Density functional theory in solution: Implementing an implicit solvent model for CASTEP and ONETEP

Coding Solvation Workshop, Livorno, Italy
August 2017

James C. Womack, a Jacek Dziedzic, a,b Lucian Anton, c Phil J. Hasnip, d Matt I. J. Probert, d Chris-Kriton Skylaris a

aUniversity of Southampton, Southampton, UK
bGdańsk University of Technology, Gdańsk, Poland
cCray U.K. Limited, Bristol, UK
dUniversity of York, York, UK
The cast

\[ \mathcal{O} \quad \text{ONETEP} \]
Linear-scaling density-matrix DFT code

\[ \varphi_\alpha(\mathbf{r}) = \sum_{m \in I_\alpha} D(\mathbf{r} - \mathbf{r}_m)c_{m\alpha} \]

\[ \rho(\mathbf{r}, \mathbf{r'}) = \sum_{\alpha \beta} \varphi_\alpha(\mathbf{r}) K^{\alpha \beta} \varphi^*_\beta(\mathbf{r'}) \]

www.onetep.org

\[ \nabla \cdot (\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -4\pi n(\mathbf{r}) \]

Parallel multigrid solver for 3-D Poisson equation

\[ \hat{A}_h u_h = f_h \]

N(\mathbf{r}) = \sum_{m,k} f_{mk} |\psi_{m,k}(\mathbf{r})|^2

www.castep.org

\[ \psi_{m,k}(\mathbf{r}) = \sum_{G} c_{Gmk} e^{i(\mathbf{G}+\mathbf{k}) \cdot \mathbf{r}} \]


www.castep.org


www.onetep.org


www.onetep.org
DFT in solution: SCRF

Continuum dielectric Self-Consistent Reaction Field

General method for self-consistent solution for the solute charge and reaction potential:

- Reaction potential due to polarization of environment by solute charge
- Solvent environment represented by polarizable dielectric medium

\[
E_{\text{KS}-\text{SCRF}}[n] = T_s[n] + E_{\text{ext}}[n] + E_{xc}[n] + E_{\text{es}}[n] + \Delta G_{\text{non-es}}[n]
\]

Initial guess

\[n(r)\]

Solve KS equations

Determine reaction potential

\[\phi_{\text{NPE}}[n](r)\]

Obtain reaction potential by solving the Non-homogeneous Poisson Equation:

\[
\nabla \cdot (\varepsilon[n](r) \nabla \phi_{\text{NPE}}(r)) = -4\pi n(r)
\]

\[\phi_0(r) + \phi_r(r)\]

\[n_{\text{elec}}(r) + n_{\text{ion}}(r)\]

Well-established methods have been developed, e.g. PCM\(^1\), COSMO\(^2\) and variants\(^3\).

Minimal Parameter Solvation Model

Solute charge
- From density functional theory
- Smeared ionic core charges

\[ n_{\text{elec}}(\mathbf{r}) = \sum_i \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}) \]

Cavity
- Density dependent
- Smoothly varying dielectric function
- Two fitted parameters: \( \beta, n_0 \)

\[ \varepsilon(\mathbf{r}) = 1 + \frac{\varepsilon_{\infty} - 1}{2}\left(1 + \frac{1 - \left(n_{\text{elec}}(\mathbf{r})/n_0\right)^{2\beta}}{1 + \left(n_{\text{elec}}(\mathbf{r})/n_0\right)^{2\beta}}\right) \]

Non-electrostatic part
- From cavity surface area
- Effective surface tension accounts for dispersion-repulsion

\[ \Delta G_{\text{non-es}} = \gamma_{\text{eff}} S[n_{\text{elec}}] \]

\[ \nabla \cdot (\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})) = -4\pi n(\mathbf{r}) \]

Self-consistently solve KS equations subject to solvent
- Electrostatic/non-electrostatic terms depend on density

