



Introduction to ONETEP

Chris-Kriton Skylaris

c.skylaris@soton.ac.uk

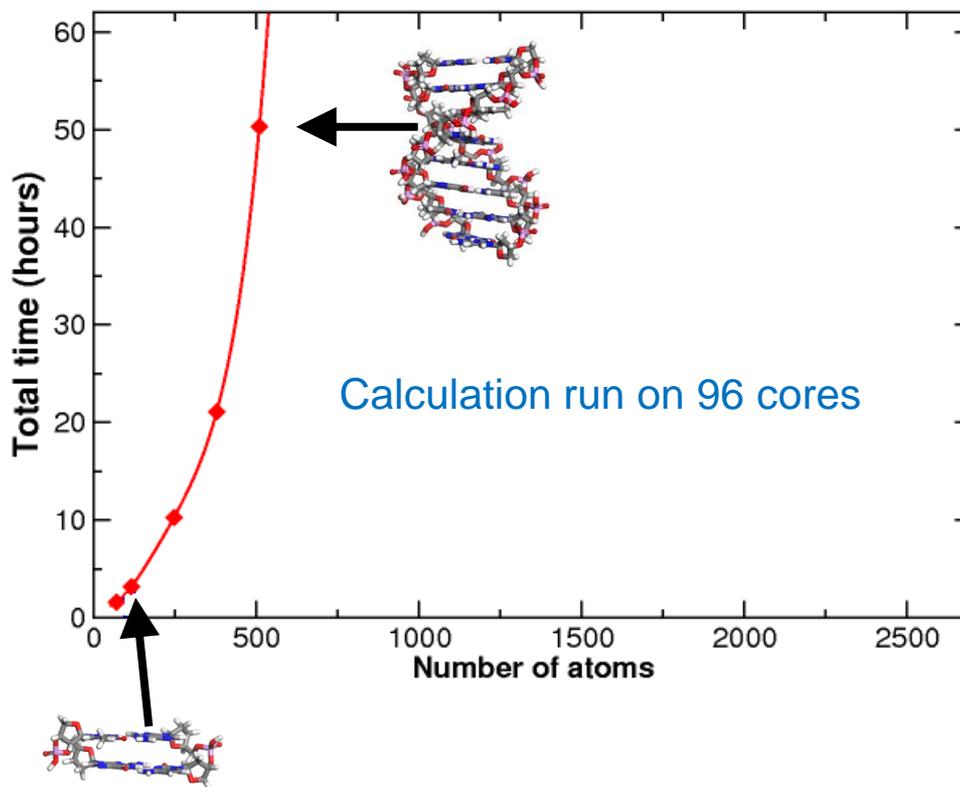
UNIVERSITY OF
Southampton

School of Chemistry

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**
 - **Psinc basis set**
 - **FFT box**
 - **Linear-scaling examples**
 - **Parallel scaling**
- **Compilation and hardware requirements**
- **Running a simple calculation**
- **Functionality available**

Computational cost of DFT: cubic-scaling



- Not a feasible approach for biomolecules with **thousands of atoms**
- A new, **linear-scaling reformulation of DFT**, is needed

Linear-scaling DFT based on the Density Matrix (DM)

- Physical principle

Nearsightedness of electronic matter

W. Kohn, *Phys. Rev. Lett.* **76**, 3168 (1996)

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

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

- Linear-scaling approaches

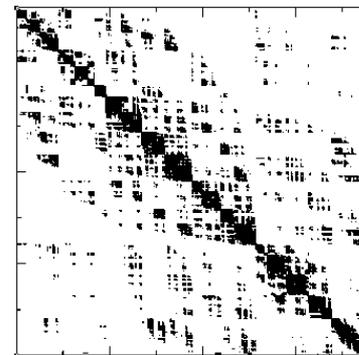
Truncate exponential “tail”

$$\rho(\mathbf{r}, \mathbf{r}') = 0 \quad \text{when} \quad |\mathbf{r} - \mathbf{r}'| > r_{\text{cut}}$$

- Practical implementation

Localised functions, sparse matrix algebra

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



DFT with the one-particle Density Matrix instead of Molecular Orbitals

Density matrix

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i f_i \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}')$$

Density

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

$$n(\mathbf{r}) = \rho(\mathbf{r}, \mathbf{r})$$

Energy expressions

With molecular orbitals

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

With density matrix

$$E[\rho] = \int \left[-\frac{1}{2}\nabla_{\mathbf{r}'}^2 \rho(\mathbf{r}', \mathbf{r}) \right]_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + \int V_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n]$$

