Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets

Zachary C. Holden, Ryan M. Richard, and John M. Herbert[a)]
Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

(Received 21 September 2013; accepted 4 December 2013; published online 27 December 2013)

An implementation of Ewald summation for use in mixed quantum mechanics/molecular mechanics (QM/MM) calculations is presented, which builds upon previous work by others that was limited to semi-empirical electronic structure for the QM region. Unlike previous work, our implementation describes the wave function's periodic images using "ChElPG" atomic charges, which are determined by fitting to the QM electrostatic potential evaluated on a real-space grid. This implementation is stable even for large Gaussian basis sets with diffuse exponents, and is thus appropriate when the QM region is described by a correlated wave function. Derivatives of the ChElPG charges with respect to the QM density matrix are a potentially serious bottleneck in this approach, so we introduce a ChElPG algorithm based on atom-centered Lebedev grids. The ChElPG charges thus obtained exhibit good rotational invariance even for sparse grids, enabling significant cost savings. Detailed analysis of the optimal choice of user-selected Ewald parameters, as well as timing breakdowns, is presented. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4850655]

I. INTRODUCTION

Periodic boundary conditions (PBC) are an essential ingredient in condensed-phase simulations. Without them, one is relegated to cluster-based approaches whose convergence must be tested, and where molecular dynamics simulations may require the use of artificial confining potentials. The aim of this work is to develop a method for applying PBC to mixed quantum mechanics/molecular mechanics (QM/MM) simulations. We are interested in electronic process such as solution-phase ionization potentials and electronic excitation energies, where long-range Coulomb forces may be especially important, and where the use of cutoffs or a minimum-image convention may be questionable. In such cases, Ewald summation is the preferred approach.[1,2] Although certain artifacts with Ewald summation have been noted,[3] and alternative approaches have been developed,[4] many of these artifacts are due simply to the artificial nature of introducing PBC in a disordered system, and therefore disappear in the limit of a large simulation cell. The important role of long-range electrostatics in obtaining correct macromolecular structures is well documented.[5–7]

A method for performing Ewald summation in semi-empirical QM/MM calculations has been described by Nam *et al.*[8] and, independently, by Riccardi *et al.*[9] In this approach, periodic images of the QM electron density are collapsed onto Mulliken atomic point charges, and then standard Ewald summation is used to model the interaction between these image charges. A correction term introduced by the presence of the images is incorporated into the Fock matrix, such that the self-consistent field (SCF) procedure remains variational (similar to the charge embedding used in the variational explicit polarization or "XPol" method[10]). Nam *et al.*[8] demonstrated that this QM/MM-Ewald approach leads to correct long-range behavior in the potential of mean force for solvated ions, unlike calculations in which cutoffs replace Ewald summation. The QM/MM-Ewald method has since been implemented in both the AMBER[11] and CHARMM[9,12] simulation packages, but only for semi-empirical wave functions.

Our goal is to generalize this methodology for an arbitrary description of the QM region, which might include a correlated *ab initio* wave function or any of the arsenal of modern density functionals, including hybrid and long-range corrected functionals whose Hartree-Fock exchange component remains expensive to evaluate in a plane-wave basis, despite recent advances.[13] We use atom-centered Gaussian basis sets exclusively, but recognize that basis sets of triple-ζ quality may be necessary, especially for correlated wave functions or double-hybrid functionals, and furthermore diffuse basis functions are needed for anions and are also an effective way to reduce basis set superposition error. However, these requirements are problematic for a method based on Mulliken charges. Semi-empirical implementations of the QM/MM-Ewald method suffer no such problems, as they are based on minimal basis sets for which Mulliken charges are well-behaved. Indeed, we find the method of Ref. 8 to be perfectly stable for Hartree-Fock and density-functional theory (DFT) calculations, provided that minimal basis sets are used.

For larger basis sets, however, the charge derivatives $\partial Q_\alpha/\partial P_{\mu\nu}$ (where Q_α is a Mulliken charge and $P_{\mu\nu}$ is a density matrix element) become unstable, leading to SCF convergence failure. Our group has observed similar instabilities in the past, in attempts to use Mulliken or Löwdin charges for self-consistent charge embedding,[14] in the context of XPol calculations. The solution in that context[14,15] is the same one pursued here, namely, the use of "ChElPG" charges[16] derived from the QM electrostatic potential evaluated on a grid. This approach leads to stable – albeit relatively expensive

a)herbert@chemistry.ohio-state.edu

– derivatives $\partial Q_\alpha / \partial P_{\mu\nu}$, and therefore extends the QM/MM-Ewald method developed by Nam *et al.*[8] to arbitrary basis sets. As such, our implementation is appropriate for use with correlated QM wave functions. To reduce the cost of computing the derivatives $\partial Q_\alpha / \partial P_{\mu\nu}$, we modify the ChElPG algorithm for use with atom-centered grids, which dramatically reduces the required number of grid points without significantly deteriorating the rotational invariance of the total energy or the ChElPG charges themselves.

II. THEORY

A. Ewald summation

Charge–charge interactions decay very slowly with distance and may not become negligible in a calculation until the distance is on the order of hundreds of nanometers. As such, the pairwise sum over such interactions is slowly convergent. In fact, it is only conditionally convergent in a periodically replicated simulation cell,[2] which is the problem that Ewald summation is designed to overcome. This section provides a brief overview of the Ewald summation technique, which also serves to introduce the notation that we will use.

Traditional charge-charge Ewald summation splits the pairwise summation into two parts: a real-space portion, based on a short-range interaction potential whose pairwise sum converges quickly; and a long-range portion based on a slowly varying interaction potential whose pairwise sum converges relatively quickly in reciprocal space. The Coulomb potential is partitioned using the error function (erf) and complementary error function (erfc), according to

$$\frac{1}{r} = \frac{\mathrm{erf}(\eta r)}{r} + \frac{\mathrm{erfc}(\eta r)}{r} \,. \tag{1}$$

Note that $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$. The Ewald parameter η controls the length scale ($\sim\eta^{-1}$) on which the short-range function $\mathrm{erfc}(\eta r)/r$ decays, and thus controls how much of the pairwise Coulomb sum is performed in real space. As η increases, more of the summation is performed in reciprocal space, whereas setting $\eta = 0$ is the same as performing the pairwise sum entirely in real space.

In the context of Ewald summation, the Coulomb energy the simulation cell, E_{cell}, is traditionally partitioned as

$$E_{\mathrm{cell}} = E_{\mathrm{real}} + E_{\mathrm{self}} + E_{\mathrm{recip}} + E_{\mathrm{charge}} + E_{\mathrm{dipole}} \,. \tag{2}$$

The energy components include the real-space energy,

$$E_{\mathrm{real}} = \sum_{\mathbf{n}} \sum_{j}^{N_{\mathrm{MM}}} \sum_{k>j}^{N_{\mathrm{MM}}} q_j q_k \frac{\mathrm{erfc}(\eta |\mathbf{r}_{jk} + \mathbf{n}L|)}{|\mathbf{r}_{jk} + \mathbf{n}L|} \tag{3}$$

(where $\mathbf{r}_{jk} = \mathbf{r}_j - \mathbf{r}_k$), the Coulomb self-energy,

$$E_{\mathrm{self}} = -\frac{\eta}{\sqrt{\pi}} \sum_{j}^{N_{\mathrm{MM}}} q_j^2, \tag{4}$$

and the reciprocal-space energy,

$$E_{\mathrm{recip}} = \sum_{\mathbf{m}\neq\mathbf{0}} \frac{|S(\mathbf{m})|^2}{2L\pi |\mathbf{m}|^2} \exp\left(\frac{-\pi^2 |\mathbf{m}|^2}{\eta^2 L^2}\right). \tag{5}$$

In these expressions, L denotes the length of the (cubic) simulation cell, and $\mathbf{n}$ and $\mathbf{m}$ are real-space lattice vectors, sums over which extend from negative infinity to positive infinity. The quantity $S(\mathbf{m})$ in Eq. (5) is known as the *structure factor*,[1] and is discussed below. The quantities q_j and q_k are point charges, the total number of which is denoted N_{MM}. For brevity, we have presented these equations for the case that the unit cell is cubic. This is not a fundamental limitation of the Ewald formalism, but the non-cubic case would require, *e.g.*, splitting the sum over lattice vectors $\mathbf{n}$ in Eq. (3) into separate sums over n_x, n_y, and n_z, with lengths L_x, L_y, and L_z for each side of the simulation cell.

The final two terms in Eq. (2) warrant some additional comments. The quantity

$$E_{\mathrm{charge}} = -\frac{Q_{\mathrm{tot}}^2 \pi}{2L^3 \eta^2} \tag{6}$$

is known as the *surface charge term*, where

$$Q_{\mathrm{tot}} = \sum_{j}^{N_{\mathrm{MM}}} q_j \tag{7}$$

represents the total charge of the simulation cell. Since the Coulomb energy is divergent if $Q_{\mathrm{tot}} \neq 0$, Ewald summation can only be used to compute the Coulomb energy for a neutral simulation cell, and E_{charge} represents the energy required to surround a charged cell with a charge-compensating membrane of opposite charge. (If the cell is electrically neutral, then $E_{\mathrm{charge}} = 0$.) Artifacts due to Ewald simulation of a charged unit cell have been noted,[17–19] but will not concern us here.

