Introduction to ONETEP

Chris-Kритon Skylaris

University of Southampton
Outline

• Density matrix reformulation of DFT
• Localised function representation of density matrix
• Linear-scaling with localised functions
• Linear-scaling with large basis set accuracy
  • NGWFs, density kernel
  • Plane waves and psinc basis set
  • FFT box
  • Parallel scaling
• Compilation and hardware requirements
• Running a simple calculation
• Functionality available
• Examples of ONETEP applications
Computational bottlenecks in DFT

Non-interacting electrons

\[ \left( -\frac{\hbar^2}{2m} \nabla^2 + v_{\text{eff}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \]

Density of interacting electrons

\[ n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \]

Minimise energy w.r.t.:

Molecular Orbitals

\{ \psi_i \}

O(N^3) Computational bottlenecks

- Solving eigenvalue problem / imposing MO orthogonality \( \langle \psi_i | \psi_j \rangle = \delta_{ij} \)
- Building the Hamiltonian matrix
Computational cost of DFT: cubic-scaling

- Not suitable for biomolecules/nanostructures with **thousands of atoms**
- A linear-scaling reformulation of DFT is needed
Linear-scaling DFT

- Physical principle
  - Nearsightedness of electronic matter

- Linear-scaling approaches

- Practical implementation
  - Localised orbitals
  - No diagonalisation
  - Energy minimisation
  - Sparse matrices and algorithms – O(N) memory and CPU cost

In molecules with non-zero band gap, the density matrix decays exponentially

$$\rho(r, r') \sim e^{-\gamma|r-r'|} \to 0 \text{ as } |r - r'| \to \infty$$

Truncate exponential “tail”

$$\rho(r, r') = 0 \text{ when } |r - r'| > r_{cut}$$
Density matrix DFT

DFT energy with molecular orbitals

\[ E_{\text{DFT}} = \sum_i f_i \int \psi_i^*(\mathbf{r}) \left( -\frac{1}{2} \nabla^2 \right) \psi_i(\mathbf{r}) d\mathbf{r} + \int V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{xc}}[n] \]

Density matrix

\[ \rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}') = \sum_{\alpha, \beta} \phi_\alpha(\mathbf{r}) K_{\alpha \beta} \phi_\beta^*(\mathbf{r}') \]

Density

\[ n(\mathbf{r}) = \rho(\mathbf{r}, \mathbf{r}) \]

Molecular orbitals

Localised orbitals

DFT energy with density matrix

\[ E_{\text{DFT}} = -\sum_{\alpha, \beta} K_{\alpha \beta} \int \phi_\beta^*(\mathbf{r}) \frac{1}{2} \nabla^2 \phi_\alpha(\mathbf{r}) d\mathbf{r} + \int V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{xc}}[n] \]
Idempotency condition on the density matrix

Density matrix

\[ \rho(r, r') = \sum_n f_n \psi_n(r) \psi_n^*(r') = \sum_{\alpha, \beta} \phi_\alpha(r) K^{\alpha \beta} \phi_\beta^*(r') \]

- Molecular orbitals
- Localised orbitals

Idempotency condition

\[ KSK = K \]

MO orthonormality

\[ \langle \psi_n | \psi_m \rangle = \delta_{nm} \]

Orbital occupancies

\[ f_n = 1 \text{ or } 0 \]

Orbital occupancies matrix

\[ S_{\alpha \beta} = \langle \phi_\alpha | \phi_\beta \rangle \]
Maintaining idempotency: Purification transformation

\[ f_{\text{new}} = 3f_{\text{old}}^2 - 2f_{\text{old}}^3 \]


- Improves the idempotency of a nearly idempotent density matrix
- Forces occupancies in interval [0, 1] (“weak” idempotency)
- Quadratic convergence. Example:
  \[ \begin{align*}
  f^{(1)} &= 1.1 \\
  f^{(2)} &= 3 \times 1.1^3 - 2 \times 1.1^2 = 0.968 \\
  f^{(3)} &= 3 \times 0.968^3 - 2 \times 0.968^2 = 0.997 
  \end{align*} \]
- But, diverges if the initial occupancies are not within certain bounds
Linear-scaling energy expression which maintains idempotency

\[
L[\sigma] = E \left[ \frac{N_e}{\text{tr}[\rho]} \rho \right] = E \left[ \frac{N_e}{\text{tr}[3\sigma^2 - 2\sigma^3]} (3\sigma^2 - 2\sigma^3) \right]
\]


