JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Asymptotic Expansion for Electrostatic Embedding Integrals in QM/MM Calculations

Aurelio Alvarez-Ibarra,*[,†] Andreas M. Köster,[†] Rui Zhang,[‡] and Dennis R. Salahub[‡]

[†]Departamento de Química, CINVESTAV. Avenida Instituto Politécnico Nacional 2508, A.P. 14-740 México, D.F. 07000, México

[‡]Department of Chemistry and Institute for Biocomplexity and Informatics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

ABSTRACT: In QM/MM studies with large MM regions, the calculation of electrostatic embedding integrals can become a computational bottleneck. To overcome this problem, an asymptotic expansion for nuclear attraction-type integrals is developed. As a result, the long-range interactions between the QM and MM atoms reduce to atom-centered multipole moment-like expansions. The algorithm uses a natural spatial division of the molecular structure. To further improve the computational performance, a cutoff radius for the multipole moment-like expansion is introduced. The new code was validated and benchmarked with deMon2k/CHARMM QM/MM calculations on an RNA polymerase II model with almost 350 000 atoms. It is shown that the computational time for the calculation of the embedding integrals in this system can be reduced below 200 s on a small parallel architecture (eight cores) without a loss of accuracy.

INTRODUCTION

Long-range electrostatic interactions have been demonstrated to be crucial in many processes involving materials and biological molecules. Their importance has been shown in the growth of nanomaterials,[1] surface adsorption,[2] colloidal aggregation,[3] protein association,[4] protein folding,[5] and membrane protein function.[6] In agreement with experimental studies, theoretical calculations have also shown long-range electrostatics to be an essential component of simulations.[7,8] Ignoring these interactions can result in serious qualitative errors.[9,10]

Methods to calculate long-range electrostatic interactions have been well developed in classical molecular mechanics (MM) simulations. Their chief use is to mimic the solvent effects correctly. Typically, atoms involved in long-range interactions are either treated explicitly as point charges or implicitly as a potential. A representative of the explicit treatments is the widely used particle mesh Ewald (PME) method,[11] which is an adaptation of the regular Ewald sum method[12] for calculating the full electrostatic energy of a unit cell in a macroscopic lattice of repeated images. An example for the implicit treatments is the recently developed generalized solvent boundary potential (GSBP) method,[13] which is modified from the spherical solvent boundary potential (SSBP) method[14] for treating the outer range of a macromolecular system in a single cell half implicitly and half explicitly. As a comparison, PME calculates the long-range interactions accurately while its cost increases dramatically with the size of the system, whereas GSBP scales well, but its implicit approximation is not as accurate as that of PME.

Adopting successful examples from the MM realm, there have been some encouraging efforts in the quantum mechanics/molecular mechanics (QM/MM) framework to account for the long-range electrostatic interactions between the QM and MM atoms. Following the line of PME, two major classes of developments have been introduced in QM/MM. In one of these, in the framework of density functional theory (DFT) by Laino et al.,[15] a multigrid approach is employed where the long-range electrostatic potentials of the QM and MM atoms are represented in terms of Gaussian functions with different cutoffs. The other approach uses Mulliken charges for the QM atoms when calculating interactions between images of adjacent unit cells under periodic boundary conditions.[16−18] While the multigrid method accounts for the long-range electrostatics accurately, its construction of the grids from the MM Gaussian functions is completed at high cost. In the other class, although the computational cost is considerably reduced by introducing Mulliken charges, there are defects to represent the QM atoms with this type of charge considering its drawbacks due to its basis set dependency.[19] However, when the unit cells are large enough, this approximation is nonetheless suitable and well-justified for systems with periodic boundary conditions.

As an extension of treating the long-range charges implicitly in MM, the GSBP method was first introduced in QM/MM by Cui and co-workers,[20] where all the atoms in the "outer region," beyond a certain distance from the QM region, are mimicked by a boundary potential to reduce the computational cost. This was later improved by Thiel and co-workers by making the boundary potential polarizable[21] and employing representative point charges at the boundary to accelerate the SCF calculation.[22] While GSBP is more economical for sizable systems, one needs to be careful about choosing a proper dielectric constant for a particular macromolecular environment. Since this parameter could be system-dependent, benchmark calculations may be entailed in some cases. When the choice of this parameter is justified, this approach is quite efficient and yields adequate results.

Received: July 17, 2012
Published: August 30, 2012

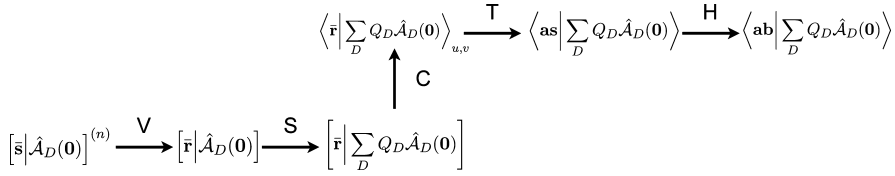

Figure 1. Pathway diagram for the calculation of the electrostatic embedding integral sum.