The final component of E_{cell} is the *surface dipole term*:[20]

$$E_{\mathrm{dipole}} = -\frac{\pi}{(2\varepsilon + 1)L^3} \sum_{j,k}^{N_{\mathrm{MM}}} q_j q_k |\mathbf{r}_{jk}|^2. \tag{8}$$

If $Q_{\mathrm{tot}} = 0$ (which is often assumed when this term is discussed in the literature, as for example in Ref. 4), then E_{dipole} is proportional to the square of the dipole moment of the simulation cell. In Eq. (8), we imagine placing the *supercell* (the simulation cell along with all of its periodic images) into a dielectric medium, whose dielectric constant is denoted by ε in Eq. (8).[2,20,21] Often, one assumes "tin foil" boundary conditions ($\varepsilon = \infty$, corresponding to placing the supercell inside of a conductor), in which case $E_{\mathrm{dipole}} = 0$.

The last bit of notation to explain is the quantity $|S(\mathbf{m})|^2$ in Eq. (5). In its most general form, this quantity is defined as

$$|S(\mathbf{m})|^2 = \sum_{j,k}^{N_{\mathrm{MM}}} q_j q_k \exp\left(\frac{2\pi i}{L}(\mathbf{m}\cdot\mathbf{r}_{jk})\right) \,. \tag{9}$$

Despite the appearance of $i = \sqrt{-1}$, the quantity $|S(\mathbf{m})|^2$ is real, as the squared-modulus notation indicates, and can be rewritten in a way that makes this obvious:

$$|S(\mathbf{m})|^2 = \sum_{j,k}^{N_{\mathrm{MM}}} q_j q_k \cos\left(\frac{2\pi}{L}(\mathbf{m}\cdot\mathbf{r}_{jk})\right). \tag{10}$$

More often, this quantity is further simplified by separating $\mathbf{r}_j$ and $\mathbf{r}_k$, which allows the double summation to be recast as

two identical single summations, with the result[1]

$$|S(\mathbf{m})|^2 = \left[\sum_j^{N_{\mathrm{MM}}} q_j \cos\left(\frac{2\pi}{L}(\mathbf{m}\cdot\mathbf{r}_j)\right)\right]^2 . \tag{11}$$

The latter form requires fewer operations to compute, and is therefore preferred. For QM/MM applications, however, some of the $\mathbf{r}_j$ vectors correspond to QM atoms and some to MM atoms, hence the simplifications leading to Eq. (11) will not be possible and Eq. (10) must be used instead.

B. QM/MM and PBC

The QM/MM-Ewald technique introduced by Nam *et al.*[8] is based upon the reasonable assumption that the simulation cell is large compared to the spatial extent of the QM wave function. As such, a large MM "buffer" screens the interaction between the electron density and its periodic images, so collapsing this density onto point charges for the purpose of computing these long-range Coulomb interactions should not engender serious error. Once the density is reduced to point charges, classical Ewald summation can be applied. In this section, we describe the basic theory behind obtaining PBC corrections to the SCF energy and Fock matrix. In developing this theory, we are concerned only with electrostatic interactions, as other QM/MM interactions such as non-bonded Lennard-Jones interactions operate on shorter length scales and PBC implementations based on smooth cutoffs should be fine. Thus, "total" energy will refer to the QM electronic structure energy plus all MM and QM/MM electrostatic interactions; other MM interactions can simply be tacked on to the formulas appearing below.

1. Energy corrections

We first write the total QM/MM supersystem (SS) energy, which includes the interactions between all periodic images, as

$$E_{\mathrm{total}} = E^{\mathrm{SS}}_{\mathrm{QM\text{-}QM}} + E^{\mathrm{SS}}_{\mathrm{QM\text{-}MM}} + E^{\mathrm{SS}}_{\mathrm{MM\text{-}MM}} . \tag{12}$$

The final term, $E^{\mathrm{SS}}_{\mathrm{MM\text{-}MM}}$, can be evaluated using a standard, classical Ewald summation and need not be discussed further. It is helpful to partition the other two terms into interactions between atoms in the simulation cell with other atoms in the simulation cell, which we will call the real-space (RS) interactions, and also interactions between the simulation cell and atoms contained in the periodic images (PI). The SS energies in Eq. (12) can thus be broken down into RS and PI parts:

$$\begin{aligned} E_{\mathrm{total}} =& E^{\mathrm{RS}}_{\mathrm{QM\text{-}QM}} + \Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}QM}} + E^{\mathrm{RS}}_{\mathrm{QM\text{-}MM}} \\ &+ \Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}MM}} + E^{\mathrm{SS}}_{\mathrm{MM\text{-}MM}}. \end{aligned} \tag{13}$$

The term $E^{\mathrm{RS}}_{\mathrm{QM\text{-}QM}}$ (interaction between QM atoms in the simulation cell with other QM atoms in the simulation cell) is simply the result of some QM electronic structure calculation. The term $E^{\mathrm{RS}}_{\mathrm{QM\text{-}MM}}$ results from some QM/MM interaction scheme; note that Eq. (12) tacitly assumes an "additive" QM/MM scheme, as opposed to a "subtractive" scheme such as ONIOM.[22] For the latter, there are no QM periodic images so Ewald summation involves the MM system only, and is therefore straightforward.

The remaining terms in Eq. (13) are calculated as differences between a SS calculation and a RS calculation:

$$\Delta E^{\mathrm{PI}} = E^{\mathrm{SS}} - E^{\mathrm{RS}}. \tag{14}$$

In particular, $\Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}MM}}$ is obtained using a QM region embedded in a periodically replicated supercell of MM regions, but without replication of the QM region. This interaction energy can be decomposed into real- and reciprocal-space parts, the latter of which will involve only MM atoms provided that the QM region is fully enveloped by the short-range part of the Coulomb potential in Eq. (1). The term $\Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}QM}}$ in Eq. (13) is obtained from a periodic array of point charges obtained from collapsing the QM electron density onto atom-centered charges, as described below.

Applying the fundamental assumption that the QM images are far apart and screened by a wide buffer of MM charges, the calculation effectively reduces to a series of pairwise Coulomb interactions between the atoms in the simulation cell and those contained in the periodic images. It is therefore expected that $\Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}QM}}$ and $\Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}MM}}$ will have similar forms:

$$\Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}QM}} = \frac{1}{2}\sum_{\alpha,\beta}^{N_{\mathrm{QM}}} Q_\alpha Q_\beta\, \omega(\mathbf{r}_{\alpha\beta}) \tag{15}$$

and

$$\Delta E^{\mathrm{PI}}_{\mathrm{QM\text{-}MM}} = \sum_{\alpha}^{N_{\mathrm{QM}}}\sum_{j}^{N_{\mathrm{MM}}} Q_\alpha q_j\, \omega(\mathbf{r}_{\alpha j}) . \tag{16}$$

For clarity, we use Q_α to denote the partial charge on a QM atom and q_j to denote the partial charge on an MM atom. As in standard Ewald summation, both quantities can be described by a potential function, $\omega(\mathbf{r})$. For a neutral simulation cell ($Q_{\mathrm{tot}} = 0$) with tin-foil boundary conditions ($\varepsilon = \infty$), this potential is

$$\begin{aligned} \omega(\mathbf{r}_{\alpha\beta}) =& \sum_{\mathbf{m}\neq\mathbf{0}} \frac{e^{-\pi^2|\mathbf{m}|^2/\eta^2 L^2}}{L\pi\,|\mathbf{m}|^2} \cos\left(\frac{2\pi}{L}(\mathbf{m}\cdot\mathbf{r}_{\alpha\beta})\right) \\ &+ \sum_{\mathbf{n}\neq\mathbf{0}} \frac{\mathrm{erfc}(\eta|\mathbf{r}_{\alpha\beta}+\mathbf{n}L|)}{|\mathbf{r}_{\alpha\beta}+\mathbf{n}L|} - \frac{\mathrm{erf}(\eta|\mathbf{r}_{\alpha\beta}|)}{|\mathbf{r}_{\alpha\beta}|}. \end{aligned} \tag{17}$$

In obtaining this result, the form of $|S(\mathbf{m})|^2$ in Eq. (10) has been used in the reciprocal (first) term in Eq. (17). The reason to prefer this form, as opposed to that in Eq. (11), is that the latter requires that each of the indices runs over the same sum, which is not the case for QM/MM interactions where one of the summation indices in Eq. (9) represents QM atoms while the other represents MM atoms.

The potential in Eq. (17) warrants some comments. First, the term containing a sum over $\mathbf{m} \neq \mathbf{0}$ is directly analogous to the reciprocal term in Eq. (5), whereas the erf and erfc terms are analogous to the real space term in Eq. (3). The different appearance of the erf term (representing the $\mathbf{n} = \mathbf{0}$ vector) is due to the fact that an energy $E^{\mathrm{RS}}_{\mathrm{real}}$, with a Coulomb potential of $1/r$, has been subtracted out of $E^{\mathrm{SS}}_{\mathrm{real}}$ with a Coulomb

potential of $\mathrm{erfc}(\eta r)/r$, to afford $\Delta E^{\mathrm{PI}}_{\mathrm{real}}$, with a Coulomb potential of $-\mathrm{erf}(\eta r)/r$.

It should also be noted that nothing analogous to the self-energy is immediately apparent in Eq. (17). In the QM-QM PI correction [Eq. (15)], there is no restriction on the sum and so $\alpha = \beta$ is allowed. For $\alpha = \beta$, the Coulomb interaction is given by

$$\lim_{r\to 0} \frac{\mathrm{erf}\,(\eta r)}{r} = \frac{2\eta}{\sqrt{\pi}}\ , \tag{18}$$

which is in fact the self term. Furthermore, there is no self energy corresponding to the QM-MM PI correction [Eq. (16)], since the atom types in the two summations are different.