A refined “Fattebert-Gygi-Scherlis” (FGS) solvation model

MPSM in ONETEP: Implementation highlights

Model originally implemented in ONETEP\textsuperscript{1,2}
- Linear-scaling density-matrix DFT code
- Strictly localized orbitals (NGWFs)
- Direct energy minimization approach
- Capable of DFT calculations on 1000s of atoms

\[
\rho(r, r') = \sum_{\alpha\beta} \varphi_\alpha(r) K^{\alpha\beta} \varphi^*_\beta(r')
\]
\[
E_{\text{min}} = \min_{\{\varphi_\alpha\}} \left( \min_K E[K, \{\varphi_\alpha\}] \right)
\]
\[
n_{\text{elec}}(r) = \rho(r, r)
\]

Solve NPE using a multigrid solver (DL\textsubscript{MG})\textsuperscript{3}
- Efficient real-space solution to second-order
- Highly parallelised: MPI + OpenMP
- Higher-order accuracy obtained by “defect correction”
- Input and output on real-space Cartesian grid

\[
\hat{A} = \nabla \cdot \varepsilon \nabla
\]
\[
\hat{A}^h_2 \text{ is } 2^{\text{nd}} \text{ order FD discretization}
\]
\[
f = -4\pi n_{\text{tot}}(r)
\]

2 J. Dziedzic, H.H. Helal, C.-K. Skylaris, A.A. Mostofi, and M.C. Payne, EPL 95, 43001 (2011)
3 L. Anton, J. Dziedzic, C.-K. Skylaris, and M. I. J. Probert, “Multigrid Solver Module for ONETEP, CASTEP and Other Codes” (dCSE, 2013)
MPSM in ONETEP: Implementation highlights

Smeared ionic cores\textsuperscript{1,2}

- Represent ionic core charge with Gaussians
- Solve NPE for total molecular charge density
- Avoids numerical issues with point-charges
- Single width parameter, $\sigma$

\[
E_{es}[n_{\text{tot}}] = \frac{1}{2} \int \, d\mathbf{r} \, n_{\text{tot}}(\mathbf{r}) \phi_{es}[n_{\text{tot}}](\mathbf{r})
\]

\[
n_{\text{elec}}(\mathbf{r}) + n_{\text{si}}(\mathbf{r})
\]

\[
n_{\text{si}}(\mathbf{r}) = \sum_I n_I(\mathbf{r})
\]

\[
n_I(\mathbf{r}) = -\frac{Z_I}{(\sigma \pi^{1/2})^3} \exp \left( -\frac{|\mathbf{r} - \mathbf{R}_I|^2}{\sigma^2} \right)
\]

\[
\phi_{es}^{\text{BC}}(\mathbf{r}) = \frac{1}{\varepsilon_{\infty}} \int_{\Omega} d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}
\]

\[
\phi_{es}^{\text{BC}}(\mathbf{r}) \approx \frac{1}{\varepsilon_{\infty}} \sum_i n_{\text{tot}}^{\text{CG}}(\mathbf{R}_i) \frac{|\mathbf{r} - \mathbf{R}_i|}{|\mathbf{r} - \mathbf{R}_i|}
\]

Coarse-grained boundary conditions\textsuperscript{2}

- Evaluation of BC integral is computationally costly
- Replace integral with sum over point charges
- Point charges represent summed charge of block of space
- Assumes homogeneous dielectric permittivity for entire cell
- Significant reduces prefactor with negligible error penalty