- **Direct minimisation** of interacting energy – no need for density mixing
- Purification never allowed to break down
  - Occupancy maxima and minima tracked
  - Occupancies restored by penalty functional if out of safe range
- Electron number conserved without need to know the chemical potential
Linear-scaling DFT in practice

Conflicting requirements

- Localised functions
- Sparse matrices

Large basis set accuracy

- Delocalised functions
- Dense matrices

Linear-scaling cost

How can we have linear-scaling cost with “cubic-scaling accuracy”?
The ONETEP linear-scaling approach

- Use a minimal number of $\{\phi_\alpha\}$
- Optimise both $K$ and $\{\phi_\alpha\}$
- Aim is to have linear-scaling DFT with large basis set accuracy

NGWFs confined to spherical regions
Density kernel $K$ sparse by truncation

$$\rho(r, r') = \sum_{\alpha\beta} \phi_\alpha(r) K^{\alpha\beta} \phi_\beta(r')$$

Non-orthogonal Generalised Wannier Functions (NGWFs)

Near-complete basis set accuracy

ONETEP (psinc basis set, K.E. cutoff 800eV)

NWChem (Gaussian basis set, including BSSE correction)

<table>
<thead>
<tr>
<th>Basis set</th>
<th># AOs</th>
<th>BE (kcal/mol)</th>
<th>BE + BSSE (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STO-3G</td>
<td>195</td>
<td>-23.17</td>
<td>-7.98</td>
</tr>
<tr>
<td>3-21G</td>
<td>361</td>
<td>-46.48</td>
<td>-12.55</td>
</tr>
<tr>
<td>6-31G*</td>
<td>535</td>
<td>-27.77</td>
<td>-8.95</td>
</tr>
<tr>
<td>6-311+G*</td>
<td>817</td>
<td>-17.71</td>
<td>-8.79</td>
</tr>
<tr>
<td>6-311++G**</td>
<td>1017</td>
<td>-12.49</td>
<td>-7.39</td>
</tr>
<tr>
<td>cc-pVDZ</td>
<td>685</td>
<td>-33.26</td>
<td>-7.28</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>1765</td>
<td>-19.59</td>
<td>-7.04</td>
</tr>
<tr>
<td>cc-PVQZ</td>
<td>3780</td>
<td>-12.41</td>
<td>-7.22</td>
</tr>
</tbody>
</table>

No B.S.S.E. correction needed with psinc basis


Plane waves: solutions of Schrödinger equation for particle in periodic box

Schrödinger equation

\[-\frac{1}{2} \frac{d^2}{dx^2} \psi_n(x) = \varepsilon_n \psi_n(x)\]

Energy eigenvalues

\[\varepsilon_n = \frac{g_n^2}{2} = \frac{2\pi^2 n^2}{L^2}\]

Eigenfunctions: plane waves

\[\psi_n(x) = \frac{1}{\sqrt{L}} e^{i g_n x} = \frac{1}{\sqrt{L}} \left[ \cos(g_n x) + i \sin(g_n x) \right]\]

\[n = 0, \pm 1, \pm 2, \pm 3, \ldots \quad g_n = \frac{2\pi n}{L}\]
Plane wave basis set

L: length of simulation cell
a: grid spacing
N: number of grid points, L/a

Example in 1 dimension: expanding a molecular orbital $\psi$ in plane waves

$$\psi(x) = \sum_{n=-N/2}^{N/2} \tilde{\psi}(g_n) e^{i g_n x}$$

$$g_n = n \frac{2\pi}{L}$$

$$g_{\text{max}} = g_{N/2} = \frac{\pi}{a}$$
Plane wave basis set

• The quality of the basis set is controlled by a single parameter: the grid spacing $\alpha$