Finally, it is worth noting that the potential in Eq. (17) differs from the Ewald potential given by Nam *et al.*,[8] insofar as the term in Eq. (17) containing the sum over $\mathbf{n} \neq \mathbf{0}$ is absent in Ref. 8. The authors of Ref. 8 assume that the Ewald parameter η has been chosen such that only the simulation cell must be considered in the real-space portion of the Ewald sum. This is a reasonable assumption but is not assumed *a priori* in this work, on the basis that cost considerations for more general QM/MM calculations might favor a different partition of the effort. In the case of a charged system, the charge term [Eq. (6)] is included in the MM Ewald summation. We henceforth assume tin-foil boundary conditions and therefore omit the dipole term in Eq. (2).

2. Fock matrix corrections

The corrections above must now be incorporated into the Fock matrix, which is computed by taking the derivative of the energy with respect to the density matrix. Using the chain rule, this correction can be expressed as

$$\Delta F^{\mathrm{PI}}_{\mu\nu} \equiv \frac{\partial \Delta E^{\mathrm{PI}}}{\partial P_{\mu\nu}} = \sum_{\alpha}^{N_{\mathrm{QM}}} \frac{\partial \Delta E^{\mathrm{PI}}}{\partial Q_\alpha} \frac{\partial Q_\alpha}{\partial P_{\mu\nu}}, \tag{19}$$

where $\Delta E^{\mathrm{PI}} = E^{\mathrm{PI}}_{\mathrm{QM\text{-}QM}} + E^{\mathrm{PI}}_{\mathrm{QM\text{-}MM}}$. The energy derivative with respect to an atomic point charge can be evaluated directly from Eqs. (15) and (16):

$$\frac{\partial \Delta E^{\mathrm{PI}}}{\partial Q_\alpha} = \sum_{\beta}^{N_{\mathrm{QM}}} Q_\beta\ \omega(\mathbf{r}_{\alpha\beta}) + \sum_{j}^{N_{\mathrm{MM}}} q_j\ \omega(\mathbf{r}_{\alpha j})\ . \tag{20}$$

The cost of evaluating Eq. (20) can be significantly reduced by recognizing that the Ewald potential depends upon the positions of the atoms (both QM and MM), but not on any details of the electronic structure. Those details are encoded into the QM charges Q_β, which are the only quantities in Eq. (20) that change from one SCF cycle to the next. Thus, we can pre-compute the Ewald potential at the relevant interatomic distances prior to entering the SCF iterations. In anticipation of doing this, let us define a column vector

$$\boldsymbol{\omega}_{\boldsymbol{\alpha}} = \left[\ \omega(\mathbf{r}_{\alpha 1})\ \omega(\mathbf{r}_{\alpha 2}) \cdots \omega(\mathbf{r}_{\alpha N_{\mathrm{QM}}})\ \textstyle\sum_j^{N_{\mathrm{MM}}} q_j \omega(\mathbf{r}_{\alpha j})\ \right]^\dagger. \tag{21}$$

The final entry in this vector is identical to the second term in Eq. (20). Next, define another column vector

$$\mathbf{Q} = [\ Q_1\ Q_2 \cdots Q_{N_{\mathrm{QM}}}\ 1\ ]^\dagger. \tag{22}$$

(Save for the final entry, the vector $\mathbf{Q}$ consists simply of the QM atomic charges.) Using this new notation, we can rewrite Eq. (20) as

$$\Delta F^{\mathrm{PI}}_{\mu\nu} = \sum_{\alpha}^{N_{\mathrm{QM}}} \frac{\partial Q_\alpha}{\partial P_{\mu\nu}}\ \mathbf{Q}^\dagger \boldsymbol{\omega}_{\boldsymbol{\alpha}}. \tag{23}$$

It remains to evaluate the charge derivatives $\partial Q_\alpha / \partial P_{\mu\nu}$. The form of these derivatives depends upon the charge scheme that is used (Mulliken, Löwdin, ChElPG, etc.). For Mulliken or Löwdin charges, these derivatives are quite simple. The Mulliken atomic charges, for example, are defined as

$$Q_\alpha = Z_\alpha - \sum_{\mu\nu\in\alpha} P_{\mu\nu} S_{\mu\nu}\ , \tag{24}$$

hence the requisite derivatives are nothing more than overlap matrix elements:

$$\frac{\partial Q_\alpha}{\partial P_{\mu\nu}} = -S_{\mu\nu}\ \delta_{\mu\nu\in\alpha}\ . \tag{25}$$

Here, $\delta_{\mu\nu\in\alpha}$ is a Kronecker delta-type symbol signifying that both atomic orbital (AO) basis functions μ and ν must be centered on atom α, else the derivative is zero by definition. It is not terribly surprising to discover that the Mulliken charges are unstable in extended basis sets, and because in this context these charges make their way into the Fock matrix, we find that the Mulliken-based QM/MM-Ewald scheme is difficult or impossible to converge in extended basis sets. (Data to this effect are provided in Sec. III; we encounter similar difficulties in attempting to use Mulliken or Löwdin charges in the context of the self-consistent XPol charge-embedding procedure.[14]) ChElPG charges, on the other hand, appear to be stable and robust, but the derivatives $\partial Q_\alpha / \partial P_{\mu\nu}$ are far more costly in the ChElPG case. These derivatives will be given explicitly below, following a discussion of the basic theory behind ChElPG charges.

C. ChElPG charges

1. Basic theory

By construction, the ChElPG atomic charges minimize, in a least-squares sense, the difference between the QM electrostatic potential (evaluated on a grid) and the electrostatic potential derived from a set of atom-centered point charges (evaluated on the same grid), subject to the constraint that the atomic charges must sum to the molecular charge.[16] A complete discussion of the ChElPG formalism, using the same notation that is used here, can be found in Ref. 15. Briefly, the ChElPG charges are given by

$$Q_A = \sum_{B}^{N_{\mathrm{QM}}} e_B \left(\mathbf{G}^{-1}\right)_{BA} - \left(\frac{\sum_{BC}^{N_{\mathrm{QM}}} e_B (\mathbf{G}^{-1})_{BC} - Q_{\mathrm{tot}}}{\sum_{BC}^{N_{\mathrm{QM}}} (\mathbf{G}^{-1})_{BC}} \right) \sum_{B}^{N_{\mathrm{QM}}} (\mathbf{G}^{-1})_{BA}, \tag{26}$$

where

$$e_B = \sum_k^{N_{\text{grid}}} \frac{w_k}{|\mathbf{r}_k - \mathbf{r}_B|} \left(\sum_J^{N_{\text{QM}}} \frac{Z_J}{|\mathbf{r}_k - \mathbf{r}_J|} - \sum_{\mu\nu}^{N_{\text{basis}}} (\mathbf{I}_k)_{\mu\nu} P_{\mu\nu} \right) \quad (27)$$

and

$$G_{AB} = \sum_k^{N_{\text{grid}}} \frac{w_k}{|\mathbf{r}_k - \mathbf{r}_A||\mathbf{r}_k - \mathbf{r}_B|}. \quad (28)$$

The indices A, B, ... index nuclei and k indexes the grid points on which the electrostatic potential is evaluated; the quantity w_k is the weight assigned to the kth grid point. We have previously introduced a weighting scheme that ensures that the charges are continuous functions of the atomic coordinates,[14,15] although a different weighting scheme will be used in this work, as described below. Finally, the quantity

$$(\mathbf{I}_k)_{\mu\nu} = \left\langle \mu \left| \frac{1}{|\mathbf{r} - \mathbf{r}_k|} \right| \nu \right\rangle_{\mathbf{r}} \quad (29)$$

is a charge–density Coulomb integral. (The subscript $\mathbf{r}$ indicates that the electron coordinate $\mathbf{r}$ is the integration variable.)

Historically, the nature of the grid on which to evaluate the electrostatic potential was a source of debate, with various incarnations of the least-squares fitting algorithm using different types of grids. One early algorithm[23] (originally called "ChElP") used a set of concentric, atom-centered spherical grids. However, the charges thus obtained were shown to be sensitive to molecular conformation,[16] which was problematic because a main goal was to use ChElP charges to parameterize force fields. The ChElPG algorithm[16] (so called to distinguish it from ChElP) consists in replacing these spherical grids with a Cartesian grid, deleting points within the van der Waals (vdW) region in order to fit to the long-range parts of the electrostatic potential. Although this reduced the conformational dependence of the charges, it was later demonstrated that atom-centered grids (including those with icosahedral symmetry) lead to far better rotational invariance of the charges, as compared to Cartesian grids.[24]

In the present context, contraction of the integrals $(\mathbf{I}_k)_{\mu\nu}$ with certain quantities that arise in the construction of $\partial Q_\alpha / \partial P_{\mu\nu}$ proves to be a serious bottleneck, and it therefore behooves us to reduce the number of grid points (N_{grid}) as much as possible. Since atom-centered Lebedev grids of octahedral symmetry are already ubiquitous in Gaussian-orbital-based DFT codes, this was the natural choice to explore.

