EMPIRE: a highly parallel semiempirical molecular orbital program: 2: periodic boundary conditions

Johannes T. Margraf 1 · Matthias Hennemann 1 · Bernd Meyer 1 · Timothy Clark 1

Received: 11 April 2015 / Accepted: 27 April 2015 / Published online: 17 May 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Introduction

The increasing role of quantum chemical calculations in drug and materials design has led to a demand for methods that can describe the electronic structures of large and complex systems. Semiempirical methods based on the neglect of diatomic differential overlap (NDDO) approximation (e.g., the MNDO [1, 2], MNDO/d [3], AM1 [4], AM1* [5], and PMx [6–8] methods) are important representatives of such approaches. Many of these methods have been implemented in the massively parallel program EMPIRE [9], which makes the full quantum-mechanical treatment of systems containing 100,000 atoms or more possible.

Periodic boundary conditions (PBC) enable quantum chemical programs to treat condensed-phase systems, such as proteins in a periodic water box or solids. This allows molecular materials to be studied in their “native” environment, instead of comparing experimental bulk properties with gas-phase monomer calculations. For semiempirical methods, the most practical way of implementing PBC is the cyclic-cluster approach [10–12] in which the system is approximated by a supercell and by imposing Born–von Karman boundary conditions [13]. Using a large unit cell allows the calculation to be performed entirely in real space. This is easily affordable because of the generally low computational cost of NDDO calculations. The main advantage of this technique is that program features like the calculation of local properties [14] or excited states are directly transferable from nonperiodic calculations [15]. We have, for example, used periodic EMPIRE calculations to model amorphous carbon [16].

EMPIRE, which was especially designed for calculations on systems with very many atoms, is also suitable for use on systems with very large unit cells (e.g., disordered and amorphous systems). EMPIRE can, for example, be used in combination with a classical molecular dynamics (MD) code to perform electronic structure calculations on snapshots from an MD run on a periodic system. In the first section of this paper, we discuss the implementation of periodic boundary conditions in EMPIRE. In the second, the program performance is discussed briefly. Finally, some exemplary applications of large-scale periodic NDDO calculations are shown.

Implementation

Periodic calculations in EMPIRE are performed entirely in real space. Therefore, no major changes to the NDDO SCF algorithm were required. Only small adjustments are necessary in the treatment of two-electron two-center integrals: the exchange energy and electrostatic interactions. These adjustments will be discussed below. For more background information, we refer the reader to [10, 12].

Two-electron two-center integrals

The values of the two-electron two-center integrals γ_{AB} used in NDDO calculations (and the associated potential) quickly

Electronic supplementary material

The online version of this article (doi:10.1007/s00894-015-2692-3) contains supplementary material, which is available to authorized users.

Timothy Clark
tim.clark@fau.de

1 Computer-Chemie-Centrum and Interdisciplinary Center for Molecular Materials, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany
decrease with the distance between the centers but remain nonzero. Since these small values add up to unphysical, infinite potentials in a periodic system, they must be corrected [12]. This is achieved by introducing a Gaussian damping function that sets in at a cutoff value \( c_{\text{cut}} \) (the default is 10.0 bohrs). The functional form for these integrals is (in atomic units)

\[
\gamma_{AB} = \frac{1}{r + e^{-0.25(r-c_{\text{cut}})^2}} \quad r > c_{\text{cut}},
\]

where \( G_A \) and \( G_B \) are parameterized constants for elements \( A \) and \( B \), respectively, and \( r \) is the distance between the two centers.

### Exchange interaction energy

The next adjustment is required with respect to the two-electron two-center exchange integrals \( h_{\mu\nu} \), which appear in the Fock matrix. These terms depend on the density matrix elements \( P_{\mu\nu} \). In a periodic calculation, the exchange interactions for an orbital centered on a given atom are only evaluated within the Wigner–Seitz cell surrounding it. The net result is the neglect of very weak exchange interactions with distant electrons, which causes no loss in accuracy [10].

### Electrostatic interactions

MNDO-like NDDO methods describe electrostatic electron–electron and electron–core interactions using multipole–multipole interactions. In a periodic system, small interactions with an infinite number of distant charges lead to unphysical results. This can be alleviated by introducing a simple, one-parameter screening function.