• The basis set is often defined by its kinetic energy cut-off, $E_{\text{max}}$, which is the (particle in a periodic box) energy corresponding to $g_{\text{max}}$

$$E_{\text{max}} = \frac{g_{\text{max}}^2}{2} = \frac{\pi^2}{2a^2}$$

• In 3 dimensions the plane wave basis set is made up by all the $g$-vectors in reciprocal space that are contained within a sphere of radius $g_{\text{max}}$

• Provides a uniform description of space - Basis Set Superposition Error (BSSE) does not occur

• Pulay forces are not needed as plane waves do not depend on atomic positions

But,

• It is practically impossible to use plane waves to describe the wavefunction at the nucleus – so pseudopotentials are needed to remove the Coulomb singularity
Discrete Fourier Transform

• The expansion coefficients of a periodic function $f(x)$ in terms of plane waves is the discrete Fourier transform (FT) of $f(x)$

$$f(x) = \sum_{g_n}^{g_{\text{max}}} \tilde{f}(g_n) e^{i g_n x}$$

Discrete FT of $f(x)$

• We can represent $f(x)$ by its values on a uniform grid in $x$ (the real space)
• The FT of $f(x)$ is on a grid in $g$ (the reciprocal space)
• We can use FTs to compute efficiently various quantities, such as for example the Coulomb potential of the electronic density which is obtained by solving the Poisson equation:

$$\nabla^2 \phi_{el}(r) = -4\pi n(r)$$

\[ n(r) \xrightarrow{\text{FT}} \tilde{n}(g) \quad g \neq 0 \]

\[ \tilde{\phi}_{el}(g) = 4\pi \frac{\tilde{n}(g)}{g^2} \quad \text{back-FT} \]

\[ \phi_{el}(r) \]
Turning plane waves into highly localised functions: Psinc basis set

- Real linear combinations of plane waves
- Highly localised
- Orthogonal

\[ D_j(r) = \frac{1}{N} \sum_{G} e^{iG \cdot (r - r_j)} \]

in 2D:

\[ D_j(r) = D_j(x)D_j(y)D_j(z) \]

NGWF localisation sphere

\[ \phi_{\alpha}(r) = \sum_{i \in \mathbb{L}R_{\alpha}} D_i(r) C_{i,\alpha} \]

\( r_i \) centre of \( D_i(r) \)

Pseudopotentials

- Replace the strong Coulomb nuclear potential $V$ by a much weaker *pseudopotential* $V_{ps}$ which mimics the effect of the core electrons
- Only keep the valence electrons, in smooth valence wave functions $\Psi_{ps}$ without oscillations in the core region
- Eliminating the $1/r$ singularity at the nucleus and the rapid oscillations of the valence wavefunctions near it make it possible to use a plane wave basis set and converge with a reasonable kinetic energy cut-off
- Even hydrogen atoms (obviously, no core electrons) have pseudopotentials in plane wave calculations!

- Beyond the core radius $r_c$, the $V_{ps}$ and $\Psi_{ps}$ match the all-electron calculation
- Determining $r_c$, as well as which electrons to remove as “core” and which to keep as “valence” is an art and requires experience
- Small $r_c$ leads to “hard” pseudopotentials which are more transferable but require more plane waves (higher $E_{max}$) and therefore more computational effort
- There are several recipes for generating pseudopotentials and libraries of these are publicly available, see for example [http://opium.sourceforge.net/](http://opium.sourceforge.net/) and [http://www.sas.upenn.edu/rappegroup/research/pseudo-potential-gga.html](http://www.sas.upenn.edu/rappegroup/research/pseudo-potential-gga.html)
- Before selecting a pseudopotential it is crucial to test it thoroughly, ideally against all-electron calculations
Energy optimisation in ONETEP

\[ E = E[K, \{\phi_\alpha\}] \]

- **Inner loop:** Optimise total (interacting) energy \( E \) w.r.t. \( K \) for fixed \( \{\phi_\alpha\} \) while imposing idempotency and normalisation.

- **Outer loop:**
  Optimise total (interacting) energy \( E \) w.r.t. \( K \) and \( \{\phi_\alpha\} \).

Guess \( K^{\alpha\beta} \) and \( \{\phi_\alpha\} \)

Iteratively improve \( K^{\alpha\beta} \)

Converged?