2. Lebedev grid implementation

This section documents our implementation of Eq. (26) using atom-centered Lebedev grids. (Once the grid is constructed, this works like any other ChElPG algorithm, but there are some numerical aspects worthy of discussion.) It could be argued that the charges thus obtained should no longer be called "ChElPG" charges, since the only difference between the ChElP and ChElPG algorithms is how the grid is constructed. However, the ChElPG acronym is widely known and emphasizes the fact that there is a grid-based aspect to the calculation. Thus, we refer to our algorithm as a Lebedev grid-based implementation of the ChElPG charges. We will retain some of the terminology from the original paper on ChElPG charges.[16] Namely, the *head space* refers to the distance from the vdW surface to the outermost radial shells that constitute the grid. (There are no points within the vdW surface.) Also, let Δx denote the spacing between radial shells. These two parameters, along with the number of Lebedev points per shell (N_p), serve to define the ChElPG grid. The Lebedev grid with N_p points on the unit sphere is constructed on each atom and then its radius is scaled by a factor

$$\alpha_i = \mathbf{r}_{\text{vdW}} + (i - 1)\Delta x \quad (30)$$

for the ith shell. Radial shells are constructed from $\mathbf{r}_{\text{vdW}}$ out to the head space distance. Aside from the symmetry of the grids, this procedure is similar to that used by Spackman[24] to evaluate Cartesian versus atom-centered grids, and also to one of the original ChElP algorithms.[23]

Although it is possible to perform a weighted least-squares fit of the electrostatic potential using the weights w_k in Eq. (26), the original ChElPG paper of Breneman and Wiberg[16] sets all $w_k = 1$, and the authors in fact emphasize the importance of using an isotropic grid to reduce conformational dependence of the charges. The use of atom-centered Lebedev grids leads to a highly anisotropic coverage of real space, as is evident from the Lebedev grid for H_2, which is shown in Fig. 1(a). In Fig. 1(b), we plot the number of grid points contained in each 0.25 Å × 0.25 Å cell, which shows how each such cell contributes to the ChElPG fit when all $w_k = 1$. Given the cylindrical nature of the point density we expect radial anisotropy, which is readily apparent in Fig. 1(b), but what is perhaps less intuitive is the fact that there is also anisotropy orthogonal to the bonding plane. The latter arises from the presence of "seams" where the two atom-centered grids meet.

To ameliorate this anisotropy, we propose a simple weighting scheme in which $w_k = 1/n_k$, where n_k is the number of grid points contained within the cell where the point k resides. Figure 1(c) shows that this scheme significantly reduces the anisotropy of the grid.

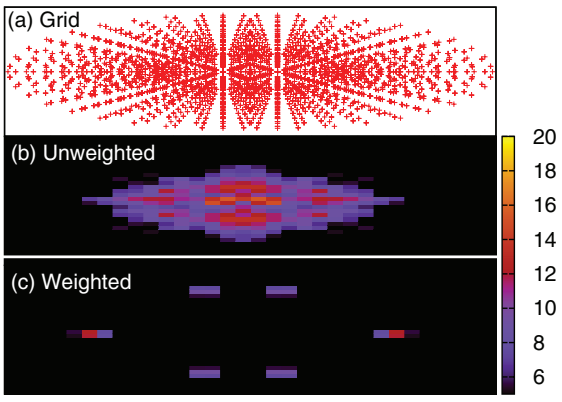

FIG. 1. Plots of the Lebedev grid for H_2 in a plane containing the internuclear axis. (a) Actual grid using $N_p = 302$ points per radial shell and nine shells per atom. (b) Number of grid points contained in each 0.25 Å × 0.25 Å cell, equivalent to the contribution of each cell to the least-squares fit when all the weights w_k are identical. (c) Each cell's contribution when a simple weighting scheme ($w_k = 1/n_k$) is used.

3. Charge derivatives

As compared to the Mulliken or Löwdin charge scheme, the derivatives $\partial Q_\alpha/\partial P_{\mu\nu}$ are significantly more complicated in the case that Q_α is a ChElPG atomic charge. To evaluate these derivatives in the ChElPG case, note that e_B is the only quantity in Eq. (26) that is dependent on $P_{\mu\nu}$. Using the notation of Herbert *et al.*,[15]

$$\frac{\partial Q_A}{\partial P_{\mu\nu}} = -\sum_B^{N_{\rm QM}} (\mathbf{G}^{-1})_{BA} \sum_k^{N_{\rm grid}} \frac{w_k}{|\mathbf{r}_k - \mathbf{r}_B|} (\mathbf{I}_k)_{\mu\nu} + \gamma_A \sum_{BC}^{N_{\rm QM}} (\mathbf{G}^{-1})_{CB} \sum_k^{N_{grid}} \frac{w_k}{|\mathbf{r}_k - \mathbf{r}_C|} (\mathbf{I}_k)_{\mu\nu}, \tag{31}$$

where

$$\gamma_A = \frac{\sum_B^{N_{\rm QM}} (\mathbf{G}^{-1})_{BA}}{\sum_{BC}^{N_{\rm QM}} (\mathbf{G}^{-1})_{BC}}. \tag{32}$$

The quantity $(\mathbf{I}_k)_{\mu\nu}$ in Eq. (31) contains the integrals in this derivative; therefore, it is desirable to rearrange the derivative in such a way so that it needs to be calculated only once for each atom. Defining

$$\lambda_{Ak} = \sum_B^{N_{\rm QM}} \left(\mathbf{G}^{-1}\right)_{BA} \frac{w_k}{|\mathbf{r}_k - \mathbf{r}_B|}, \tag{33}$$

we obtain the following compact result:

$$\frac{\partial Q_A}{\partial P_{\mu\nu}} = \sum_B^{N_{\rm QM}} (\mathbf{\Omega}_B)_{\mu\nu} (\gamma_A - \delta_{BA}). \tag{34}$$

The quantity

$$(\mathbf{\Omega}_B)_{\mu\nu} = \sum_k^{N_{\rm grid}} \lambda_{Bk} (\mathbf{I}_k)_{\mu\nu} \tag{35}$$

consists of charge–density integrals in the AO basis, with "charges" λ_{Bk} located at points $\mathbf{r}_k$. Combining this with Eq. (23) and rearranging the order of summations, one obtains a correction to the Fock matrix in which $(\mathbf{\Omega}_B)_{\mu\nu}$ is evaluated just once for each B, and thus the integrals $(\mathbf{I}_k)_{\mu\nu}$ are calculated exactly once per QM atom.

D. Parameters for the Ewald sums

In addition to the charge scheme, the user-controlled Ewald parameter, η, can greatly influence the calculation time. This parameter controls how much of the pairwise sum is performed in real space, and thus controls how many vectors are required to converge the vector summations in the real-space term [Eq. (3)] and the reciprocal-space term [Eq. (5)]. Both summations converge as Gaussian functions.[2] Following Ref. 2, we thus choose a constant, C, such that $\exp(C^2)$ is within a specified convergence threshold. We take this to be the same as the threshold (drop tolerance) used for the one- and two-electron integrals:

$$C = \sqrt{-\ln(\text{Integral Threshold})}. \tag{36}$$

Unless otherwise stated, the integral threshold will be set to 10^{-8} here.

In real space, the argument of the complimentary error function [see Eq. (3)] controls the convergence, hence we want

$$\exp(C^2) \leq \exp(\eta^2 |\mathbf{r}_{\alpha\beta} + \mathbf{n}_{\rm max} L|^2), \tag{37}$$

where the vector $\mathbf{n}_{\rm max} = (n_{\rm max}, 0, 0)$ specifies how many periodic boxes one must use in the calculation to achieve a required level of accuracy. This is equivalent to figuring out how far away two atoms must be before their pairwise interaction contributes less than the integral threshold. Thus, we obtain

$$C \leq \eta |\mathbf{r}_{\alpha\beta} + \mathbf{n}_{\rm max} L|. \tag{38}$$

Each of the components of $\mathbf{r}_{\alpha\beta}$ must be less than the box length, L. Replacing $\mathbf{n}_{\rm max} + 1$ with $\mathbf{n}_{\rm max}$ for convenience, one obtains

$$C \leq \eta |\mathbf{n}_{\rm max} L| = \eta\, n_{\rm max} L \tag{39}$$

and therefore

$$n_{\rm max} = \text{ceiling}\left(\frac{C}{\eta L}\right). \tag{40}$$

Equation (40) specifies the largest vector that must be included in the real-space sum in order to achieve a certain drop tolerance. If the integers n_x, n_y, n_z are run from $-n_{\rm max}$ to $n_{\rm max}$, however, there are unnecessary vectors that are included in this "supercube." The farthest *distance* that needs to be considered is actually $|\mathbf{n}_{\rm max}|$, so we need include only those lattice vectors satisfying the condition $|\mathbf{n}| \leq |\mathbf{n}_{\rm max}|$. Enforcing the condition creates a "supersphere" where some lattice vectors from the corners of the supercube have been excluded. Note from Eq. (40) that $n_{\rm max} = 0$ when $C/\eta L < 1/2$. This condition leads to a cutoff radius, $R_c = C/\eta < L/2$, so that all significant interactions are included using the minimum-image convention (cutoff at half the box length) in the real-space sum. In this case, the real-space sum is calculated only within the simulation cell, with the result that the $\mathbf{n} \neq \mathbf{0}$ term in Eq. (17) is zero. This is in accordance with the assumption made by Nam *et al.*[8] and is often the case for large simulation cells.

The requisite number of reciprocal-space vectors $\mathbf{m}$ is calculated in the same manner. Since the argument of the exponential function in Eq. (5) is $\pi^2|\mathbf{m}|^2/\eta^2 L^2$, this quantity replaces $(\eta|\mathbf{r}_{\alpha\beta} + \mathbf{n}_{\rm max} L|)^2$ in an inequality similar to Eq. (37), with the result

$$m_{\rm max} = \text{ceiling}\left(\frac{CL\eta}{\pi}\right). \tag{41}$$

Each of the elements in $\mathbf{m}$ runs from $-m_{\rm max}$ to $m_{\rm max}$ with $\mathbf{0}$ excluded and subject to a constraint that $|\mathbf{m}| < |\mathbf{m}_{\rm max}|$.