In view of the disadvantages of PME-related and GSBP-related methods, we report here an asymptotic integral expansion (AIE) approach to calculate the long-range electrostatic interactions between the QM and MM atoms. The application of the AIE is based on the definition of the extension of the orbitals of the QM atoms. If the orbital spatial extension does not contain the point charge, then the AIE can be applied without loss of accuracy. The asymptotic expansion of the nuclear attraction-type embedding integrals yields atom-centered multipole moments that interact with the point charges of all the MM atoms. The higher accuracy of multipoles compared with other simple models like Mulliken charges for atomic charge distributions has been demonstrated in the auxiliary density functional theory (ADFT)-based developments of polarizable force fields.[23,24] The computational efficiency of multipole-based long-range interactions has been shown in both QM calculations[25,26] and MM calculations.[27,28] In this paper, we will show results of QM/MM calculations with deMon2k[30] and CHARMM[31,32] employing our new AIE approach for systems with 124 QM atoms and up to 343 176 MM atoms. We then discuss these results in terms of accuracy and efficiency.

THEORY AND IMPLEMENTATION

Calculation of Electrostatic Embedding Integrals. With an electrostatic point charge embedding the energy expression in deMon2k takes the following form:

$$E = E^{(0)} - \sum_{a,b} P_{ab} \sum_{D} Q_D \langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle + \sum_{A}\sum_{D} \frac{Z_A Q_D}{|\vec{A} - \vec{D}|} \tag{1}$$

where $E^{(0)}$ is the energy of the QM part of the system, Q_D is an embedding charge at position $\vec{D}$, and $\hat{\mathcal{A}}$ is a nuclear attraction-type operator defined as

$$\hat{\mathcal{A}}_D(\mathbf{k}) = \left(\frac{\partial}{\partial D_x}\right)^{k_x}\left(\frac{\partial}{\partial D_y}\right)^{k_y}\left(\frac{\partial}{\partial D_z}\right)^{k_z}\frac{1}{|\vec{r} - \vec{D}|} \tag{2}$$

with $\mathbf{k}$ being an array of three non-negative integer numbers (k_x,k_y,k_z). With this, $\hat{\mathcal{A}}_D(\mathbf{0})$ denotes $1/(|\vec{r}-\vec{D}|)$. The atomic orbitals $\mathbf{a}$ and $\mathbf{b}$ are contracted Cartesian Gaussian type orbitals (GTOs).[33] P_{ab} denotes an element of the density matrix defined by the molecular orbital coefficients $\{c_i\}$ and the occupation numbers $\{n_i\}$ as

$$P_{ab} = \sum_{i}^{\text{occ}} n_i c_{ai} c_{bi} \tag{3}$$

The last term in eq 1 represents the classical Coulomb interaction between the nuclei A with charge Z_A and the embedding charges Q_D. Since this term can be calculated analytically, it is not considered in the self-consistent field (SCF) procedure. Knowing this, the resulting core Hamiltonian matrix element for the embedding SCF is given by

$$H_{ab} = H_{ab}^{(0)} - \sum_{D} Q_D \langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle \tag{4}$$

where $H_{ab}^{(0)}$ is the core Hamiltonian element of the QM part by itself. Thus, the calculation of the integral sum

$$\sum_{D} Q_D \langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle \tag{5}$$

is the difference between an embedded and an isolated QM calculation. Although efficient recurrence relations for this integral sum have been developed and implemented,[34] increasing the number of embedding charges to many thousands turns this part of the calculation into a computational bottleneck. Due to this, it is necessary to improve the algorithm for the calculation of this integral sum, particularly in large scale QM/MM simulations, where the embedding box can contain many hundreds of thousands of MM atoms.

As a reference, we depict in Figure 1 the pathway diagram for the calculation of the integral sum in deMon2k. For the notation and a detailed discussion on this integral sum, several references are available.[25,34,35] The square and angular brackets denote integrals of primitive and contracted GTOs, respectively. The pathway diagram is read from bottom left to top right. The calculation of the integral sum (5) as a whole rather than the calculation of the individual integrals is characteristic of this algorithm. As a result, the contraction (C), transformation (T), and horizontal (H) recurrence steps are outside the three-center loops and, therefore, scale quadratically with the number of basis functions. Thus, only the vertical recurrence (V) and the summation (S) steps remain in the three-center loop. A detailed analysis shows that the computational bottleneck in the calculation of the integral sum (5) arises mainly from the calculation of the incomplete gamma function for the basic primitive integrals (lower left corner in Figure 1). Thus, avoiding this step is crucial for improving the computational performance of large scale QM/MM calculations.