Yes

Iteratively improve \( \{\phi_\alpha\} \)

No

Converged?

Yes

finished

No

Guess \( K^{\alpha\beta} \) and \( \{\phi_\alpha\} \)

Iteratively improve \( K^{\alpha\beta} \)

Converged?

Yes

Iteratively improve \( \{\phi_\alpha\} \)

No

Converged?

Yes

finished

NGWF optimisation

formaldehyde, H$_2$CO
NGWF optimisation

BaTiO₃

Initial

Ba $\rho$

Ti $d$

Final

O $s$
Psinc basis energy cut-off

Basis set variational approaches:
FFT box technique

simulation cell
FFT box technique

True linear scaling: kinetic energy preconditioning

Linear-scaling: Amyloid fibrils

Structures of the amyloid fibril kindly provided by the authors of
Parallelisation

Graphical Processor Units (GPUs)

“Porting ONETEP to Graphical Processing Unit based Coprocessors. 1. FFT box Operations”

Combining MPI and OpenMP

“Hybrid MPI-OpenMP Parallelism in the ONETEP Linear-Scaling Electronic Structure Code: Application to the Delamination of Cellulose Nanofibrils”

- Tests on "Blue Joule", 1.2 PetaFLOPS, IBM Blue Gene/Q
- 114,688 compute cores
Compiling ONETEP

Simple multi-platform build system, needs:

• Fortran 2003 compiler
• BLAS and LAPACK (or SCALAPACK) numerical libraries
• FFT library: vendor-supplied or FFTw
  – www.fftw.org
• MPI library for parallel version
Running ONETEP

• Parallel computer
  – Minimum 1 GB per processor (core)
  – Typically distribute 10-100 atoms per processor
  – Cross-over >100 atoms

• Prepare input file: free format
  – Documentation at www.onetep.org

• Supply pseudopotential files (.recpot format)
Input file

• Keywords of different types:
  – Integer
  – Boolean
  – String
  – Real
  – Physical (real + unit)
  – Block data e.g. atomic positions, delimited by \%block and \%endblock

• Atomic units by default (hartree and bohr)
Example input file: formaldehyde

! Example input file for the ONETEP program
! Formaldehyde molecule

cutoff_energy 600 eV

%block lattice_cart
  48.00 0.00 0.00
  0.00 48.00 0.00
  0.00 0.00 48.00
%endblock lattice_cart

%block positions_abs
  O  24.887507  23.896975  22.647313
  C  27.731659  23.667449  22.643306
  H  28.655157  21.721170  22.637547
  H  28.955467  25.440371  22.646039
%endblock positions_abs

%block species
  O  O  8  4  6.5
  C  C  6  4  6.5
  H  H  1  1  6.5
%endblock species

%block species_pot
  O  oxygen.recpot
  C  carbon.recpot
  H  hydrogen.recpot
%endblock species_pot
ONETEP calculation outline

- Initialisation phase:
  - Construct initial NGWFs (STOs or PAOs)
  - Construct initial charge density (atomic superposition) and effective potential
  - Construct initial Hamiltonian
  - Obtain initial density kernel using iterative approach
ONETEP calculation outline continued

- Main optimisation phase:
  - Combination of nested self-consistent loops
  - Outer loop optimises the NGWFs (density kernel fixed)
  - Inner loop optimises the density kernel (NGWFs fixed) using Density Matrix Minimisation approaches

Guess $K_{\alpha\beta}$ and $\{\phi_\alpha\}$

Iteratively improve $K_{\alpha\beta}$

Converged?

Yes

Iteratively improve $\{\phi_\alpha\}$

No

Converged?