MPSM in ONETEP: Demonstrating the model

Testing the model in ONETEP for 71 neutral molecules*

<table>
<thead>
<tr>
<th>Approach</th>
<th>XC functional</th>
<th>Error (RMS)</th>
<th>Error (Max)</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPSM⁶</td>
<td>PBE</td>
<td>3.8</td>
<td>8.3</td>
<td>0.83</td>
</tr>
<tr>
<td>MPSM⁷</td>
<td>PBE</td>
<td>4.1</td>
<td>9.1</td>
<td>0.83</td>
</tr>
<tr>
<td>PCM</td>
<td>PBE</td>
<td>10.9</td>
<td>23.3</td>
<td>0.53</td>
</tr>
<tr>
<td>SMD</td>
<td>M05-2X</td>
<td>3.4</td>
<td>14.5</td>
<td>0.87</td>
</tr>
<tr>
<td>AMBER</td>
<td>(classical)</td>
<td>5.1</td>
<td>19.9</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Free energies of solvation in $\text{H}_2\text{O}$, kcal/mol

*Self-consistent cavity, †Fixed cavity

PCM,⁴ SMD⁵ and AMBER⁸ force field Poisson-Boltzmann are competing implicit solvent models.

See Ref 7 for more calculation details.

The MPSM demonstrates comparable accuracy to other methods using only two fitted parameters.

*Taken from blind tests in
Experimental energies of solvation from
3 Minnesota solvation database, version 2009

4 S. Miertuš et al., Chem. Phys. 55, 117 (1981)
7 J. Dziedzic et al., EPL 95, 43001 (2011)
Extending and improving the model

**ARCHER eCSE project**
“Implementation and optimisation of advanced solvent modelling functionality in CASTEP and ONETEP”

- Improvements to multigrid solver (DL_MG)
- Extensions to solvent model in ONETEP
- Implementation of solvent model in CASTEP

Existing implementation of minimal parameter solvent model in ONETEP
J. Dziedzic, H. H. Helal, C.-K. Skylaris, L. Anton, A. A. Mostofi, M. C. Payne

- Implement defect correction in DL_MG
- Periodic boundary conditions (BCs)

**Software design**

- Solvent model
- DL_MG / CASTEP interface
- Supporting functionality

1. A multigrid solver capable of > 2nd-order solution of the NPE
2. Periodic systems can be treated using the solvation model
3. Implicit solvation functionality available in CASTEP
DL_MG and the defect correction

- DL_MG is a 2\textsuperscript{nd}-order multigrid solver
- 2\textsuperscript{nd}-order solutions can produce significant errors in calculated energies\textsuperscript{1}.
- Use “high-order defect correction”\textsuperscript{1,2,3} to iteratively improve the solution
  - Higher order finite differences reduce discretization error

\begin{align*}
\hat{A}_2^h \phi^{(0)} &= f \\
\hat{A}_2^h \phi^{(i+1)} &= \phi^{(i)} + e_{2,d}^{(i)}
\end{align*}

\begin{align*}
r_d^{(i)} &= f - \hat{A}_d^h \phi^{(i)} \\
\hat{A}_2^h e_{2,d}^{(i)} &= r_d^{(i)} \\
\end{align*}

\vspace{0.5cm}

- Procedure ported from ONETEP to DL_MG
- High-order defect correction
- Extension and optimization of ported defect correction
  - Support for periodic BCs
  - 3-D parallel data decomposition
  - Optimization of loops in FD derivative
  - Non-blocking MPI communications for halo exchange

DL_MG and the defect correction

- Error wrt reciprocal space solution
- 448 atom graphene sheet
- 800 eV KE cutoff
- Vacuum with periodic BCs
- Solve Poisson equation for $n_{\text{elec}}$
- $256 \times 264 \times 240$ grid for DL_MG

\[ \nabla^2 \phi_0(\mathbf{r}) = -4\pi n_{\text{elec}}(\mathbf{r}) \]

\[ \tilde{\phi}_0(\mathbf{G}) = 4\pi \frac{\tilde{n}_{\text{elec}}(\mathbf{G})}{|\mathbf{G}|^2} \]

- Time spent in defect correction + solver
- 2615 atom T4 lysozyme-catechol complex
- 826.827 eV KE cutoff
- Vacuum and solvent (H$_2$O) with open BCs
- $513 \times 513 \times 513$ grid for DL_MG
- ARCHER UK national supercomputer
- 6 MPI per node, 4 OpenMP per MPI