Asymptotic Expansion. When two electronic densities do not overlap significantly, i.e., their respective centers are far away from each other, they can be considered in the asymptotic region of each other, and an asymptotic expansion can be used to accurately calculate their interaction energy.[36] This situation is obviously found in molecular systems largely extended in space, such as solids, solutions, and large biological molecules which are the main targets of QM/MM calculations. To employ the asymptotic expansions, it is mandatory to separate the contributions to the integral sum (5) into a close neighbors region (called the near-field) and a distant neighbors region (called the far-field):

$$\sum_D Q_D\langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle = \sum_D^{\text{Near}} Q_D\langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle + \sum_D^{\text{Far}} Q_D\langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle \quad (6)$$

Then, the two contributions can be treated and calculated in separate ways. Here, we will focus on the far-field contribution. For the accuracy of the whole approach, the reliable definition of the asymptotic region is of utmost importance. In order to apply the asymptotic expansion, the orbitals **a** and **b** must have a negligible overlap with the embedding charge Q_D. To probe this situation, asymptotic radii are defined for atomic orbitals in deMon2k as[25]

$$r_\chi = \sqrt{\frac{\ln \frac{d_{\min}}{\tau}}{\zeta_{\min}}} \quad (7)$$

where $d_{\min}$ denotes the contraction coefficient for the primitive basis function with the smallest exponent contributing to the contracted orbital χ, and τ is the desired accuracy for the integral calculated over the asymptotic expansion (default value 10^{-10} au). For a given atom, the asymptotic atomic radius is defined as the largest of its asymptotic orbital radii. The multipole expansion of the embedding integral $\langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle$ is used when the distance $\overline{AD}$ is larger than the asymptotic radius of orbital **a** and distance $\overline{BD}$ is larger than the asymptotic radius of orbital **b**.

For the asymptotic integral expansion, the nuclear attraction-type operator in the far-field summation in eq 6 is expanded in a Taylor series. On the basis of the primitive product center, the asymptotic expansion has its origin on either atom A or B. The corresponding operators are rewritten as

$$\frac{1}{|\vec{r}-\vec{D}|} \equiv \frac{1}{|(\vec{A}-\vec{D})+\vec{r}_A|} \text{ or } \frac{1}{|\vec{r}-\vec{D}|} \equiv \frac{1}{|(\vec{B}-\vec{D})+\vec{r}_B|} \quad (8)$$

with $\vec{r}_A = \vec{r} - \vec{A}$ and $\vec{r}_B = \vec{r} - \vec{B}$. Since the orbital centers (on $\vec{A}$ and $\vec{B}$) are both far away from the embedding point charge (on $\vec{D}$), $|\vec{r}_A| \ll |\vec{A}-\vec{D}|$ and $|\vec{r}_B| \ll |\vec{B}-\vec{D}|$ hold always. This ensures convergence of the series. Using the Taylor series expansion for the operator at atom A (eq 8), we find

$$\frac{1}{|(\vec{A}-\vec{D})+\vec{r}_A|} \sim \sum_{m_x,m_y,m_z} \hat{M}_A(m_x, m_y, m_z)\, T_{AD}(m_x, m_y, m_z) \quad (9)$$

with

$$\hat{M}_A(m_x, m_y, m_z) = \frac{(-1)^m}{m_x!m_y!m_z!}(x-A_x)^{m_x}(y-A_y)^{m_y}(z-A_z)^{m_z} \quad (10)$$

and

$$T_{AD}(m_x, m_y, m_z) = \left(\frac{\partial}{\partial D_x}\right)^{m_x}\left(\frac{\partial}{\partial D_y}\right)^{m_y}\left(\frac{\partial}{\partial D_z}\right)^{m_z}\frac{1}{|\vec{A}-\vec{D}|} \quad (11)$$

with $m = m_x + m_y + m_z$. Inserting eqs 10 and 11 into eq 9 and the latter in the far-field summation in eq 6 yields

$$\sum_D^{\text{Far}} Q_D\langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle \sim \sum_D^{\text{Far}} Q_D \left\langle \mathbf{ab}\left| \sum_{m_x,m_y,m_z}\left(\frac{\partial}{\partial D_x}\right)^{m_x}\left(\frac{\partial}{\partial D_y}\right)^{m_y}\left(\frac{\partial}{\partial D_z}\right)^{m_z}\frac{1}{|\vec{A}-\vec{D}|} \times \frac{(-1)^m}{m_x!m_y!m_z!}(x-A_x)^{m_x}(y-A_y)^{m_y}(z-A_z)^{m_z}\right.\right\rangle \quad (12)$$