Yes

No

finished
Example output file: formaldehyde

Linear-Scaling Ab Initio Total Energy Program

Release for academic collaborators of ODG
Version 4.3.1.3

Authors:
Jacek Dziedzic, Peter D. Haynes, Nicholas D. M. Hine,
Arash. A. Mostofi, Mike C. Payne and Chris-Kriton Skylaris

Contributors:
J. Aarons, L. Andrinopoulos, P. W. Avraam, R. A. Bell,
S. J. Clark, G. C. Constantinescu, F. Corsetti, N. Corsini,
O. Dieguez, S. M. M. Dubois, J. M. Escartin, H. H. Helal,
M. J. S. Phipps, C. J. Pickard, M. I. J. Probert,
L. Ratcliff, M. Robinson, A. Ruiz Serrano, J. S. Spencer,
E. W. Tait, G. Teobaldi, D. Turban, V. Vitale,
K. A. Wilkinson and T. J. Zuehlsdorff

Copyright (c) 2004-2015
Example output file: formaldehyde

• Parallel strategy, calculation parameters

Reading parameters from file "h2co.dat" ...... done

Checking processes and threads...

  Default threads: 2
  Running with 2 MPI processes.
  There are 2 MPI processes running on the same node as the root process.
  Each MPI process is using:
    2 threads for simulation cell FFTs.
    2 threads for parallel FFT box operations.
    2 threads for loops over batched FFT box operations.
    2 threads in other parallel regions.

  ... done

Basic input checks...... done

Reading geometry and species blocks from file "h2co.dat" ... 
  <species_atomic_set> block not found: NGWF initialisation set to SOLVE
  ... done

--------------------------------------------------------------------------------
---------------------------------- INPUT FILE ----------------------------------
--------------------------------------------------------------------------------
cutoff_energy 600 eV
threads_max 2
Example output file: formaldehyde

------------------------- Atom counting information -------------------------
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Natoms</th>
<th>Nngwfs</th>
<th>Nprojs</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Totals: 4 10 2

================================ PSINC grid sizes =================================
Simulation cell: 84 x 84 x 84
FFT-box: 75 x 75 x 75
PPD: 6 x 6 x 1
Grid space d1= 0.571428571429a0 (KE cutoff= 23.26377Eh = 633.03940eV)
Grid space d2= 0.571428571429a0 (KE cutoff= 23.26377Eh = 633.03940eV)
Grid space d3= 0.571428571429a0 (KE cutoff= 23.26377Eh = 633.03940eV)

• Simulation cell, FFT box, grid spacing
Example output file: formaldehyde

• NGWF initialisation to pseudo-atomic orbitals (PAOs)

Atom SCF Calculation for O : Z (AE atom) = 8 : Z (PS atom) = 6
Config String:
Orbitals (num,occ): 2 2.00 4.00
Orbitals (num,l): 2 0 1
Atom SCF converged after 33 iterations to a total energy of -15.71589674

Atom SCF Calculation for C : Z (AE atom) = 6 : Z (PS atom) = 4
Config String:
Orbitals (num,occ): 2 2.00 2.00
Orbitals (num,l): 2 0 1
Atom SCF converged after 33 iterations to a total energy of -5.33577996

Atom SCF Calculation for H : Z (AE atom) = 1 : Z (PS atom) = 1
Config String:
Orbitals (num,occ): 1 1.00
Orbitals (num,l): 1 0
Atom SCF converged after 30 iterations to a total energy of -0.43623914
NGWF initialisation ...... done
Example output file: formaldehyde

• Initialise and refine density kernel

Up spin density kernel initialisation ...... done

Writing density kernel to file "h2co.dkn" ... done

>>> Optimising kernel for current NGWFs:

<table>
<thead>
<tr>
<th>iter</th>
<th>energy (Eh)</th>
<th>rms gradient</th>
<th>commutator</th>
<th>dE (Eh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.230601586111E+01</td>
<td>5.3808E-02</td>
<td>4.4840E-02</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.233073648394E+01</td>
<td>3.8281E-02</td>
<td>3.1901E-02</td>
<td>-2.4721E-02</td>
</tr>
<tr>
<td>3</td>
<td>-2.235530086009E+01</td>
<td>1.1889E-02</td>
<td>9.9074E-03</td>
<td>-2.4564E-02</td>
</tr>
<tr>
<td>4</td>
<td>-2.235864046835E+01</td>
<td>6.8691E-03</td>
<td>5.7242E-03</td>
<td>-3.3396E-03</td>
</tr>
<tr>
<td>5</td>
<td>-2.235945547230E+01</td>
<td>3.9727E-03</td>
<td>3.3106E-03</td>
<td>-8.1500E-04</td>
</tr>
</tbody>
</table>