One-particle density matrix

Operator representation

$$\hat{\rho} = \sum_n f_n |\psi_n\rangle \langle \psi_n|$$

Position representation

$$\langle \mathbf{r} | \hat{\rho} | \mathbf{r}' \rangle = \rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}')$$

Conditions

- **Idempotency** (from orbital orthonormality and occupancies 1 or 0)

$$\rho^2(\mathbf{r}, \mathbf{r}') = \int \rho(\mathbf{r}, \mathbf{r}'') \rho(\mathbf{r}'', \mathbf{r}') d\mathbf{r}'' = \sum_n f_n^2 \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}') = \rho(\mathbf{r}, \mathbf{r}')$$

- **Normalisation** (preserving the number of electrons)

$$\text{tr}[\rho] = \int \rho(\mathbf{r}, \mathbf{r}) d\mathbf{r} = \int \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}) d\mathbf{r} = \sum_n f_n \int |\psi_n(\mathbf{r})|^2 d\mathbf{r} = \sum_n f_n = N_e$$

Linear-scaling DFT: Density matrix in localised functions

$$\psi_n(\mathbf{r}) = \phi_\alpha(\mathbf{r}) M_n^\alpha$$

Molecular
Orbital (MO)

Atomic
Orbital (AO)

Overlap matrix

$$S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$$

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}') = \phi_\alpha(\mathbf{r}) \left(\sum_n M_n^\alpha f_n M_n^{\dagger\beta} \right) \phi_\beta^*(\mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}')$$

Conditions

• Idempotency

$$\mathbf{KSK} = \mathbf{K}$$

• Normalisation

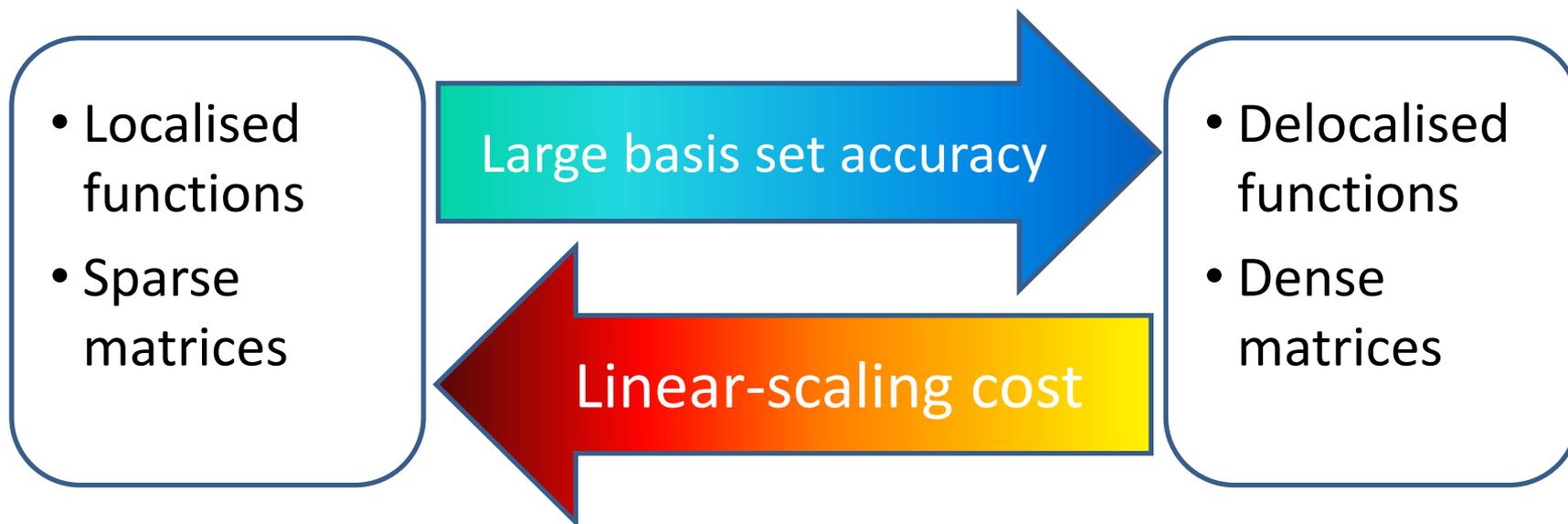
$$\text{tr}[\rho] = \text{tr}[\mathbf{KS}] = N_e$$

$$K^{\alpha\beta} = \sum_n M_n^\alpha f_n M_n^{\dagger\beta}$$

Density kernel

Linear-scaling DFT in practice

Conflicting requirements



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

Linear-scaling cost with “cubic-scaling accuracy”