The factors $(i-A_i)^{m_i}$ will increase the angular momentum index i of orbital **a**. Rearranging and simplifying eq 12 leads to the following asymptotic form of the far-field embedding integral sum in eq 6:

$$\sum_D^{\text{Far}} Q_D\langle \mathbf{ab}|\hat{\mathcal{A}}_D(\mathbf{0})\rangle \sim \sum_{m_x,m_y,m_z}\frac{(-1)^m}{m_x!m_y!m_z!}\langle \mathbf{a}+\mathbf{m}|\mathbf{b}\rangle \times \sum_D^{\text{Far}} Q_D T_{AD}(m_x, m_y, m_z) \quad (13)$$

where **m** denotes the momentum index vector (m_x,m_y,m_z). The $\langle \mathbf{a}+\mathbf{m}|\mathbf{b}\rangle$ integrals are overlap-type integrals that can be efficiently calculated via the Obara–Saika recurrence relations.[37] They represent the multipole moment analogs as obtained from the asymptotic integral expansions. When the nuclear attraction-type operator (eq 8) is redefined, the multipole expansion center is shifted to either center $\vec{A}$ or center $\vec{B}$. This improves the performance of the code since the expansion center is always located at an atom and no extra space division algorithm is required. However, the shift of the primitive orbital product center to an atom requires higher order multipoles in the sum (eq 9). In our experience, expanding this sum up to the eighth order keeps the accuracy of the original deMon2k code (i.e., the total energy obtained is consistent within 10^{-6} au). Thus, we employed eighth order expansions in eq 9. For larger distances between the orbital centers and the embedding charge Q_D, faster convergences of the asymptotical integral expansion can be found. With this in mind, we define a cutoff distance, d, for the multipole sum (eq 9) as

$$d = \sqrt[(m+1)]{\frac{Q_{\max}S_{\max}}{\tau}} \quad (14)$$

Here, m refers to the cutoff order, i.e., the (reduced) asymptotic expansion order used for all charges beyond the distance d. In our actual implementation, we employ a cutoff order of 2, i.e., analogous to a multipole expansion up to the quadrupole moment contribution. $Q_{\max}$ refers to the largest embedding charge magnitude in the embedding box. $S_{\max}$ refers to the largest tensor norm generated by the overlap-like integrals $\langle \mathbf{a}+\mathbf{m}|\mathbf{b}\rangle$ in the far-field region. As for the asymptotic orbital radii, τ denotes the threshold for the desired integral accuracy. The above-defined cutoff distance has proven to conserve total energies within 10^{-6} au compared to the original integral sum.

Implementation Details. Analysis of eq 13 shows that the asymptotic expansion of the far-field contributions of the embedding integral sum leads to a factorization into two parts,

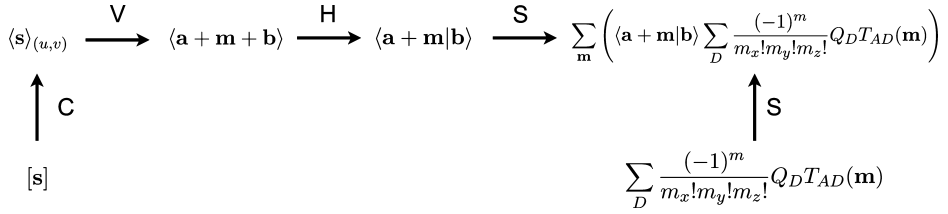

Figure 2. Pathway diagram for the calculation of the asymptotic embedding integral sum.

namely, an atom-dependent part that consists of the embedding charges and the T_{AD} derivatives and an orbital-dependent part that consists of overlap-type integrals. As a result, the sum over the embedding charges is free of any integral calculation. Obviously, this is an enormous computational advantage if a large number of embedding charges has to be processed. The other major simplification arises from the replacement of nuclear attraction-type integrals by overlap-type integrals. This eliminates the need to calculate the incomplete gamma function in the asymptotic expansion of the integral sum. As a consequence, the computational bottleneck will shift away from the embedding sum calculation. In comparison to fast multipole methods (FMM),[29] the asymptotic integral sum expansion, as implemented here, does not rely on an extra division of space. Instead, the molecular structure by itself, here represented by the QM and MM atoms, defines naturally the expansion centers since the embedding integral sum division (eq 6) is performed on the basis of the asymptotic radii of the atoms, defined by their largest asymptotic orbital radii (eq 7). The near-field integral sum is calculated according to eq 5 and the pathway diagram depicted in Figure 1, whereas the far-field integral sum is calculated via the asymptotic expansion, eq 13. The corresponding pathway diagram is depicted in Figure 2. It consists of the calculation of modified overlap integrals by contraction (C), vertical (V), and horizontal (H) recurrence steps that are summed (S) together with an atomic embedding charge sum. As already discussed and shown in this diagram, the overlap-type integral calculation and the embedding charge summation are separated in the asymptotic integral sum calculation. As a consequence, the embedding charge summation can be moved to the outer atomic loop as shown in the pseudocode in Figure 3. To further improve the computational performance, both embedding charge integral sums, near- and far-field, have been parallelized.

```
DO ATOM_A
  DO CHARGE_D
     Calculate Σ_D (-1)^m/(m_x! m_y! m_z!) Q_D T_AD(m)
  END DO
  DO ATOM_B
     DO SHELL_A
        DO SHELL_B
           Calculate ⟨a+m|b⟩
           Build Σ_m Σ_D (-1)^m/(m_x! m_y! m_z!) Q_D T_AD(m) ⟨a+m|b⟩
        END DO
     END DO
  END DO
END DO
```