Now it is possible to determine the exact number of vectors that will be needed for the calculation, given a particular value of η. The number of total vectors in the supercube is

$$v_{\rm tot} = (2n_{\rm max} + 1)^3 + (2m_{\rm max} + 1)^3. \tag{42}$$

The number of total vectors that satisfy the constraints (*i.e.*, the supersphere) is not so easily computed but can be determined through recursion relations. In order to find these

numbers the reader should consult Sloane's handbook of integer sequences (specifically, series A000605).[25,26]

In this work, we employ standard Ewald summation as opposed to the particle-mesh Ewald technique[27,28] that is more common in strictly classical simulations. As such, the cost of both the real- and reciprocal-space sums scales quadratically with the number of vectors. As such, the minimum number of vectors leads to the fastest calculation and the value of η that affords this minimum is the optimal Ewald parameter. This value can be determined by numerical solution of the equation

$$\frac{2CL^3\eta^3}{\sqrt{\pi^3}} + \frac{L^2\eta^2}{\sqrt{\pi}} - L\eta - 2C = 0\,. \tag{43}$$

Alternatively, one may build a table where one finds the number of total vectors in the supercube and then chooses the η value corresponding to the smallest number of vectors.

III. RESULTS

The Lebedev ChElPG method and the QM/MM-Ewald method described above have been implemented in a locally modified version of Q-CHEM v. 4.0.[29,30] Here, we describe various numerical tests designed to evaluate the numerical performance of the method.

A. Charge schemes

In attempting to implement the algorithm in Ref. 8, we encountered serious SCF convergence problems that we suspected were due to the use of extended basis sets in conjunction with Mulliken image charges. Since no such difficulties have been reported in previous minimal-basis implementations of the algorithm,[8,11] we first wanted to verify that the Mulliken version [with charge derivatives given in Eq. (25)] does indeed work in a minimal basis set. To test this, calculations were run at the QM = Hartree-Fock (HF)/STO-3G level of theory, for a single QM water molecule in a box of 215 TIP3P water molecules,[31] with $L = 18.643$ Å corresponding to ambient liquid density. This calculation converged rapidly using Mulliken charges, and in comparison to the corresponding calculation using a Cartesian ChElPG grid (head spacing of 5 Å and $\Delta x = 1$ Å), essentially the same energy is obtained. Moreover, if we compute Mulliken charges using the density matrix obtained from the ChElPG Ewald calculation, we obtain values within 0.003 a.u. of the Mulliken charges obtained from the Mulliken Ewald calculation. This confirms that various charge schemes work equally well in minimal basis sets.

In order to test extended basis sets, the same calculation was preformed using the 6-31(x+,y+)G* and 6-311(x+,y+)G* basis sets, where x and y range from 0 to 3, except that the 6-31(3+,+)G*, 6-31(3+,2+)G*, and 6-311(3+,+)G* basis sets were excluded because in these cases the Mulliken-based Ewald procedure fails to converge after 50 SCF cycles. Figure 2 shows the final, converged point charge on the oxygen atom of the QM water molecule as a function of basis set size. We converge the SCF Ewald calculation using either Mulliken image charges [Fig. 2(a)] or else ChElPG image charges [Fig. 2(b)], and in the latter case we also compute Mulliken charges using the final, converged SCF density matrix. From Fig. 2(a) we see that the use of Mulliken image charges – when the calculation can be converged – often leads to a positive partial charge on the oxygen atom in larger basis sets. Not only is this behavior not observed with ChElPG charges, but if we use ChElPG image charges to converge the SCF calculation (*i.e.*, the ChElPG charges are used to construct the Fock matrix correction $\Delta F^{\rm PI}_{\mu\nu}$), then the Mulliken charges obtained upon convergence are reasonable [see Fig. 2(b)]. This suggests that the problem lies with instabilities in the Mulliken charge derivatives as the basis set is expanded, which are exacerbated when these charges are included as part of the self-consistent iteration procedure.

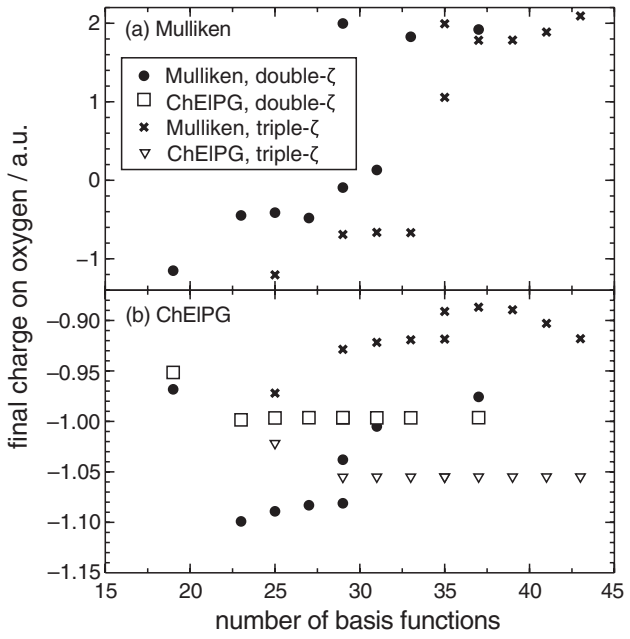

FIG. 2. Converged partial charges for the oxygen atom of a single QM water molecule in an MM water box, plotted against the number of basis functions used to describe the QM region, for 6-31(x+,y+)G* and 6-311(x+,y+)G* basis sets. In (a), the QM/MM-Ewald method uses Mulliken image charges whereas in (b) it uses ChElPG image charges. In the latter case, Mulliken charges were also computed upon SCF convergence.

These instabilities are borne out by the SCF energies, plotted as a function of basis size in Fig. 3. When Mulliken image charges are employed, the correct SCF energy of ≈ -76 hartree is obtained only in small basis sets; in larger basis sets, the "converged" SCF energy differs from this value by as much as 5 hartree. For ChElPG image charges, the SCF energy is stable with respect to basis-set expansion.

Clearly, Mulliken charges cannot be used for QM/MM-Ewald calculations in non-minimal basis sets. The remainder of this work explores the use of ChElPG image charges. In that case, one must determine electrostatic grid parameters to ensure that the charges are converged. Tests of how the ChElPG charges converge with respect to grid parameters are presented in Sec. III B.

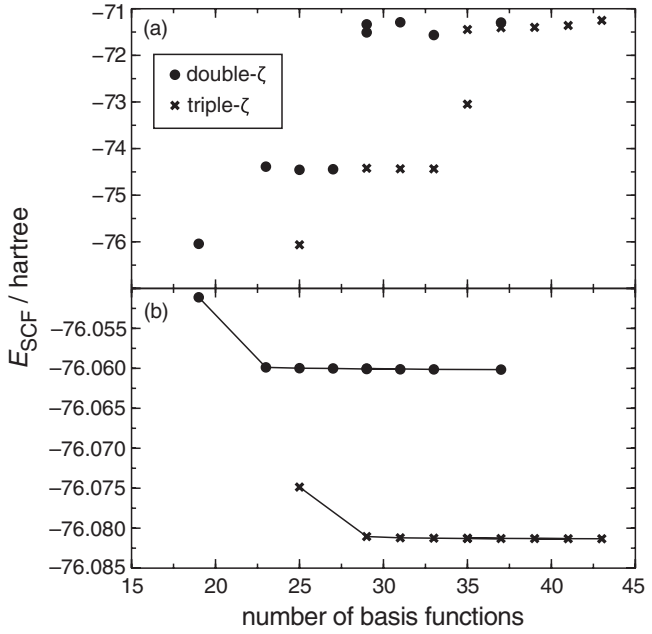

FIG. 3. Total SCF energy versus the number of basis functions, for a single QM water molecule in a box of MM water molecules, using either (a) Mulliken image charges or (b) ChElPG image charges.

B. Lebedev ChElPG charges

1. Gas phase

Our Lebedev grid-based implementation of the ChElPG algorithm is new, and here we seek to test it against the original Cartesian grid-based version of Breneman and Wiberg.[16] The Lebedev version is inherently much more efficient, as it uses far fewer grid points for the same head space and grid spacing, so we seek to understand how sparse we can make the Lebedev grid without adversely affecting the charges that are obtained. In these tests, we leave the head space set to 2.8 Å (the value recommended by Breneman and Wiberg[16]), and use Bondi radii[32] to define the vdW surface. Atom-centered radial Lebedev shells with $N_p = 590$ points per shell extend from the atomic Bondi radius out to 2.8 Å away from that surface, in radial increments of Δx. (This value of N_p has previously been shown to provide good rotational invariance, in the context of polarizable continuum model calculations where the vdW cavity is discretized with atom-centered Lebedev grids.[33]) Choosing bins of volume $(\Delta x)^3$, this leaves only Δx as a parameter to test convergence of the ChElPG charges.

We first aim to determine whether the Lebedev ChElPG charges provide a reasonable representation of the electrostatic potential. To that end, we first examine the convergence behavior of the Cartesian ChElPG charges, in order to establish a baseline. Note that the Cartesian ChElPG charges provide the best possible representation of the electrostatic potential, in a least-squares sense, in the limit that $\Delta x \to 0$, and we will take Cartesian ChElPG charges computed using $\Delta x = 0.05$ Å to be the "true" ChElPG charges. (This choice is justified by the fact that the charges change by only $\sim 10^{-3}$ a.u. when Δx is increased to 0.10 Å.) Convergence of the Cartesian ChElPG charges towards these "true" values, as a function of Δx, is plotted in Fig. 4 for several small molecules.