Finished density kernel iterations (  5)

Writing density kernel to file "h2co.dkn" ... done

>>> Density kernel optimised for the current NGWF basis:

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Total energy                =  -2.23594554723032E+01  Eh
RMS occupancy error         =   5.0218E-13
[H,K] commutator            =   3.3106E-03
Occupancy bounds            = [  0.000:  1.000]
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Example output file: formaldehyde

- SCF iterations

>>> Checking for convergence of NGWFs:  NOT CONVERGED

NGWF RMS gradient = 5.2368E-04 > 2.0000E-06 | above tolerance

>>> Improving NGWFs using line search:

Predicted total energy = -2.26058828338741E+01 Eh
Predicted gain in energy = -6.9072241115630E-03 Eh

Writing NGWFs to file "h2co.tightbox_ngwfs"... done

>>> Optimising kernel for current NGWFs:

<table>
<thead>
<tr>
<th>iter</th>
<th>energy (Eh)</th>
<th>rms gradient</th>
<th>commutator</th>
<th>dE (Eh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.260585039399E+01</td>
<td>4.3060E-03</td>
<td>3.5884E-03</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.260614202529E+01</td>
<td>8.8690E-04</td>
<td>7.3908E-04</td>
<td>-2.9163E-04</td>
</tr>
<tr>
<td>3</td>
<td>-2.260616377801E+01</td>
<td>2.2375E-04</td>
<td>1.8646E-04</td>
<td>-2.1753E-05</td>
</tr>
<tr>
<td>4</td>
<td>-2.260616700389E+01</td>
<td>5.2027E-05</td>
<td>4.3355E-05</td>
<td>-3.2259E-06</td>
</tr>
<tr>
<td>5</td>
<td>-2.260616709879E+01</td>
<td>1.7239E-05</td>
<td>1.4366E-05</td>
<td>-9.4900E-08</td>
</tr>
</tbody>
</table>

Finished density kernel iterations (5)

Writing density kernel to file "h2co.dkn" ... done

>>> Density kernel optimised for the current NGWF basis:

-2.26061670987929E+01 Eh

RMS occupancy error = 8.1973E-14
[H,K] commutator = 1.4366E-05
Occupancy bounds = [0.000: 1.000]

Writing density kernel to file "h2co.dkn" ... done
Example output file: formaldehyde

- End of SCF calculation

```plaintext
** NGWF optimisation converged

RMS NGWF gradient = 0.00000126815059

Criteria satisfied:

- RMS NGWF gradient lower than set threshold.

<table>
<thead>
<tr>
<th>Energy Components (Eh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinetic               :  14.91662618649395</td>
</tr>
<tr>
<td>Pseudopotential (local): -75.55456214905318</td>
</tr>
<tr>
<td>Pseudopotential (non-local): 3.08867291023381</td>
</tr>
<tr>
<td>Hartree               :  29.55304104455138</td>
</tr>
<tr>
<td>Exchange-correlation : -5.51703677593763</td>
</tr>
<tr>
<td>Ewald                 :  10.90369328705708</td>
</tr>
<tr>
<td>Total                 : -22.60956549665460</td>
</tr>
</tbody>
</table>

Integrated density : 11.99999999999943

<table>
<thead>
<tr>
<th>CALCULATION SUMMARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITER</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>
```
Summary of Functionality

Total energies

- Various exchange-correlation functionals: LDA (Ceperley-Alder-Perdew-Zunger, Vosko-Wilk-Nusair, PW92), GGA (PW91, PBE, revPBE, RPBE, BLYP, XLYP, WC). Also interface with LIBXC which can provide a large variety of functionals.