Figure 3. Pseudocode for the calculation of the asymptotic embedding integral sum.

Computational Method. All calculations were performed within auxiliary density functional theory[38] as implemented in the deMon2k program.[30] The Coulomb energy was calculated by the variational fitting procedure proposed by Dunlap et al.[39,40] For the fitting of the density, the auxiliary function set A2[41] was used in all calculations. All electrons were treated explicitly using the double-ζ valence-polarization (DZVP) basis.[41] The generalized gradient approximation, as proposed by Perdew et al.,[42] has been applied. The exchange-correlation potential was numerically integrated on an adaptive grid.[43] The grid accuracy was set to 10^{-5} au in all calculations.

The embedding charges are taken from the CHARMM 27 force field parameters.[44,45] Positions of the embedding charges and QM atoms are taken from a snapshot of a 1-ns-long molecular dynamics trajectory of the RNA polymerase II system.[46] The QM system is capped with hydrogen atoms at the boundary with the embedding charges.

RESULTS AND DISCUSSION

QM/MM System. The tested system for our asymptotic expansion implementation is the model we have built for studying the nucleotidyl transfer reaction involving RNA polymerase II. The entire system contains 343 300 atoms including solvent molecules. The QM system is defined based on the analysis of the ligand and the protein, which results in a QM subsystem of 124 atoms with a charge of −2 and a singlet multiplicity. The rest of the system is defined as the MM subsystem. To study the long-range electrostatics, we vary the size of the MM part and calculate the electrostatic interaction between the QM subsystem and MM atoms. The number of MM atoms are 1145, 2479, 3309, 4527, 5946, 6851, 7677, 8573, 9382, 10 614, 14 034, 59 259, and ultimately 343 176, which consists of the entire MM subsystem.

In Figure 4, some representative QM/MM systems calculated in this work are depicted. First, the ribbon model of the whole RNA polymerase II (A) with the QM part in blue and the MM part in orange is shown. Following this, some of the embedded systems are shown. These correspond to 7677 (B), 9382 (C), 14 024 (D), 59 259 (E), and 343 176 (F) embedding charges.

Performance. In Table 1, the number of integrals in the embedding calculations is shown. The far-field integrals are separated into *full sum* (with the asymptotic expansion order up to eighth) and *cutoff sum* (with the asymptotic expansion order up to second). Even for the second smallest test system, with 2479 embedding charges, the near-field integrals are already saturated as the comparison with the other calculations shows. Thus, from this system on, the number of far-field integrals increases. This number grows rapidly as Table 1 shows. From 5946 embedding charges, some far field integrals are evaluated with the cutoff sum, i.e., with asymptotic expansion sums (eq 13) only up to second order. The number of these integrals increases moderately until the second largest system with

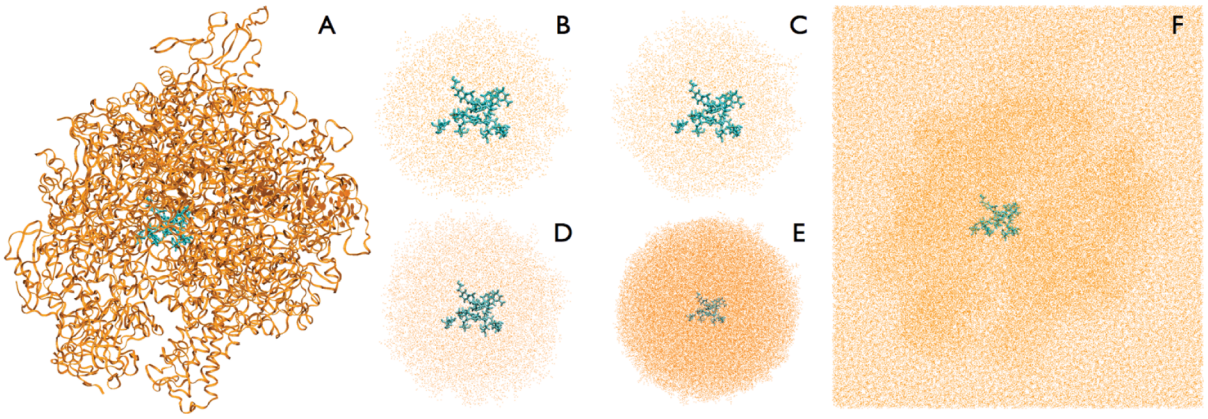

Figure 4. Selected QM/MM systems used for the validation of the asymptotic embedding integral sum calculation.