Two levels of resolution

- Coarse
 - Minimal set of localised non-orthogonal functions
- Fine
 - Large, near-complete basis set
 - *In situ* optimisation of the minimal set of non-orthogonal functions

Optimal basis density matrix minimization (OBDMM) approaches

(S. Goedecker, *Rev. Mod. Phys.*, **71**, 1085 (1999))

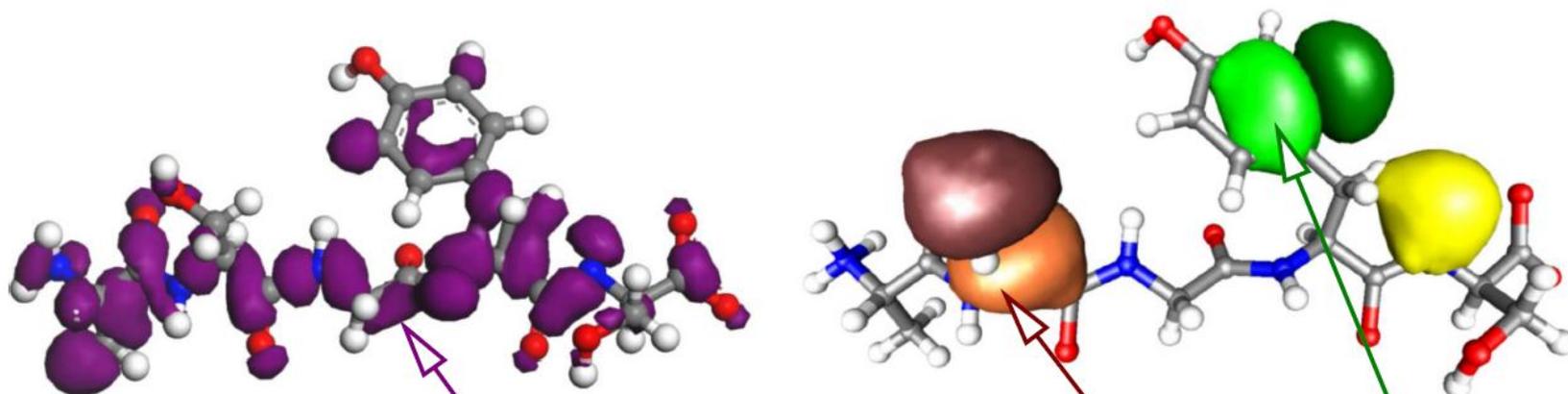
E. Hernandez and M. J. Gillan, *Phys. Rev. B* **51**, 10157 (1995).

J.-L. Fattebert and J. Bernholc, *Phys. Rev. B* **62**, 1713 (2000).

C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Dieguez and M. C. Payne, *Phys. Rev. B* **66**, 035119 (2002).

The ONETEP approach

Skylaris, Haynes, Mostofi & Payne, *J. Chem. Phys.* **122**, 084119 (2005)



$$\rho(\mathbf{r}, \mathbf{r}') = \sum_n f_n \psi_n(\mathbf{r}) \psi_n^*(\mathbf{r}')$$

Molecular orbitals (MOs)