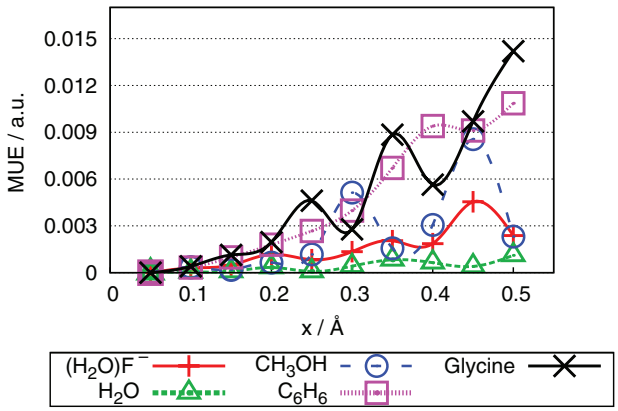

FIG. 4. Convergence of the Cartesian ChElPG charges (atomic units) computed at the HF/aug-cc-pVDZ level, as a function of the Cartesian grid spacing Δx. The reference values were computed using $\Delta x = 0.05$ Å. The vertical axis plots the mean unsigned error (MUE) for all of the charges in the molecule.

Even for $\Delta x = 0.5$ Å, the charges are already converged to about two decimal places.

The convergence of the weighted and unweighted Lebedev ChElPG charges, as a function of Δx, is shown in Fig. 5. Use of the weighting scheme tends to afford better agreement with the Cartesian ChElPG charges, suggesting that an approximately isotropic grid is indeed important for reproducing Cartesian ChElPG charges. Interestingly, the slope of errors with respect to Δx is about the same regardless of whether the weighting scheme is used or not. We take this to mean that the charges converge at about the same rate with respect to Δx, but converge to different values depending on whether the weighting is used. The data in Fig. 5 suggest that it is reasonable to expect errors of the same order of magnitude, or maybe only slightly larger, as those seen for Cartesian ChElPG charges when using Lebedev ChElPG charges.

Given that we can converge the Lebedev ChElPG charges to about the same values as their Cartesian counterparts, we now turn our attention to the rotational invariance of the

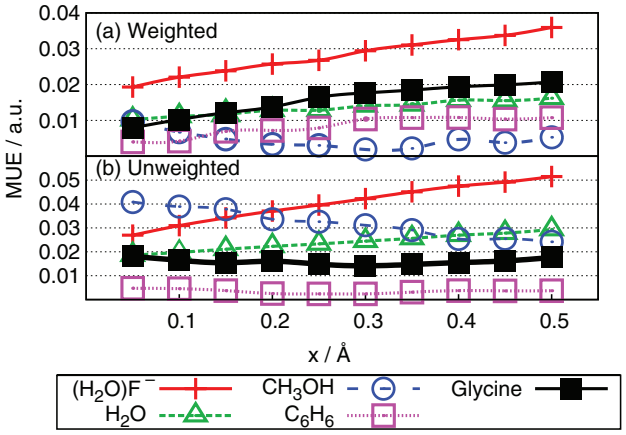

FIG. 5. Convergence of the Lebedev ChElPG charges (atomic units) computed at the HF/aug-cc-pVDZ level, as a function of the grid spacing Δx. The reference values were computed using a Cartesian grid with $\Delta x = 0.05$ Å. The vertical axis plots the mean unsigned error (MUE) for all of the charges in the molecule. In (a), the weighting scheme discussed in Sec. II C 2 is employed ($w_k = 1/n_k$), whereas in (b) the weights are all equal.

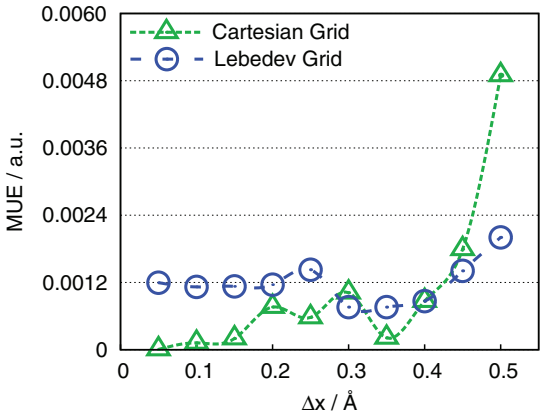

FIG. 6. Rotational invariance (a.u.) of the Lebedev and Cartesian ChElPG charges on the hydroxyl oxygen of glycine as the molecule is rotated by 180° about an axis. Charges are computed at the HF/6-31G* level of theory.

ChElPG charges. To this end, we have computed the HF/6-31G* ChElPG charges of a glycine molecule in the standard nuclear orientation (principal axes of nuclear charge) and also after rotation around the x axis from 0 to π rad, in increments of $\pi/12$ rad. We examine the ChElPG charge on the hydroxyl oxygen atom as a function of this rotation angle, as compared to the value obtained at 0°. The mean unsigned error (MUE), which is a measure of rotational invariance, is plotted in Fig. 6 as a function of Δx.

For comparison, the actual ChElPG charge on the hydroxyl oxygen ranges from −0.78 to −0.80 a.u., whereas the data in Fig. 6 suggest that both the Cartesian and Lebedev grids afford charges that are rotationally invariant to within 0.01 a.u. or better, even for $\Delta x = 0.5$ Å. For grid spacings $\Delta x < 0.15$ Å, no further improvement in the rotational invariance is observed. One interesting point is that the Lebedev grid exhibits better rotational invariance when Δx is large, but (slightly) worse invariance when Δx is small. This is seemingly at odds with Spackman's results for spherical and icosahedral grids.[24] However, Spackman[24] took care to only compare grids of relatively the same density.

In this particular glycine example, the Cartesian grid has sides of ≈21 Å, meaning the Cartesian grids occupy a volume of ≈9300 Å^3 and the Lebedev grid (assuming it is constructed from spheres) has a volume of ≈4800 Å^3. At $\Delta x = 0.5$ Å, the Cartesian grid contains about 15 800 points and the Lebedev grid about 9000 grid points, or point densities of 1.7 Å^{-3} and 1.9 Å^{-3}, respectively. On the other hand, at $\Delta x = 0.05$ Å the Cartesian grid has about 1.5×10^7 points for a point density of about 1600 Å^{-3}, while the Lebedev grid has 84 000 points for a density of 17.5 Å^{-3}. In other words, the Cartesian grid is about 100 times more dense than the Lebedev grid for $\Delta x = 0.05$ Å. We conclude that for Lebedev and Cartesian grids of similar densities (*e.g.*, the $\Delta x = 0.5$ Å case), the Lebedev grids exhibit better rotational invariance. For cases where the point densities are very different (*e.g.*, the $\Delta x = 0.05$ Å case), the more dense grid exhibits the better rotational invariance. This is consistent with Spackman's results.[24]

2. Condensed phase

The results above show that the use of weighted Lebedev grids affords ChElPG charges that are nearly identical to those obtained using Cartesian grids, but can do so with far fewer grid points. However, the benchmarks above use $N_p = 590$ points per radial shell, which will be expensive in calculations with larger QM regions. Reducing this number to $N_p = 50$, we have tested the rotational invariance of the ChElPG charges in the context of a QM/MM-Ewald calculation, taking as a test system a QM region composed of five water molecules (B3LYP/6-31+G* level) in a periodic cell containing 211 MM water molecules ($L = 18.643$ Å). The ChElPG charges were computed using a head space of 3.0 Å with $\Delta x = 0.5$ Å. We carried out single-point energy calculations after rotating the entire simulation cell in increments of 90°, leaving fixed the axis system that defines the ChElPG unit spheres. (That is, the axes of the simulation cell are rotated with respect to the axes that define the grid.)

The results, which are summarized in Tables I and II, use a sparser grid than was used for the gas-phase calculations, yet good rotational invariance of both energies (Table I) and forces (Table II) is observed. The variation in the SCF energy as a function of rotation angle is smaller than the SCF convergence threshold (10^{-5} hartree, for the calculations in Table I). The convergence threshold was tightened to 10^{-7} hartree for the gradient calculations (Table II), yet the variation in different components of the force is no larger than 1.3×10^{-7} a.u. These results suggest that rotational invariance can be achieved in condensed-phase systems using grids that are far sparser than those used in the gas-phase calculations presented above.

C. Timings

One drawback to the use of ChElPG charges is the expense associated with computing the charge derivatives

TABLE I. SCF energies and Lebedev ChElPG charges as the simulation cell (containing 5 QM and 211 MM water molecules) is rotated with respect to the axes that define the Lebedev unit spheres. The final column is the difference between the maximum and minimum values for the various quantities in each row. The SCF convergence threshold was set to 10^{-5} hartree.

	0°	90°	180°	270°	$\lvert\Delta_{\text{max-min}}\rvert$
E_{SCF} (a.u.)	− 382.2489807663	− 382.2489808032	− 382.2489806988	− 382.2489807549	0.000000
ϵ_{HOMO} (a.u.)	− 0.300498	− 0.300498	− 0.300498	− 0.300498	0.000000
Q_{oxy} (a.u.)	− 1.026878	− 1.030694	− 1.035725	− 1.027292	0.008847
Q_{hyd1} (a.u.)	0.465895	0.468642	0.467734	0.466975	0.001839
Q_{hyd2} (a.u.)	0.559032	0.560517	0.564779	0.557386	0.007393

TABLE II. Magnitude of the force on a single oxygen atom as the simulation cell (containing 5 QM and 211 MM water molecules) is rotated with respect to the axes that define the Lebedev unit spheres. The final column is the difference between the maximum and minimum values for the various quantities in each row. The SCF convergence threshold was set to 10^{-7} hartree and the integral threshold to 10^{-10}, with the corresponding value of C determined from Eq. (36).