Table 1. Number of Integrals in the Embedding Calculations (DZVP Basis Set)

embedding charges in box	number of near-field integrals	number of far-field integrals (full sum)	number of far-field integrals (cutoff sum)
1145	63 766 082	340 788 753	0
2479	67 686 417	808 201 300	0
3309	67 762 848	1 101 382 959	0
4527	67 762 848	1 531 730 373	0
5946	67 762 848	2 033 093 450	2260
6851	67 762 848	2 352 777 039	75 986
7677	67 762 848	2 644 578 206	119 617
8573	67 762 848	2 960 792 766	482 465
9382	67 762 848	3 245 326 149	1 787 389
10614	67 762 848	3 676 953 910	5 453 564
14034	67 762 848	4 837 018 040	53 754 094
59259	67 762 848	10 498 227 669	10 371 577 140
343176	67 762 848	10 498 321 422	110 689 422 808

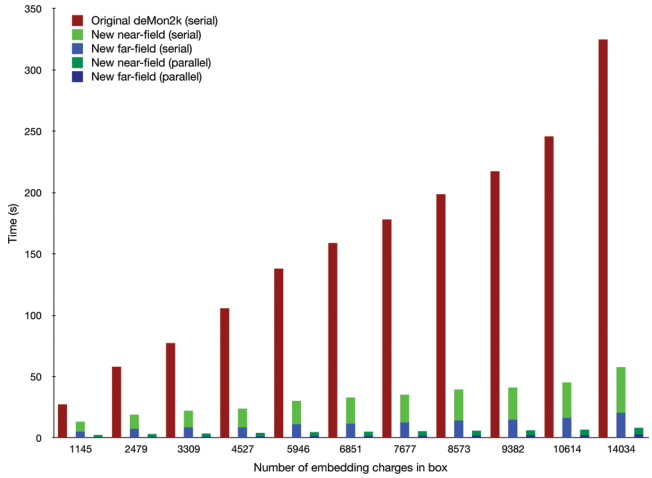

Figure 5. Comparison of computational timings for the embedding integral calculation in small QM/MM systems.

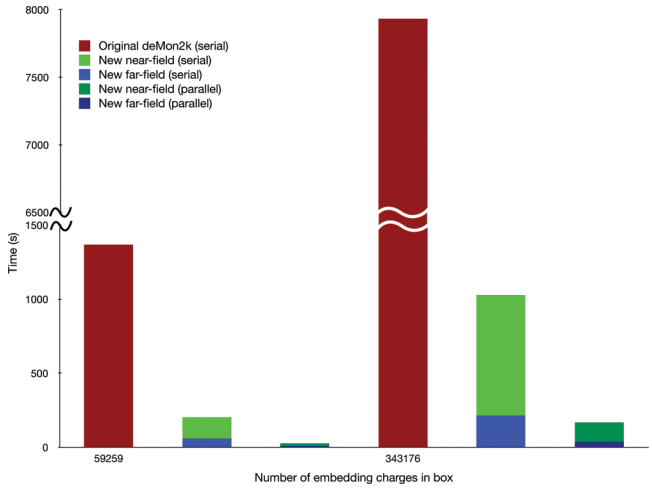

Figure 6. Comparison of computational timings for the embedding integral calculation in large QM/MM systems.

59 259 embedding charges. In this system, the number of far-field integrals with expansions up to eighth and second order are practically the same. In the largest system, the number of far-field integrals with eighth order expansion remains constant, indicating saturation in this region, whereas the number of far-field integrals with second order expansion increases by 1 order of magnitude. This indicates that a further increase of the embedding box will add only cutoff sum integrals. Certainly a second cutoff radius could be introduced beyond which only zeroth order (charge) contributions are considered. This shows that our asymptotic integral expansion ultimately turns into a point charge picture as used in most MM approaches. As a consequence, every acceleration method for point charge interactions, e.g., fast multipole methods, can be used in combination with the here discussed asymptotic integral expansion without a loss of accurracy. This will be relevant for systems with many millions of MM atoms.