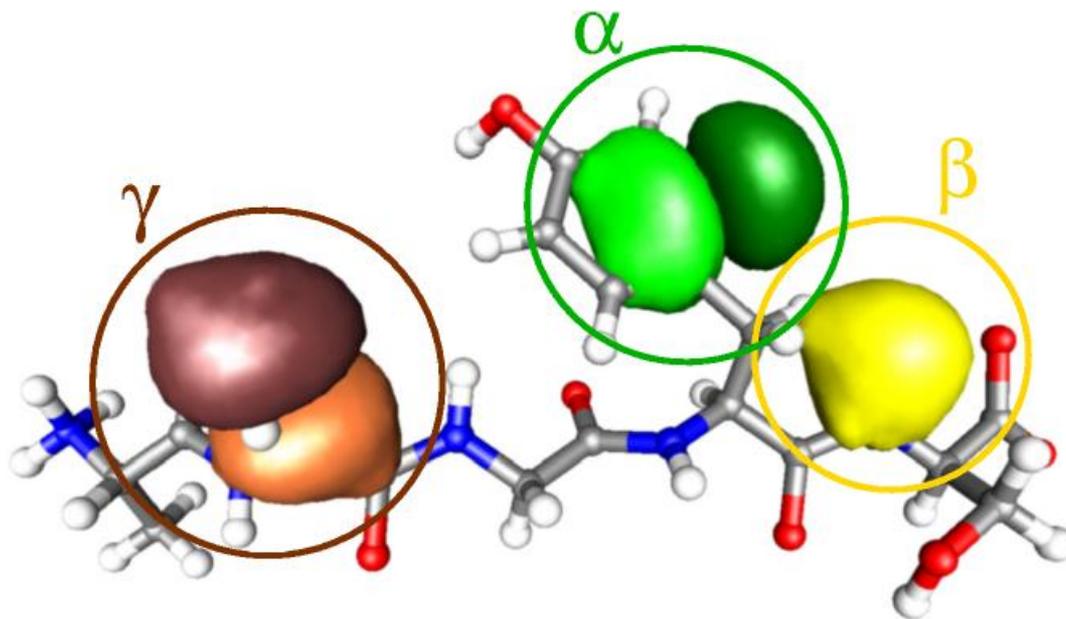
$$= \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}')$$

Non-orthogonal
Generalised Wannier
Functions (NGWFs)

Density
kernel

- Coarse level: The NGWFs are the localised orbitals
- Fine level: NGWFs expanded in a basis set of psinc functions

Density matrix localisation



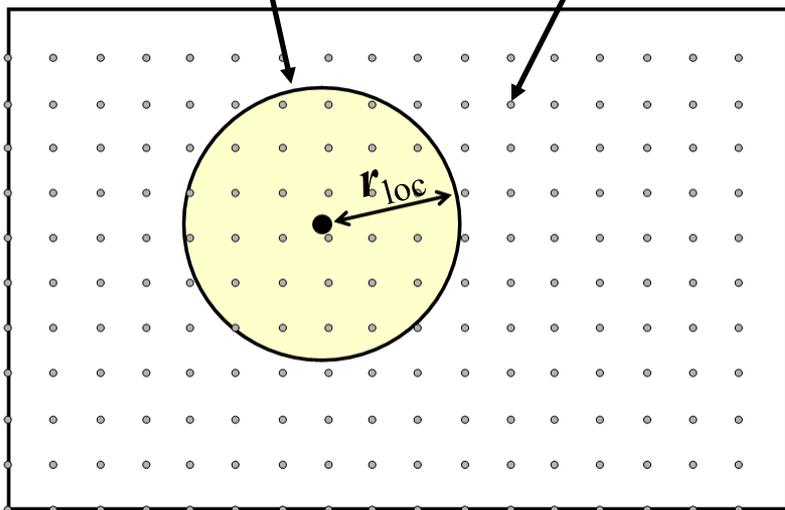
$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_{\alpha}(\mathbf{r}) K^{\alpha\beta} \phi_{\beta}(\mathbf{r}')$$

- Impose spatial cut-offs:
 - NGWFs confined to spherical regions
 - Sparse density kernel \mathbf{K} by truncation

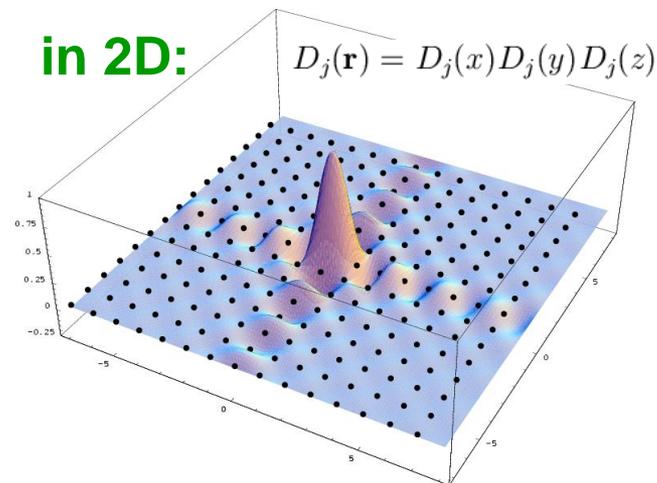
Basis set: Psinc functions

- Real linear combinations of plane waves $D_j(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{G}}^{\mathbf{G}_{\max}} e^{i\mathbf{G}\cdot(\mathbf{r}-\mathbf{r}_j)}$
- Highly localised
- Orthogonal