	Force (a.u.)				
	0°	90°	180°	270°	$\lvert\Delta_{\text{max-min}}\rvert$
$\lvert\mathbf{F}_{\text{total}}\rvert$	0.0318286779	0.0318296463	0.0318290238	0.0318289497	0.0000010
$\lvert\mathbf{F}_{\text{QM/MM}}\rvert$	0.0318488287	0.0318484627	0.0318490519	0.0318486728	0.0000006
$\lvert\mathbf{F}_{\text{Ewald}}\rvert$	0.0000201508	0.0000188163	0.0000200282	0.0000197231	0.0000013

$\partial Q_\alpha/\partial P_{\mu\nu}$, especially the tensor $\mathbf{\Omega}_B$ in Eq. (35). The choice of the Ewald parameter η can also make a large difference in calculation time, as it controls the number of vectors used in the real- and reciprocal-space sums. A poor choice for η can double the calculation time, in our experience.

In order to understand how the Ewald parameter affects the calculation time, two systems were analyzed. The first is intended to be indicative of a fairly small QM region, consisting of 11 QM water molecules (B3LYP/6-31+G* level of theory, for a total of 253 basis functions) in a $L = 18.643$ Å simulation cell containing 205 TIP3P water molecules.[31] The second calculation is much larger, and consists of a QM region containing an aqueous cytidine molecule and all 27 water molecules that reside within 6 Å of the cytidine molecule. The QM region is described at the B3LYP/6-31+G* level (970 basis functions) and placed in a $L = 50.0$ Å simulation cell containing 4122 TIP3P water molecules. In both cases, the ChElPG parameters are set to 3.0 Å for the head space, $N_p = 50$, and $\Delta x = 0.5$ Å, since these values afford good rotational invariance for the test case in Sec. III B 2. We compute the SCF energy of both systems as a function of η. For the first system, all values of η afford the same energy to within 10^{-10} hartree, while for the larger system the variation is no greater than 2×10^{-7} hartree. This implies that we have indeed converged both the real- and the reciprocal-space sums for each value of η, which should be the case if one follows the recommendations in Sec. II D.

Figure 7 shows the CPU time required to compute the $\partial \Delta E^{\text{PI}}/\partial Q_\alpha$ [Eq. (20)], as a function of η, for these two test systems. Note that in a single-point calculation, this is essentially a one-time cost insofar as the main cost is in calculating the Ewald potential, which is done outside of the SCF iterations. Unless the Ewald parameter is chosen poorly, the cost of this step is small in comparison to the time required to compute the ChElPG charge derivatives, $\partial Q_\alpha/\partial P_{\mu\nu}$. This is demonstrated in Fig. 8, where we compare (as a function of η) the fraction of the total job time that is consumed in computing derivatives $\partial \Delta E^{\text{PI}}/\partial Q_\alpha$ versus the fraction required to compute the derivatives $\partial Q_\alpha/\partial P_{\mu\nu}$. Note that all of these calculations are exact (within the integral drop tolerance), insofar as we use the criteria given in Sec. II D to decide how many vectors are necessary to converge the real- and reciprocal-space sums.

As can be seen in Figs. 7 and 8, a poor choice for η can make a large difference in the calculation time. This issue is less important in smaller systems where the time to compute the Ewald potential is small; however, in a large system (such

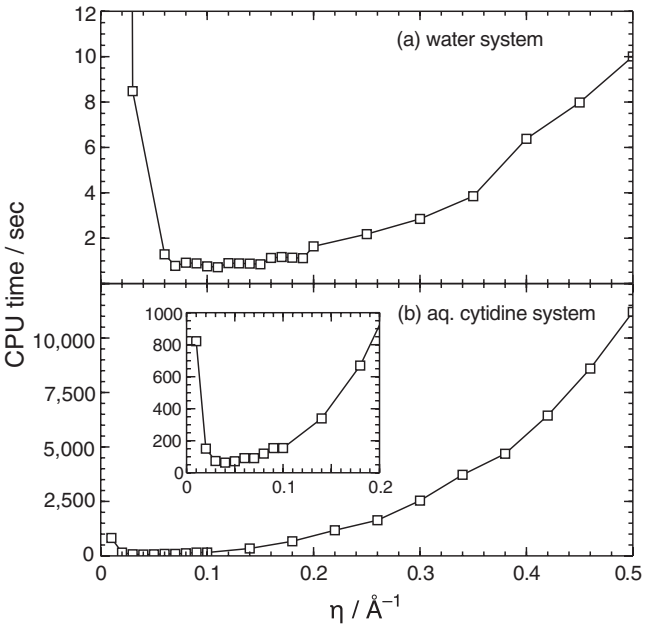

FIG. 7. The total CPU time required to calculate $\partial \Delta E^{\text{PI}}/\partial Q_\alpha$ [Eq. (20)] for (a) 11 QM water molecules in a large MM water box and (b) 27 water molecules plus cytidine in the QM region, surrounded by a larger box of MM water molecules.

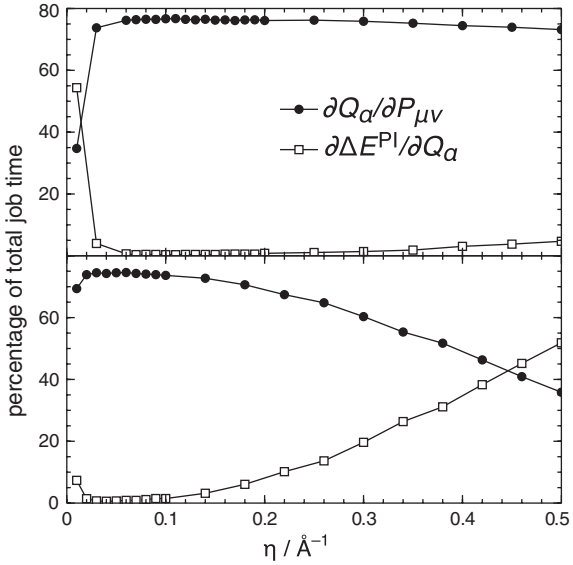

FIG. 8. Percentage of the total single-point energy calculation time that is spent in calculating derivatives $\partial \Delta E^{\text{PI}}/\partial Q_\alpha$ and $\partial Q_\alpha/\partial P_{\mu\nu}$, for a water test system (small QM region, upper panel) and an aqueous cytidine test system (large QM region, lower panel). The time required to calculate $\partial Q_\alpha/\partial P_{\mu\nu}$ is independent of η but becomes a smaller percentage as the time to compute $\partial \Delta E^{\text{PI}}/\partial Q_\alpha$ increases.

TABLE III. Timings (in seconds) for a single-point QM/MM calculation of cytidine (QM) in water (MM), with periodic boundary conditions using $\eta = 0.04$ Å^{-1}. The QM region consists of 30 atoms and 349 basis functions (B3LYP/6-31+G*) and the MM region consists of 12 840 point charges (4160 TIP3P water molecules). The Lebedev and Cartesian ChElPG grids consist of 885 and 32 598 points, respectively. For the steps that must be repeated at each SCF calculation, two columns of timing data are provided, corresponding to the second (left column) and ninth (right column) SCF cycles. The row labeled "other" includes the XC quadrature step and any remaining overhead associated with the Fock build.

	Lebedev		Cartesian	
One-electron integrals	6.4		6.4	
Ewald potential ($\partial\Delta E^{\mathrm{PI}}/\partial Q_\alpha$)	31.5		31.4	
Two-electron integrals	17.6	12.2	17.7	12.4
ChElPG charges	0.4	0.4	3.9	3.8
$\partial Q_\alpha/\partial P_{\mu\nu}$	16.2	16.1	176.4	176.2
Other	6.1	4.8	6.1	4.7
Total Fock build	71.7	33.4	235.4	197.1
Total SCF	348.6		1658.9	

as cytidine in 27 QM water molecules), this step can become the bottleneck if η is chosen too large. This point has not been emphasized previously in the context of Ewald summation for QM/MM calculations.

Most classical implementations of Ewald summation are based on the particle-mesh Ewald method,[27,28] in which the reciprocal-space summation is faster (scaling as $N_{\mathrm{vec}} \log N_{\mathrm{vec}}$ with respect to the number of reciprocal lattice vectors, N_{vec}) than the real-space summation (which scales as N_{MM}^2 with respect to the number of point charges). For this reason, a larger Ewald parameter is generally selected, in order to perform more of the summation in reciprocal space, which may not be an effective strategy for the present implementation, where the cost of the reciprocal-space sum scales as N_{vec}^2. Although a particle-mesh implementation of QM/MM-Ewald may be interesting to consider (especially in view of the recent quantum Ewald mesh for evaluation of electron repulsion integrals[34]), at present the $\partial\Delta E^{\mathrm{PI}}/\partial Q_\alpha$ term is often not the bottleneck of the calculation, as can be seen in Fig. 8. As such, there seems to be little need to accelerate this part of the calculation at present.

It is beneficial to analyze the complete timings of the QM/MM-Ewald calculations, and we will present timings for a variety of aqueous cytidine calculations performed at the B3LYP/6-31+G* level, using ChElPG charges with a head space of 3.0 Å, $N_p = 26$, and $\Delta x = 0.5$ Å. (Although the convergence tests of the ChElPG charges reported in Sec. III B 1 used a much larger number of Lebedev grid points, numerical tests of QM/MM-Ewald calculations, comparing $N_p = 26$ to $N_p = 590$, demonstrate that the converged SCF energies differ by less than the convergence threshold of 10^{-5} hartree.) Table III compares timing data for Lebedev and Cartesian ChElPG grids in the QM/MM-Ewald procedure, for a calculation in which only the cytidine molecule is treated at a QM level. The use of Lebedev grids reduces the number of grid points from 32 598 points to just 885 points, which substantially reduces the cost of computing the charge derivatives $\partial Q_\alpha/\partial P_{\mu\nu}$. At the same time, the difference between the SCF energies in the two calculations is only 6.663×10^{-6} hartree, which is smaller than the SCF convergence threshold of 10^{-5} hartree, so there is every reason to prefer the Lebedev-based approach.