In Figures 5 and 6, computational timings for the embedding integral calculation in the studied QM/MM systems are shown. The red bars correspond to the embedding integral sum timings of the original deMon2k code, i.e. using eq 5. Since this part of the code was not parallelized nor separated into near- and far-field contributions, only one bar represents the time for the whole integral sum calculation. The light blue and light green bars represent the serial computational timings for the calculation of the near-field and far-field embedding integral sum with the new algorithm, respectively. Since these calculations are now separated into two different steps, timings for each step are given. Both bars together represent now the full timing for the embedding integral calculation (eq 6) and thus can be directly compared to the original deMon2k timings (red bars). The dark blue and dark green bars depict the

dx.doi.org/10.1021/ct300609z | J. Chem. Theory Comput. 2012, 8, 4232−4238

parallel computational timings for the calculation of the near-field and far-field embedding integrals on eight processors.

The results for the new algorithm in Figures 5 and 6 show a considerable reduction in the time for the integral sum calculations even for the smallest test systems. This is due to the improved computational performance of the asymptotic integral expansion as already discussed above. As expected, the computational savings increases rapidly with the system size because of the rapid saturation of the number of near-field integrals (see Table 1). Comparison of our serial and parallel implementations of the asymptotic integral sum calculation indicates almost perfect scaling with respect to the number of processors. For the largest system (with 343 176 embedding charges) we find a computational speed up relative to the original embedding sum implementation in deMon2k, by a factor of 7 for serial runs and a factor of more than 45 for eight-processor parallel runs. The 200 s (Figure 6) for the embedding integral calculation represents only a fraction of the total QM/MM step time that is in the range of 5000 s. Thus, our asymptotic integral expansion removes the embedding integral bottleneck in large scale QM/MM calculations without jeopardizing the accuracy of the QM methodology. To analyze this in more detail, we list in Table 2 the total energies for the studied embedded systems calculated with the original embedding integral sum (eq 5) and the new near-field and far-field separation (eq 6). The far-field integrals in eq 6 are asymptotically expanded according to eq 13 with eighth and second order expansions as discussed. Table 2 shows that our new serial and parallel code employing the asymptotic expansion of the embedding integral sum reproduces the original embedding sum energies within the default 10^{-6} au accuracy. Note also the variations of the total energies in Table 2. Clearly, this underlies the importance of the accurate calculation of the long-range electrostatic interactions in QM/MM studies.

Table 2. Comparison of Total Energies (au) from eqs 5 and 6 for the Studied Embedded Systems (See Text for Details)

embedding charges in box	equation 5 (serial)	equation 6 (serial)	equation 6 (parallel)
1145	−4880.222132924	−4880.222132917	−4880.222132918
2479	−4880.199888324	−4880.199888324	−4880.199888324
3309	−4880.153319910	−4880.153319912	−4880.153319912
4527	−4880.345254270	−4880.345254270	−4880.345254271
5946	−4880.294915544	−4880.294915542	−4880.294915542
6851	−4880.361831981	−4880.361831929	−4880.361831929
7677	−4880.317036142	−4880.317036090	−4880.317036090
8573	−4880.674560899	−4880.674560783	−4880.674560783
9382	−4880.730888314	−4880.730887986	−4880.730887986
10614	−4880.787864272	−4880.787864196	−4880.787864197
14034	−4880.582718647	−4880.582718714	−4880.582718715
59259	−4880.430977078	−4880.430977870	−4880.430977871
343176	−4880.453535024	−4880.453535264	−4880.453535265

CONCLUSIONS

In this paper, we present the working equations for the asymptotic multipole expansion of long-range electrostatic embedding integrals. We validate this new integral algorithm with QM/MM calculations that contain up to 343 176 embedding charges. Comparisons with conventional embedding integral algorithms show computational speedups by a factor of 7. Because the new integral algorithm is well suited for parallelization, the embedding integral bottleneck is removed from the QM/MM calculations with several hundreds of thousands of MM atoms. The computational performance improvement arises from two sources. First, in contrast to conventional embedding integral algorithms, the sum over the embedding charges in the far field region is free of any integral calculation. Second, the nuclear attraction-type integrals, which require the calculation of the incomplete gamma function, are turned into analytic overlap-type integrals. Because this new embedding integral algorithm is based on asymptotic integral expansions, high accuracies can be reached. Our current implementation with the above-described performance enhancement reproduces total energies to 10^{-6} au or better compared to conventional embedding integral calculations. For even larger systems with millions of MM atoms, a reduction to a point charge interaction scheme is straightforward.

AUTHOR INFORMATION

Corresponding Author

*E-mail: aalvarezi@cinvestav.mx.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by CIAM grant 107310 and by NSERC grant 10174. A.A.-I. gratefully acknowledges a CONACyT Ph.D. student grant (210726). Part of the development was performed at Compute Canada WestGrid and the HPC resource Xiuhcóatl of CINVESTAV. We thank Prof. Sergei Noskov and Bogdan Lev for illuminating discussions and help with the deMon2k/CHARMM interface coding.

