOHCH_2COOH$ (bottom: the QM region is the COOH group on the right). Technical note: for clarity in plotting, we replace the orthogonalized AO basis by the simpler nonorthogonal AO basis, which does not have long-range tails, to generate the orbital contours.

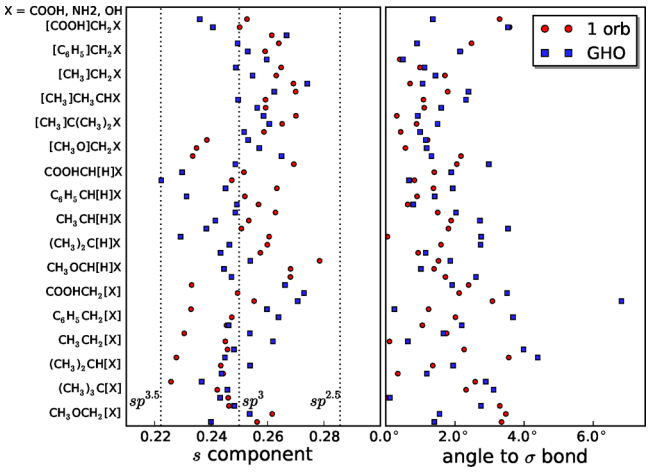


Figure 7. Similarity between the single optimal boundary orbital and the GHO active orbital. (left) s component of the boundary orbital (optimal orbital and GHO active orbital). (right) Angle between the p components of the boundary orbital and the cut bond.

Table 3. QM Fragment Mulliken Charges for the QM Boundaries with sp^2 Hybrids[a]

	1 orb	2 orb	ref.
$[CH_2]CHCHCH_2$	−0.7469	0.0009	−0.0083
$[HCO]OCH_3$	0.2032	0.0482	0.0289
$HCO[OCH_3]$	−0.1424	−0.0335	−0.0289
$[HCO]NH_2$	0.1605	−0.0067	−0.0110
$HCO[NH_2]$	−0.1060	0.0154	0.0110

[a]The atoms grouped in the brackets make up the QM fragment. The reference is the Mulliken charge of full system SCF results.

CONCLUSIONS

In summary, in this work, motivated by investigations in density matrix embedding theory, we defined a mathematically exact set of link orbitals to describe the QM/MM boundary. These orbitals are constructed from the QM solution of the full molecule. The exact link orbitals are simple to obtain (they are the natural orbitals of the secondary region) and small in number: there are (at most) as many exact link orbitals as there are orbitals in the primary region. Truncation of the exact link orbitals is easily achieved from their eigenvalues, and this leads to a subset of link orbitals that are optimal in a mathematically well-defined least-squares density matrix error sense. We have used the optimal link orbitals to analyze the general hybrid orbital (GHO) construction that is widely used for sp^3 QM/MM boundaries. In general, the GHO appears as a reasonably good proxy for the most important optimal link orbital, although our study highlights that it is somewhat artificially electronegative. Further, we find little correlation between the detailed GHO construction, that chooses the weights of valence s and p components based on the bond-angles around the C boundary atom, and the weights determined from the optimal link orbital. For cations and anions, our optimal treatment shows that a single link orbital is in general insufficient, and more link orbitals are required. The second optimal link orbital

dx.doi.org/10.1021/ct500512f | J. Chem. Theory Comput. 2014, 10, 3784−3790

takes the chemically intuitive form of a donor or acceptor orbital that is involved in charge redistribution across the boundary. Finally, using optimal link orbitals, we show that we can cut across double bonds (sp^2 boundaries) as well, using two link orbitals.

While the optimal link orbitals in this work are constructed with a full knowledge of the QM solution for the entire molecule, and thus cannot be used in QM/MM calculations a priori, we can envision several ways in which approximately optimal link orbitals can be obtained in practice. For example, approximately optimal link orbitals can always be practically constructed from a lower-level QM treatment of the molecule than is intended for the primary region. If a standard mean-field (e.g., Hartree−Fock or density functional theory) is used in the QM/MM calculation, then a more approximate QM method, such as density functional tight-binding, or semiempirical molecular orbital, could be used to treat the full molecule and construct (nearly) optimal link orbitals. Another possibility would be to construct approximate link orbitals from a library of boundaries, much as in the molecular test-suite here, which could be parametrized by the type of bond being cut. The ability to cut across arbitrary boundaries, and even treat charge transfer across the QM/MM region, using optimal link orbitals, will significantly reduce the need for very large QM primary regions. The optimal link orbitals described here provide a new interpretational tool for electron transfer across boundaries, with a rigorous identification of the donor and acceptor states. Finally, it is natural to consider the use of optimal link orbitals not only in QM/MM calculations, but also in QM/QM embedding, both in their original density matrix embedding context, as well as in recent density functional embedding schemes.[27−31]

■ ASSOCIATED CONTENT

Supporting Information

Supplementary figure of Mulliken charges (analogous to Figure 5) and the supplementary table of adiabatic ionization energies (analogous to Table 1) with a cc-pVDZ basis set. This material is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gkchan@princeton.edu.

Funding

This work was primarily supported by the Department of Energy, Office of Science, through grant no. DE-SC0010530. Additional support was provided by the Department of Energy, Office of Science, through grant no. DE-SC0008624.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Gao, J. *Rev. Comput. Chem.* **2007**, *7*, 119−185.
(2) Morokuma, K. *Philos. Trans. R. Soc., A* **2002**, *360*, 1149−1164.
(3) Lin, H.; Truhlar, D. *Theor. Chem. Acc.* **2007**, *117*, 185−199.
(4) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720−1735.
(5) König, P. H.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082−9095.
(6) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(7) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(8) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(9) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290−9295.
(10) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46−54.
(11) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534−10547.
(12) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578−9584.
(13) Amara, P.; Field, M. J. *Theor. Chem. Acc.* **2003**, *109*, 43−52.
(14) Ferenczy, G. G.; Rivail, J.-L.; Surján, P. R.; Náray-Szabó, G. *J. Comput. Chem.* **1992**, *13*, 830−837.
(15) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269−282.
(16) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100−106.
(17) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442−1457.
(18) Ferré, N.; Assfeld, X.; Rivail, J.-L. *J. Comput. Chem.* **2002**, *23*, 610−624.
(19) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714−4721.
(20) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. *Theor. Chem. Acc.* **2000**, *104*, 336−343.
(21) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632−650.
(22) Jung, J.; Choi, C. H.; Sugita, Y.; Ten-no, S. *J. Chem. Phys.* **2007**, *127*, 204102.
(23) Knizia, G.; Chan, G. K.-L. *Phys. Rev. Lett.* **2012**, *109*, 186404.
(24) Knizia, G.; Chan, G. K.-L. *J. Chem. Theory Comput.* **2013**, *9*, 1428−1432.
(25) Bulik, I. W.; Scuseria, G. E.; Dukelsky, J. *Phys. Rev. B* **2014**, *89*, 035140.
(26) Neese, F. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78.
(27) Cortona, P. *Phys. Rev. B* **1991**, *44*, 8454−8458.
(28) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050−8053.
(29) Huang, P.; Carter, E. A. *J. Chem. Phys.* **2006**, *125*, 084102.
(30) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. *J. Chem. Theory Comput.* **2012**, *8*, 2564−2568.
(31) Fux, S.; Jacob, C. R.; Neugebauer, J.; Visscher, L.; Reiher, M. *J. Chem. Phys.* **2010**, *132*, 164101.