- Abs. error in $E_{\text{Hartree}}$ / kcal mol$^{-1}$
- Defect correction FD order

24 Aug 17  Coding Solvation Workshop
Periodic BCs in ONETEP

Electrostatic energy corrected for smeared ionic cores

\[ E_{\text{es}}[n_{\text{tot}}] = \frac{1}{2} \int \mathbf{dr} \ n_{\text{tot}}(\mathbf{r}) \phi_{\text{NPE}}[n_{\text{tot}}](\mathbf{r}) + (E_{\text{locps}}[n_{\text{elec}}] - E_{\text{si-elec}}[n_{\text{elec}}]) + (E_{\text{ion-ion}} - E_{\text{si-si}}) \]

Open (Dirichlet) BCs

\[ \phi(\mathbf{r}) = \frac{1}{\epsilon_{\infty}} \int_{\Omega} \mathbf{dr}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \text{ for } \mathbf{r} \in \partial\Omega \]

- Suitable for isolated solutes
- Used in original implementation of MPSM\textsuperscript{1,2}

\[ E_{\text{ion-ion}}^{\text{OBC}} = \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \]

\[ V_{\text{si}}^{\text{OBC}}(\mathbf{r}) = -\sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \text{erf} \left( \frac{|\mathbf{r} - \mathbf{R}_I|}{\sigma_I} \right) \]

Boundary conditions must be consistently treated in each term of the electrostatic energy and potential

\[ E_{\text{si-si}} = \frac{1}{2} \int \mathbf{dr} \ n_{\text{si}}(\mathbf{r}) V_{\text{si}}(\mathbf{r}) \]

\[ E_{\text{si-elec}}[n_{\text{elec}}] = \int \mathbf{dr} \ n_{\text{elec}}(\mathbf{r}) V_{\text{si}}(\mathbf{r}) \]

Periodic BCs

\[ \phi(0, y, z) = \phi(L_x, y, z) \]

- Suitable for naturally periodic materials
- Only smeared ion terms needed modification
- Locps and Ewald terms already available

\[ E_{\text{ion-ion}}^{\text{PBC}} = E_{\text{Ewald}} \]

\[ \tilde{V}_{\text{si}}^{\text{PBC}}(\mathbf{G}) = 4\pi \frac{\tilde{n}_{\text{si}}(\mathbf{G})}{|\mathbf{G}|^2} \]

1 J. Dziedzic, H.H. Helal, C.-K. Skylaris, A.A. Mostofi, and M.C. Payne, EPL 95, 43001 (2011)
### Periodic BCs in ONETEP

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Charge</th>
<th>Free energy of solvation / kcal/mol</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OBC(^a)</td>
<td>PBC(^a)</td>
<td>PBC(^b)</td>
</tr>
<tr>
<td>Toluene</td>
<td>0</td>
<td>1.666</td>
<td>1.664</td>
<td>1.664</td>
</tr>
<tr>
<td>Benzoic acid</td>
<td>-1</td>
<td>-56.556</td>
<td>-34.231</td>
<td>-34.232</td>
</tr>
<tr>
<td>Aniline</td>
<td>1</td>
<td>-72.386</td>
<td>-49.952</td>
<td>-49.954</td>
</tr>
</tbody>
</table>

\(^a\) Molecule located at the centre of the simulation cell.

\(^b\) Molecule located at the origin of the simulation cell.