REFERENCES

(1) Kong, X. Y.; Ding, Y.; Yang, R.; Wang, S. L. *Science* **2004**, *303*, 1348−1351.
(2) Vandesteeg, H. G. M.; Stuart, M. A. C.; Dekeizer, A.; Bijsterbosch, B. H. *Langmuir* **1992**, *8*, 2538−2546.
(3) Leunissen, M. E.; Christova, C. G.; Hynninen, A. P.; Royall, C. P.; Campbell, A. I.; Imhof, A.; Dijkstra, M.; van Roij, R.; van Blaaderen, A. *Nature* **2005**, *437*, 235−240.
(4) Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol.* **1996**, *3*, 427−431.
(5) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173−191.
(6) Papazian, D. M.; Shao, X. M.; Seoh, S. A.; Mock, A. F.; Huang, Y.; Wainstock, D. H. *Neuron* **1995**, *14*, 1293−1301.
(7) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.
(8) Yeh, I. C.; Berkowitz, M. L. *J. Chem. Phys.* **1999**, *111*, 3155−3162.
(9) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(10) Monticelli, L.; Simoes, C.; Belvisi, L.; Colombo, G. *J. Phys.: Condens. Matter* **2006**, *18*, S329−S345.
(11) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(12) Ewald, P. P. *Ann. Phys. (Berlin, Ger.)* **1921**, *64*, 253−287.
(13) Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(14) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(15) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2006**, *2*, 1370−1378.
(16) Nam, K.; Gao, J. L.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2−13.
(17) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715−17733.

(18) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019−1031.
(19) Wiberg, K. B.; Rablen, P. R. *J. Comput. Chem.* **1993**, *14*, 1504−1518.
(20) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123* (14905), 1−14.
(21) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2008**, *4*, 1600−1609.
(22) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2009**, *5*, 3114−3128.
(23) Elking, D. M.; Perera, L.; Duke, R.; Darden, T.; Pedersen, L. G. *J. Comput. Chem.* **2010**, *31*, 2702−2713.
(24) Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190−202.
(25) Köster, A. M. *J. Chem. Phys.* **2003**, *118*, 9943−9951.
(26) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268−278.
(27) Ding, H. Q.; Karasawa, N.; Goddard, W. A. *J. Chem. Phys.* **1992**, *97*, 4309−4315.
(28) Esselink, K. *Comput. Phys. Commun.* **1995**, *87*, 375−395.
(29) Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1987**, *73*, 325−348.
(30) Köster, A. M.; Geudtner, G.; Calaminici, P.; Casida, M. E.; Dominguez, V. D.; Flores-Moreno, R.; Gamboa, G. U.; Goursot, A.; Heine, T.; Ipatov, A.; Janetzko, F.; del Campo, J. M.; Reveles, J. U.; Vela, A.; Zuniga-Gutierrez, B.; Salahub, D. R. *deMon2k*, version 3; The deMon developers, CINVESTAV: Mexico City, 2011.
(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(32) Lev, B.; Zhang, R.; de la Lande, A.; Salahub, D. R.; Noskov, S. Y. *J. Comput. Chem.* **2010**, *31*, 1015−1023.
(33) Saunders, V. R. Molecular integrals for Gaussian Type functions. In *Methods in Computational Molecular Physics*; Diercksen, G. H. F., Wilson, S., Eds.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1983; pp 1−36.
(34) Quintanar, C.; Caballero, R.; Köster, A. M. *Int. J. Quantum Chem.* **2004**, *96*, 483−491.
(35) Köster, A. M. *J. Chem. Phys.* **1996**, *104*, 4114−4124.
(36) Challacombe, M.; Schwegler, E.; Almlöf, J. *Chem. Phys. Lett.* **1995**, *241*, 67−72.
(37) Obara, S.; Saika, A. *J. Chem. Phys.* **1986**, *84*, 3963−3974.
(38) Köster, A. M.; Reveles, J. U.; del Campo, J. M. *J. Chem. Phys.* **2004**, *121*, 3417−3424.
(39) Dunlap, B. I.; Connolly, J. W.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 4993−4999.
(40) Mintmire, J. W.; Dunlap, B. I. *Phys. Rev. A* **1982**, *25*, 88−95.
(41) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Phys.* **1992**, *70*, 560−571.
(42) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(43) Köster, A. M.; Flores-Moreno, R.; Reveles, J. U. *J. Chem. Phys.* **2004**, *121*, 681−690.
(44) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wirkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(45) MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. *Biopolymers* **2001**, *56*, 257−265.
(46) Zhang, R.; Salahub, D. R. Unpublished, 2012.