Simply put, distant charges are relocated to an effective distance \( r_{\text{eff}} \), which is a function of the actual distance \( r \). The space around a charge is divided into three regions, delimited by a parameter \( \alpha \): at close distances \( (r<\alpha) \), the actual and effective distances are equal \( (r_{\text{eff}}=r) \). At large distances \( (r>2\alpha) \), all charges are moved to a constant radius of 1.5\( \alpha \). In this manner, their effects cancel each other out due to symmetry [12]. In the intermediate region \( \alpha<r<2\alpha \), the multipoles are scaled by a factor \( \lambda(r) \), with the boundary conditions

\[
\lambda(\alpha) = 1
\]

and

\[
\lambda(2\alpha) = 0.
\]

**Fig. 1** Top: positions of distributed multipole point charges in the exact, scaled, and unscaled scenarios. Bottom: absolute error in the Coulomb energy for the scaled and unscaled cases
The corresponding function is the derivative of (4):\[
\lambda(r) = 2 - \frac{r}{\alpha^*}.
\] (7)

The effect of scaling the multipole size can be evaluated by considering three different scenarios for an interaction between a point-charge and a dipole. The rigorous but unpractical solution is to scale the positions of the constituent point charges of the dipole individually, whereas the simplest approach would be to scale only the center of the dipole. Finally, we can scale the center of the dipole according to (4) and the distances of the distributed point charges to their center by (7).

Figure 1 shows the positions of the distributed monopoles for all three cases ($\alpha = 15.0 \text{ Å}$). Clearly, the scaling function $\lambda$ is a practical way to describe the exact scaling of the multipole distance. This can also be shown by considering the Coulomb energy for the interaction between point charge and dipole. The dependence of the absolute error (i.e., the difference between the unscaled/scaled and exact cases) for this energy on the distance is also shown in Fig. 1. Below the cutoff value, all models are by definition equivalent. At the cutoff value, the scaled and unscaled errors are identical. At increasing distances, however, the error decreases for the scaled case and increases for the unscaled case.

![Fig. 2 Energy convergence with respect to the unit cell volume for diamond, ZnO, NaCl, and adamantane](image)

### Performance

#### Setting up periodic calculations

Periodic EMPIRE calculations require little more input than nonperiodic ones. Apart from the Cartesian coordinates of all atoms, a unit-cell vector is required for each periodic direction. Calculations can be periodic in one, two, or three dimensions. Since no k-space sampling is performed, the unit cell should be sufficiently large. The MOPAC manual suggests that 7–8 Å per repeat unit vector should be sufficient for most compounds, and larger unit cells should be used for highly conjugated $\pi$ systems and small band-gap materials [17]. It is good practice to check the convergence of the calculated results with the size of the unit cell. Figure 2 shows the convergence of the calculated heat of formation with the unit-cell volume for diamond, ZnO, NaCl, and the adamantane molecular crystal. The convergence of the unit-cell size may differ for other properties, as Bredow et al. showed for excitation energies, where significantly larger cells were required than for the ground-state energy [15].

The electrostatic screening parameter can be modified via the keyword ScreeningR, which sets the value of $2\alpha$ in Å. The default is 30.0 Å. This conservative cutoff corresponds to the MOPAC default. Lower cutoffs result in lower computational cost, but whether the heat of formation is affected by the change should be checked. The energy convergences and computational costs of different values of $\alpha$ are shown for diamond and ZnO in the “Electronic supplementary material” (ESM, Figs. S1 and S2).

#### Single-node open MP scaling

Table 1 shows timings for AM1-SCF calculations of differently sized diamond and ZnO unit cells performed with the single-node OMP version of EMPIRE. The corresponding speedup for different numbers of cores is plotted in Fig. 3. The scaling is quite efficient; the speedup factor is >7 using eight cores for all systems investigated. The largest system considered here is the C$_{512}$ unit cell, for which an SCF