- Spin polarisation

- DFT+D (empirical dispersion)

- DFT+U

- Charge-constrained DFT

- Non-local exchange-correlation functionals for dispersion (e.g. Langreth and Lundqvist, VV10)

- Hartree-Fock exchange and hybrid functionals

- Finite temperature DFT for metallic systems
Summary of Functionality

Boundary conditions

• Periodic boundary conditions
• Open boundary conditions (Cut-off Coulomb, Martyna-Tuckerman or real-space open boundaries)
• Implicit solvent model
• Electrostatic embedding

Core electrons

• Norm conserving pseudopotentials
• Projector Augmented wave (PAW) approach (all electron)
Summary of Functionality

Atomic forces
- Geometry optimisation
- Transition state search
- *Ab initio* molecular dynamics

Visualisation
- NGWFs
- Molecular Orbitals
- Density and potentials

Atomic orbital basis set option
- Instead of NGWFs construct and use SZ, SZP, DZ, DZP, etc atomic orbital basis sets
Summary of Functionality

Electronic properties

• Density of states (DOS), local DOS, projected DOS
• Atomic charges
• Dipoles and higher moments
• Optimisation of separate NGWF set for accurate conduction bands and optical absorption spectra
• Natural Bond Orbital (NBO) analysis (Natural Atomic Orbitals in ONETEP and interface to NBO5.9 program)
• Electron transport
• Distributed multipole analysis
Ab initio (Born-Oppenheimer) Molecular Dynamics with ONETEP

- Peptide in water
- NGWFs fully optimised (plane wave accuracy)
- 1522 atoms
- 2256 cores
- Time step of 0.5 fs

Dipole moment partitioning

We have developed two approaches:

1) Distributed multipole analysis

\[ \mathcal{M}_{lm}(I) = \sum_{J}^{N_f/2} \sum_{S_I \neq 0} c_{ll}^{S_I} \int_{\mathcal{I}_a} \mathbf{r} |r|^l \mathcal{C}_{lm}(\mathbf{r}) f_{s_I}(\mathbf{r}) \]

1) Density kernel partitioning (DKP)
Fully anharmonic IR spectra of molecules in solution

Dipole autocorrelation function

\[ I(\omega) \propto \int dt \, e^{i\omega t} \langle \mu(0) \cdot \mu(t) \rangle \]

Example: IR spectrum of peptide in water

- DMA
- DKP

exp. at 300 K

\[ \omega \text{ cm}^{-1} \]

1600 1400 1200
Protein-ligand free energies of binding from QM-PBSA calculations

1) MD of complex in explicit water (MM)

2) Sample snapshots of trajectory

3) Energy of snapshots in implicit solvent for complex, receptor and ligand (QM)

$\Delta G_{\text{bind}}$
T4 Lysozyme L99A/M102Q protein

- 2616 atoms
- Contains a polar binding site in the core of the protein
- Buried from the solvent

Complex of T4 Lysozyme L99A/M102Q and catechol
T4 Lysozyme L99A/M102Q (2602 atoms) in implicit solvent

- First quantum chemistry study of an entire protein in implicit solvent
- Determination of optimum calculation parameters for Energy calculations (vacuum, solvent, binding)
  - Smear core charges
  - Open boundary conditions
  - Higher finite difference order correction
  - NGWF radius
  - Dispersion