---

Free energy of solvation, \(\Delta G_{\text{sol}}\)

"the reversible work required to transfer the solute in a fixed configuration from vacuum to solution"\(^1\)

---

**DL_MG grid sizes:** 145\(^3\) (OBC), 152\(^3\) (PBC)

- Solvation in \(\text{H}_2\text{O}\) in open and periodic BCs
- PBE exchange-correlation
- Norm-conserving pseudopotentials
- 800 eV KE cutoff
- Fixed dielectric cavity

<table>
<thead>
<tr>
<th>(N_{\text{atoms}})</th>
<th>BCs</th>
<th>Free energy of solvation / kcal/mol</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>per atom</td>
<td></td>
</tr>
<tr>
<td>448</td>
<td>Periodic</td>
<td>58.096</td>
<td>0.12968</td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>Periodic</td>
<td>29.073</td>
<td>0.12979</td>
<td></td>
</tr>
<tr>
<td>448</td>
<td>Open*</td>
<td>-27.258</td>
<td>-0.06084</td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>Open*</td>
<td>-32.555</td>
<td>-0.14533</td>
<td></td>
</tr>
</tbody>
</table>

* Open BC calculations were performed using the periodic repeating unit in a padded simulation cell without modification. The structure in OBCs is therefore not physically realistic.

---

Implementing the model in CASTEP

**ONETEP**

Common features

- Solute charge density on real space grid
- Hartree (electrostatic) potential on real space grid
- Pseudopotentials represent nuclei & core electrons
- Written in Fortran 2003 with modular design
- MPI and OpenMP parallelism

Challenges

- No pre-existing support for open BCs in CASTEP
- No pre-existing support for smeared ions in CASTEP
- Different parallel data distributions (1-D vs. 3-D)
- Non-contiguous real-space data in CASTEP

**Grid constraints imposed by DL_MG**

- Grid used by DL_MG must satisfy size constraints
  - Satisfaction of BCs
  - Sufficient number of MG levels
- BUT simulation cell grid size determined by $E_{\text{cutoff}}$
  - To satisfy DL_MG requirements, truncate or pad simulation cell grid

\[ q_i 2^{n_i} + \delta_i \]

\[ \delta_i = \begin{cases} 
0 & \text{for periodic BCs} \\
1 & \text{for open BCs} 
\end{cases} \]

**Multigrid region**

**Simulation cell**

- **ONETEP**
  - Strictly localized orbitals
  - No density can exist in truncated region

- **CASTEP**
  - No strict localization
  - Avoid charge density truncation by padding cell

24 Aug 17  Coding Solvation Workshop  13
Implementing the model in CASTEP

Smeared-ion-corrected energy expressions

\[
E_{\text{tot}}[n_{\text{elec}}] = E_{\text{kin}}[\{\psi_{m,k}\}] + E_{\text{nl}}[\{\psi_{m,k}\}] + E_{\text{xc}}[n_{\text{elec}}] \\
+ \{E_{\text{NPE}}^{\text{tot}}[n_{\text{elec}}] + E_{\text{locps}}^{\text{corr}}[n_{\text{elec}}] + (E_{\text{ion-ion}} - E_{\text{si-si}})\}
\]

\[
E_{\text{tot}}[n_{\text{elec}}] = E_{\text{BS}}[\{\psi_{m,k}\}] + E_{\text{xc}}^{\text{corr}}[n_{\text{elec}}] \\
+ E_{\text{NPE}}^{\text{si}}[n_{\text{elec}}] - E_{\text{NPE}}^{\text{elec}}[n_{\text{elec}}] + (E_{\text{ion-ion}} - E_{\text{si-si}})
\]

- CASTEP includes several energy optimization techniques
- Two energy expressions are used
- Smeared ion correction differs for each expression

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Charge</th>
<th>Free energy of solvation / kcal/mol</th>
<th>CASTEP</th>
<th>ONETEP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ALLBANDS</td>
<td>DM</td>
</tr>
<tr>
<td>Toluene</td>
<td>0</td>
<td></td>
<td>1.695</td>
<td>1.696</td>
</tr>
<tr>
<td>Benzoic acid</td>
<td>-1</td>
<td></td>
<td>-56.401</td>
<td>-56.401</td>
</tr>
<tr>
<td>Aniline</td>
<td>1</td>
<td></td>
<td>-72.666</td>
<td>-72.664</td>
</tr>
</tbody>
</table>

- Solvation in H₂O with open BCs
- PBE exchange-correlation
- Norm-conserving pseudopotentials
- Fixed dielectric cavity
- KE cutoff set to obtain identical fine grid sizes in CASTEP & ONETEP
- 150³ fine grid
  - Padded to 161³ for CASTEP
  - Truncated to 145³ for ONETEP
Conclusions and further work

Key outcomes

Significant new functionality!