| Unit cell | Number of orbitals | Wall-clock time (s) for N cores |
|-----------|--------------------|---------------------------------|
|           |                    | N=1    | N=2    | N=3    | N=4    | N=5    | N=6    | N=7    | N=8    |
| C$_{64}$  | 256                | 26.9   | 13.7   | 9.4    | 7.0    | 5.8    | 4.9    | 4.3    | 3.6    |
| C$_{216}$ | 864                | 111.4  | 56.7   | 38.7   | 30.1   | 24.6   | 20.0   | 17.2   | 15.2   |
| C$_{512}$ | 2048               | 428.3  | 217.7  | 145.7  | 109.6  | 91.1   | 76.4   | 66.8   | 58.6   |
| (ZnO)$_{96}$ | 768               | 69.9   | 35.7   | 24.4   | 18.5   | 14.9   | 12.5   | 11.1   | 9.7    |
| (ZnO)$_{150}$ | 1200              | 162.4  | 83.2   | 56.1   | 44.5   | 34.6   | 29.1   | 25.8   | 22.4   |
| (ZnO)$_{216}$ | 1728             | 224.0  | 114.2  | 77.9   | 59.4   | 48.0   | 41.0   | 36.5   | 31.9   |
calculation takes less than 1 min on eight cores. This shows that on a modern desktop computer, periodic calculations with EMPIRE are absolutely affordable (see Table 2 and Fig. 4).

Multi-node hybrid OMP/MPI scaling

The scaling of the hybrid OMP/MPI multi-node version of EMPIRE was tested on the LiMa cluster at the Regionales Rechenzentrum Erlangen. Here, we used differently sized diamond unit cells from \( C_{1,728} \) to \( C_{13,824} \). Please note that very large unit cells also require large amounts of memory, especially because the integrals are stored, since their calculation is relatively expensive in periodic calculations. Therefore, it is not possible to use the same reference number of nodes when determining the scaling for these systems. The speedup is always relative to the lowest number of nodes feasible for a given system. Optimizing the SCF procedure for periodic calculations may improve the performance of EMPIRE on fewer nodes. As it is, the calculations scale very impressively up to twice the minimum number of nodes. Further increasing the number of nodes leads to a plateau.

### Table 2

| Unit cell | Number of orbitals | Wall-clock time (s) for \( N \) nodes |
|-----------|--------------------|--------------------------------------|
|           | \( N=4 \) | \( N=8 \) | \( N=16 \) | \( N=24 \) | \( N=32 \) |
| \( C_{1728} \) | 6912 | 177.8 | 112.2 | 109.3 | - | - |
| \( C_{8000} \) | 32000 | - | 5149.8 | 2806.1 | 2448.5 | 2327.7 |
| \( C_{13824} \) | 55296 | - | - | 13813.6 | 9376.8 | 7741.4 |

Application

The application of NDDO methods to crystalline materials has been thoroughly tested and evaluated by Stewart, and will therefore not be discussed here in any detail [18]. Instead, we would like to focus on two aspects unique to EMPIRE: firstly, the calculation of local properties; secondly, the fact that even unit cells with thousands of atoms can be treated easily.

Local properties

A local property is any property that can be derived from the wavefunction of a structure and mapped onto a real-space grid, such as the electron density and the molecular electrostatic potential (MEP). These can be calculated with most electronic structure codes. EMPIRE (in combination with an auxiliary program) gives access to several additional local properties derived from molecular orbitals and their energies. These are the local electron affinity (EA\(_L\)), ionization energy (IE\(_L\)), electronegativity, and hardness, which have been used for biochemical QSAR studies and to predict the electron-transport properties of nanostructures [19–27].

Figure 5 shows the molecular electrostatic potentials of a pristine and a defective ZnO(10\(\bar{1}\)0) surface. The nonpolar (10\(\bar{1}\)0) surface consists of rows of ZnO dimers that are separated by trenches [28]. The most abundant atomic defects on this surface are ZnO dimer vacancies [29]. The entire geometry was re-optimized with the MNDO/d Hamiltonian for both the pristine and the defective surface. The removal of one ZnO dimer clearly affects the electrostatic potential around the...
defect, making it necessary to choose a large unit cell to avoid interactions of the defect with its periodic image. This calculation required around 1 min per optimization step and converged in 32 steps. The cell contains 766 atoms and 3,064 electrons. Note also that the MEP does not depend on simplifications such as point-charge models. The multipole formalism used in MNDO-like techniques gives a very good representation of the MEP calculated at higher levels of theory, including anisotropic distributions around heavy atoms [30].