Timing data are provided in Table IV for a sequence of related calculations in which the QM region consists of the cytidine molecule plus all water molecules containing an atom within some specified distance, R, of the glycosidic nitrogen. (All calculations contained 4160 QM + MM water molecules in the simulation cell.) These calculations were performed at the B3LYP/6-31+G* level with $\eta = 0.04$ Å^{-1}. ChElPG grid parameters are the same as those for the cytidine-only QM region discussed above.

The data in Table IV reveal that the time needed to calculate the one-electron integrals is almost negligible. (Note that the one-electron integral timings quoted in the table include only the QM-MM interactions within the simulation cell. The

TABLE IV. Timing data (rounded to the nearest second) for QM/MM calculations of aqueous cytidine in which the QM region consists of a region of specified radius, R, around the cytidine molecule, described at the B3LYP/6-31+G* level. All calculations were performed with periodic boundary conditions using $\eta = 0.04$ Å^{-1} and ChElPG charges. For the steps that must be repeated at each SCF calculation, two columns of timing data are provided, corresponding to the second (left column) and ninth (right column) SCF cycles. Timings labeled "other" includes the XC quadrature step and any remaining overhead associated with the Fock build. All calculations were run in serial on a single Intel Xeon x5650 processor with 48 GB RAM with no competing processes on the node.

	$R = 6$ Å		$R = 7$ Å		$R = 8$ Å		$R = 9$ Å	
Number of QM atoms	108		174		249		345	
Number of MM atoms	12 402		12 336		12 231		12 165	
Number of basis functions	947		1453		2028		2764	
Number of ChElPG grid points	2700		4194		5683		7642	
One-electron integrals	27		44		67		93	
Ewald potential ($\partial\Delta E^{\mathrm{PI}}/\partial Q_\alpha$)	104		166		236		324	
Two-electron integrals	291	122	943	317	2169	735	4546	1288
ChElPG charges	4	4	11	11	24	24	48	47
$\partial Q_\alpha/\partial P_{\mu\nu}$	762	763	3476	3477	10 761	10 767	29 930	29 925
Other	45	29	85	63	137	102	207	153
Total Fock build	1206	917	4681	3868	13 327	11 629	35 054	31 417
Total SCF	8363		29 795		86 921		231 248	

time required to calculate the QM-MM image interactions is included in the $\partial \Delta E^{\rm PI}/\partial Q_\alpha$ term.) Given the data in Table IV, it seems that there is little motivation at this point to work on accelerating the one-electron integral evaluation, *e.g.*, using asymptotic expansions.[35]

The data also reveal that the cost of computing $\partial \Delta E^{\rm PI}/\partial Q_\alpha$ is an order-of-magnitude less than the cost of computing electron repulsion integrals, except for the smallest QM regions. As such, the particle-mesh Ewald technique,[27] which is generally regarded as the method of choice for implementing PBC in classical simulations, does *not* appear to be a promising way forward in the present context, since the most expensive step in our QM/MM-Ewald algorithm (by a very wide margin, especially for large QM regions) is calculation of the charge derivatives $\partial Q_\alpha/\partial P_{\mu\nu}$. In particular, the matrix $\mathbf{\Omega}_B$ in Eq. (35) must be computed $N_{\rm QM}$ times (once for each QM atom, B) in order to calculate the charge derivatives $\partial Q_\alpha/\partial P_{\mu\nu}$. Each of the $\mathbf{\Omega}_B$ matrices is independent of one another so this step can be trivially parallelized across $N_{\rm MM}$ processors, and further parallelism will be as good as the parallelization of the one-electron integrals $(\mathbf{I}_k)_{\mu\nu}$ [Eq. (29)]. Even a factor of two reduction in the time to calculate $\partial Q_\alpha/\partial P_{\mu\nu}$ would reduce the total SCF time for the calculations in Table IV by a minimum of 45%, and by 70% in the case of the smallest ($R = 6$ Å) QM region.

IV. CONCLUSION

Although the theory of Ewald summation for QM/MM calculations has been described before,[8,11] in the context of semi-empirical QM methods, we have provided a robust and general way to extend this technique to extended basis sets, where earlier implementations based on Mulliken image charges for the QM electron density experience stability problems. These are alleviated by using ChElPG image charges instead, and the relatively high cost of computing such charges is mitigated somewhat by an implementation of the ChElPG algorithm using atom-centered Lebedev grids for evaluation of the QM electrostatic potential. These Lebedev ChElPG charges exhibit good rotational invariance and reproduce the QM/MM-Ewald results using traditional ChElPG charges, even for very sparse grids. This is important, because for large QM regions the cost of evaluating derivatives of the ChElPG charges with respect to the density matrix becomes the overwhelming bottleneck in the calculation. (For a more realistic QM region of 349 basis functions, this cost is comparable to the cost of building the Coulomb and exchange matrices for a hybrid density functional.) In future work, we plan to explore multipole-based charge embeddings that sidestep the need for expensive ChElPG charge derivatives. For now, however – and for QM regions more in line with contemporary QM/MM calculations – the Lebedev ChElPG-based QM/MM-Ewald procedure is a promising way to perform periodic QM/MM calculations in a Gaussian-orbital-based SCF electronic structure code. The method works for both HF and DFT calculations, including functionals of arbitrary complexity. Post-HF correlated wave functions can be built upon HF molecular orbitals and eigenvalues that are polarized by the PBC, and the fact that large basis sets can be used means that QM/MM calculations with correlated wave functions are possible.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation (NSF) CAREER award (CHE-0748448). Calculations were performed at the Ohio Supercomputer Center under Project No. PAA0003. J.M.H. is an Alfred P. Sloan Foundation Fellow and a Camille Dreyfus Teacher-Scholar.

[1] A. Y. Toukmaji and J. A. Board, Jr., Comput. Phys. Commun. **95**, 73 (1996).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 2002).
[3] C. Sagui and T. A. Darden, Annu. Rev. Biophys. Biomol. Struct. **28**, 155 (1999).
[4] C. J. Fennell and J. D. Gezelter, J. Chem. Phys. **124**, 234104 (2006).
[5] D. M. York, T. A. Darden, and L. G. Pedersen, J. Chem. Phys. **99**, 8345 (1993).
[6] D. M. York, A. Wlodawer, L. G. Pedersen, and T. A. Darden, Proc. Natl. Acad. Sci. U.S.A. **91**, 8715 (1994).
[7] D. M. York, W. Yang, H. Lee, T. Darden, and L. G. Pedersen, J. Am. Chem. Soc. **117**, 5001 (1995).
[8] K. Nam, J. Gao, and D. M. York, J. Chem. Theory Comput. **1**, 2 (2005).
[9] D. Riccardi, P. Schaefer, and Q. Cui, J. Phys. Chem. B **109**, 17715 (2005).
[10] W. Xie, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Phys. **128**, 234108 (2008).
[11] R. C. Walker, M. F. Crowley, and D. A. Case, J. Comput. Chem. **29**, 1019 (2008).
[12] B. R. Brooks, C. L. Brooks III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, C. Qui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, J. Comput. Chem. **30**, 1545 (2009).
[13] J. Dziedzic, Q. Hill, and C.-K. Skylaris, J. Chem. Phys. **139**, 214103 (2013).
[14] L. D. Jacobson and J. M. Herbert, J. Chem. Phys. **134**, 094118 (2011).
[15] J. M. Herbert, L. D. Jacobson, K. U. Lao, and M. A. Rohrdanz, Phys. Chem. Chem. Phys. **14**, 7679 (2012).
[16] C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).
[17] G. Hummer, L. R. Pratt, and A. E. García, J. Phys. Chem. **100**, 1206 (1996).
[18] S. Sakane, H. S. Ashbaugh, and R. H. Wood, J. Phys. Chem. B **102**, 5673 (1998).
[19] P. H. Hünenberger and J. A. McCammon, Biophys. Chem. **78**, 69 (1999).
[20] S. W. de Leeuw, J. W. Perram, and E. R. Smith, Proc. R. Soc. London, Ser. A **373**, 27 (1980).
[21] S. Boresch and O. Steinhauser, Ber. Bunsenges. Phys. Chem. **101**, 1019 (1997).
[22] T. Vreven and K. Morokuma, Annu. Rep. Comput. Chem. **2**, 35 (2006).
[23] L. E. Chirlian and M. M. Francl, J. Comput. Chem. **8**, 894 (1987).
[24] M. A. Spackman, J. Comput. Chem. **17**, 1 (1996).
[25] N. J. A. Sloane, *A Handbook of Integer Sequences* (Academic Press, 1973).
[26] N. J. A. Sloane, see http://oeis.org/a000605 for code to compute the number of lattice vectors contained in the supersphere.
[27] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[28] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[29] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M.

Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[30] A. I. Krylov and P. M. W. Gill, WIREs Comput. Mol. Sci. **3**, 317 (2013).
[31] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Imprey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[32] A. Bondi, J. Phys. Chem. **68**, 441 (1964).
[33] A. W. Lange and J. M. Herbert, J. Chem. Phys. **133**, 244111 (2010).
[34] C.-M. Chang, Y. Shao, and J. Kong, J. Chem. Phys. **136**, 114112 (2012).
[35] A. Alvarez-Ibarra, A. M. Köster, R. Zhang, and D. R. Salahub, J. Chem. Theory Comput. **8**, 4232 (2012).