Ligands: Binders and non-binders

2-fluoroaniline  3-chlorophenol  Toluene  2-methylphenol

Catechol  Phenol

Non-binders

2-aminophenol  1-phenylsemicarbazide

known decoys
Free energies of binding to T4 Lysozyme L99A/M102Q


<table>
<thead>
<tr>
<th>Ligand</th>
<th>( \Delta G_{\text{bind,vac}}^{\text{QM}} )</th>
<th>( \Delta G_{\text{bind,solv}}^{\text{QM}} )</th>
<th>( \Delta G_{\text{bind,solv}}^{\text{QM}} - T\Delta S )</th>
<th>( \Delta G_{\text{exp}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catechol</td>
<td>(-13.9 \pm 0.38)</td>
<td>(-9.0 \pm 0.42)</td>
<td>(-8.6 \pm 0.67)</td>
<td>(-4.4 ) [22]</td>
</tr>
<tr>
<td>3-Chlorophenol</td>
<td>(-8.1 \pm 0.26)</td>
<td>(-6.9 \pm 0.40)</td>
<td>(-7.7 \pm 0.67)</td>
<td>(-5.8 ) [11]</td>
</tr>
<tr>
<td>2-Fluoroaniline</td>
<td>(-4.3 \pm 0.25)</td>
<td>(-5.9 \pm 0.42)</td>
<td>(-4.8 \pm 0.55)</td>
<td>(-5.5 ) [11]</td>
</tr>
<tr>
<td>2-Methylphenol</td>
<td>(-8.2 \pm 0.25)</td>
<td>(-8.5 \pm 0.37)</td>
<td>(-7.0 \pm 0.61)</td>
<td>(-4.4 ) [11]</td>
</tr>
<tr>
<td>Toluene</td>
<td>(1.6 \pm 0.24)</td>
<td>(-4.8 \pm 0.35)</td>
<td>(-4.4 \pm 0.59)</td>
<td>(-5.2 ) [11]</td>
</tr>
<tr>
<td>1-Phenylsemicarbazide</td>
<td>(-19.8 \pm 0.49)</td>
<td>(-0.3 \pm 0.55)</td>
<td>(3.8 \pm 0.70)</td>
<td>Nonbinder [22]</td>
</tr>
<tr>
<td>2-Aminophenol</td>
<td>(-12.2 \pm 0.41)</td>
<td>(-6.2 \pm 0.39)</td>
<td>(-5.1 \pm 0.51)</td>
<td>Nonbinder [21]</td>
</tr>
<tr>
<td>Phenol (reference)</td>
<td>(-5.6 \pm 0.35)</td>
<td>(-5.6 \pm 0.42)</td>
<td>(-5.6 \pm 0.59)</td>
<td>(-5.6 ) [11]</td>
</tr>
<tr>
<td>Max error\textsuperscript{a}</td>
<td>19.8</td>
<td>6.2</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>RMS error</td>
<td>10.0</td>
<td>3.3</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>Catechol</td>
<td>(-16.0 \pm 0.34)</td>
<td>(-4.1 \pm 0.27)</td>
<td>(-3.7 \pm 0.60)</td>
<td>(-4.4 ) [22]</td>
</tr>
<tr>
<td>3-Chlorophenol</td>
<td>(-9.6 \pm 0.25)</td>
<td>(-8.8 \pm 0.27)</td>
<td>(-9.2 \pm 0.73)</td>
<td>(-5.8 ) [11]</td>
</tr>
<tr>
<td>2-Fluoroaniline</td>
<td>(-5.7 \pm 0.28)</td>
<td>(-8.5 \pm 0.33)</td>
<td>(-7.4 \pm 0.46)</td>
<td>(-5.5 ) [11]</td>
</tr>
<tr>
<td>2-Methylphenol</td>
<td>(-8.2 \pm 0.27)</td>
<td>(-7.2 \pm 0.23)</td>
<td>(-6.1 \pm 0.61)</td>
<td>(-4.4 ) [11]</td>
</tr>
<tr>
<td>Toluene</td>
<td>(0.8 \pm 0.22)</td>
<td>(-7.5 \pm 0.27)</td>
<td>(-7.1 \pm 0.50)</td>
<td>(-5.2 ) [11]</td>
</tr>
<tr>
<td>1-Phenylsemicarbazide</td>
<td>(-22.9 \pm 0.42)</td>
<td>(-11.6 \pm 0.37)</td>
<td>(-7.5 \pm 0.55)</td>
<td>Nonbinder [22]</td>
</tr>
<tr>
<td>2-Aminophenol</td>
<td>(-13.2 \pm 0.40)</td>
<td>(-7.1 \pm 0.39)</td>
<td>(-6.1 \pm 0.50)</td>
<td>Nonbinder [21]</td>
</tr>
<tr>
<td>Phenol (reference)</td>
<td>(-5.6 \pm 0.34)</td>
<td>(-5.6 \pm 0.34)</td>
<td>(-5.6 \pm 0.49)</td>
<td>(-5.6 ) [11]</td>
</tr>
<tr>
<td>Max error\textsuperscript{a}</td>
<td>22.9</td>
<td>11.6</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>RMS error</td>
<td>11.3</td>
<td>5.5</td>
<td>4.0</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Errors calculated relative to the binding free energy of phenol.
More information

User manuals, keywords, tutorials, publications:
www.onetep.org