- High-order defect correction in DL_MG
  - More accurate solutions of Poisson equation available directly from solver library
  - No need to re-implement correction scheme in calling software package
- Periodic BC solvation in ONETEP
  - Materials with natural periodicity can be studied in implicit solvent using ONETEP
  - Real-space periodic BC solution of Poisson equation available using DL_MG
- Implicit solvent model in CASTEP
  - Implicit solvent calculations possible in CASTEP for the first time
  - Open BC vacuum calculations also possible using components of model

Limitations and next steps

- Periodic BC MPSM needs further testing
  - Periodic BCs treated consistently within model
  - BUT significant discrepancy between charged OBC and PBC $\Delta G_{\text{solv}}$ — is this physical?
- Implement mixed open / periodic BCs
  - Better representation of surfaces and wires
- Improve parallelism
  - Single MPI process calls DL_MG and computes Dirichlet BCs (with threaded parallelism)
  - Requires contiguous representation of real-space quantities in CASTEP
- Self-consistent optimization of $\varepsilon(r)$
  - Only fixed cavity model available currently
Acknowledgements

ARCHER eCSE project collaborators

- Jacek Dziedzic (Southampton & Gdańsk, ONETEP)
- Chris-Kriton Skylaris (Southampton, ONETEP)
- Lucian Anton (Cray Inc., DL_MG)
- Matt Probert (York, CASTEP)
- Phil Hasnip (York, CASTEP)

Previous work on solvent model

- Hatem H. Helal
- Arash A. Mostofi
- Mike C. Payne
- Jacek Dziedzic
- Chris-Kriton Skylaris
- Lucian Anton

Funding

This work was funded under the embedded CSE programme of the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk).

Embedded CSE (eCSE) support provides funding to the ARCHER user community to develop software in a sustainable manner to run on ARCHER.

Paper in preparation

DL_MG: A parallel multigrid Poisson and Poisson-Boltzmann solver for electronic structure calculations in vacuum and solution
J. C. Womack, L. Anton, J. Dziedzic, P. J. Hasnip, M. I. J. Probert, and C.-K. Skylaris
Appendix: CASTEP total energy expressions

\[
E_{\text{tot}}[n_{\text{elec}}] = E_{\text{kin}}[\{\psi_{m,k}\}] + E_{\text{nl}}[\{\psi_{m,k}\}] + E_{\text{xc}}[n_{\text{elec}}] \\
+ \{E_{\text{Hart}}[n_{\text{elec}}] + E_{\text{locps}}[n_{\text{elec}}] + E_{\text{ion-ion}}\}
\]

\[
E_{\text{NPE}}[n_{\text{elec}}] = E_{\text{elec}}^\text{NPE}[n_{\text{elec}}] + E_{\text{si}}^\text{NPE}[n_{\text{elec}}] \\
= \frac{1}{2} \int \text{d}r \left( n_{\text{elec}}(r) + n_{\text{si}}(r) \right) \phi_{\text{NPE}}^\text{tot}(r)
\]

\[
E_{\text{tot}}[n_{\text{elec}}] = E_{\text{BS}}[\{\psi_{m,k}\}] + E_{\text{xc}}^\text{corr}[n_{\text{elec}}] - E_{\text{Hart}}[n_{\text{elec}}] + E_{\text{ion-ion}}
\]

- CASTEP includes several energy optimization techniques
- Two energy expressions are used
- Smeared ion correction differs for each expression

"Band structure energy" includes \(2E_{\text{NPE}}^{\text{elec}}\)