A recent application of periodic local property maps lies in the study of charge transport in organic materials [25–27, 31]. In the condensed phase, the local ionization energy (IE_L) and electron affinity (EA_L) can be interpreted as the local valence-band maximum and conduction-band minimum, respectively. They can therefore be used to visualize the anisotropic electronic properties of a molecular crystal. More recently, the local properties have been used as external potentials to simulate charge transport (see [32]; Bauer T et al., A multi-agent quantum Monte Carlo model for charge transport: application to organic field-effect transistors, submitted). Figure 6 shows the local ionization energy (IE_L) of a rubrene crystal projected onto volume slices that cut through the unit cell along its main axes.

Low IE_L values (shown in blue) correspond to electron-donating/hole-conducting pathways, whereas high IE_L values (shown in red) represent energy barriers. In Fig. 6, the IE_L maps look vastly different depending on the orientation of the volume slice. This is in line with experimental reports, which show that the field-effect mobilities in rubrene single crystals depend strongly on the orientation of the contacts [33, 34].

Large unit cells

As an example of a large system, we chose the solvated lipid bilayer membrane shown in Fig. 7. Specifically, the model consists of 128 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) and 3,840 water molecules equilibrated for 400 ns in a classical molecular dynamics simulation [35]. The unit cell contains 25,088 atoms and spans 62.502×65.506×58.441 Å³. An AM1-SCF calculation was performed on 384 cores of the LiMa cluster (64 MPI tasks on 32 nodes with 2×6 cores each). The SCF converged in 31 cycles and took a little over 3 h 7 min.

Note that periodic calculations of this size push double-precision (64-bit) arithmetic to its limit, since many small values are summed to a very large result during the energy summation. To avoid numerical inaccuracies for large systems, this step is performed in quadruple precision (128-bit), and special care is taken in the ordering of the summands.
The resulting HDF5 binary wavefunction file has a size of 21 GB and can be used to calculate local property maps. The molecular electrostatic potential across the membrane is shown in Fig. 7 (right). This clearly visualizes the polar water layer and head groups and the nonpolar lipid bilayer. Such calculations could, for instance, be used to predict the permeability of membranes to different chemicals.

Every EMPIRE calculation also includes a Coulson population analysis. Figure 8 shows a plot of the Coulson charges of all oxygen atoms as a function of their vertical position. In this plot, five charge groups are discernable, corresponding to the four chemically distinct oxygen atoms in DLPC and the oxygen atom in water. This presents an interesting perspective in the development of force fields for condensed-phase applications, since the charges can be derived directly for the solid or liquid of interest.

Conclusions

We have implemented periodic boundary conditions in the massively parallel semiempirical molecular orbital theory code EMPIRE. The standard SCF procedure of EMPIRE reliably converges the wavefunctions of a broad range of periodic systems, including covalent, ionic, and molecular crystals and surfaces as well as disordered biological systems such as a lipid bilayer. Like the nonperiodic version of EMPIRE, the program is parallelized in the single-node version via open MP, and in the multi-node version via a hybrid open MP/MPI approach.

The single-node version was shown to perform well for calculations on unit cells containing between 64 and 512 atoms, and to scale very efficiently on up to eight cores. The multi-node version allows systems with tens of thousands of atoms to be treated; the largest system described here consisted of 25,088 atoms. The program scaling is similar to that observed for nonperiodic calculations with EMPIRE [9].

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft as part of SFB 953 “Synthetic Carbon Allotropes” and EXC 315 “Engineering of Advanced Materials”, and by the Bavarian Government within the “Solar Technologies go Hybrid” (SolTech) initiative. JTM was supported by a Beilstein Fellowship.

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Dewar MJS, Thiel W (1977) Ground states of molecules, 38. The MNDO method. Approximations and parameters. J Am Chem Soc 99:4899. doi:10.1021/ja00457a004
2. Dewar MJS, Thiel W (1977) Ground states of molecules, 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. J Am Chem Soc 99:4907. doi:10.1021/ja00457a005