NGWF localisation
sphere

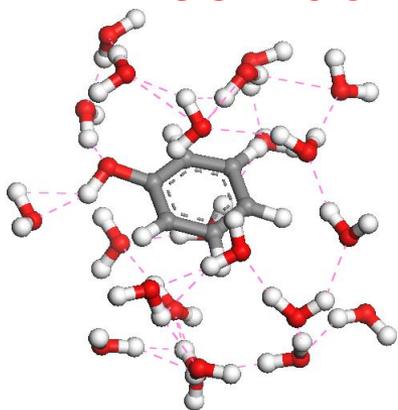


in 2D:



- A. A. Mostofi, P. D. Haynes, C.-K. Skylaris and M. C. Payne, *J. Chem. Phys.* **119**, 8842 (2003)
- D. Baye and P. H. Heenen, *J. Phys. A: Math. Gen.* **19**, 2041 (1986)

Near-complete basis set accuracy



Binding energy
calculation

NWChem (Gaussian basis set, including
BSSE correction)

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

NGWF radii (Å)	# NGWFs	BE (kcal/mol)
2.9	166	-11.93
3.2	166	-12.86
3.7	166	-8.25
4.2	166	-7.06
4.8	166	-7.04

ONETEP is basis set variational

C.-K. Skylaris, O. Dieguez, P. D. Haynes and
M. C. Payne, *Phys. Rev. B* **66**, 073103 (2002).

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

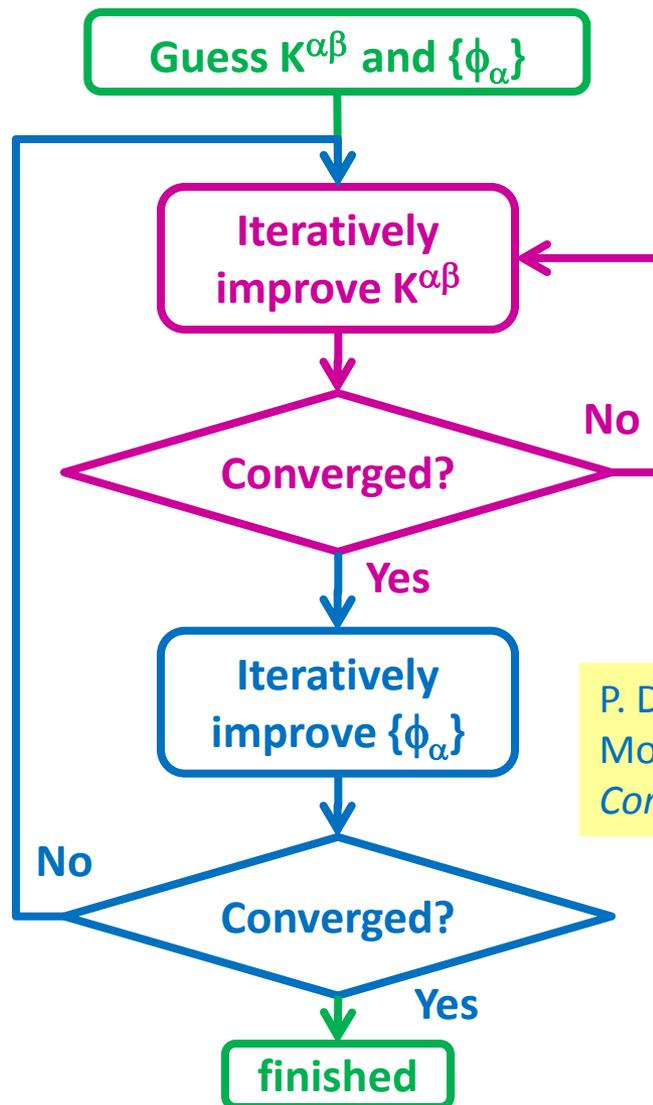
No B.S.S.E. correction needed with psinc basis

P. D. Haynes, C.-K. Skylaris, A. A. Mostofi and M. C. Payne,
Chem. Phys. Lett. **422** 345 (2006).

Energy optimisation in ONETEP

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

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

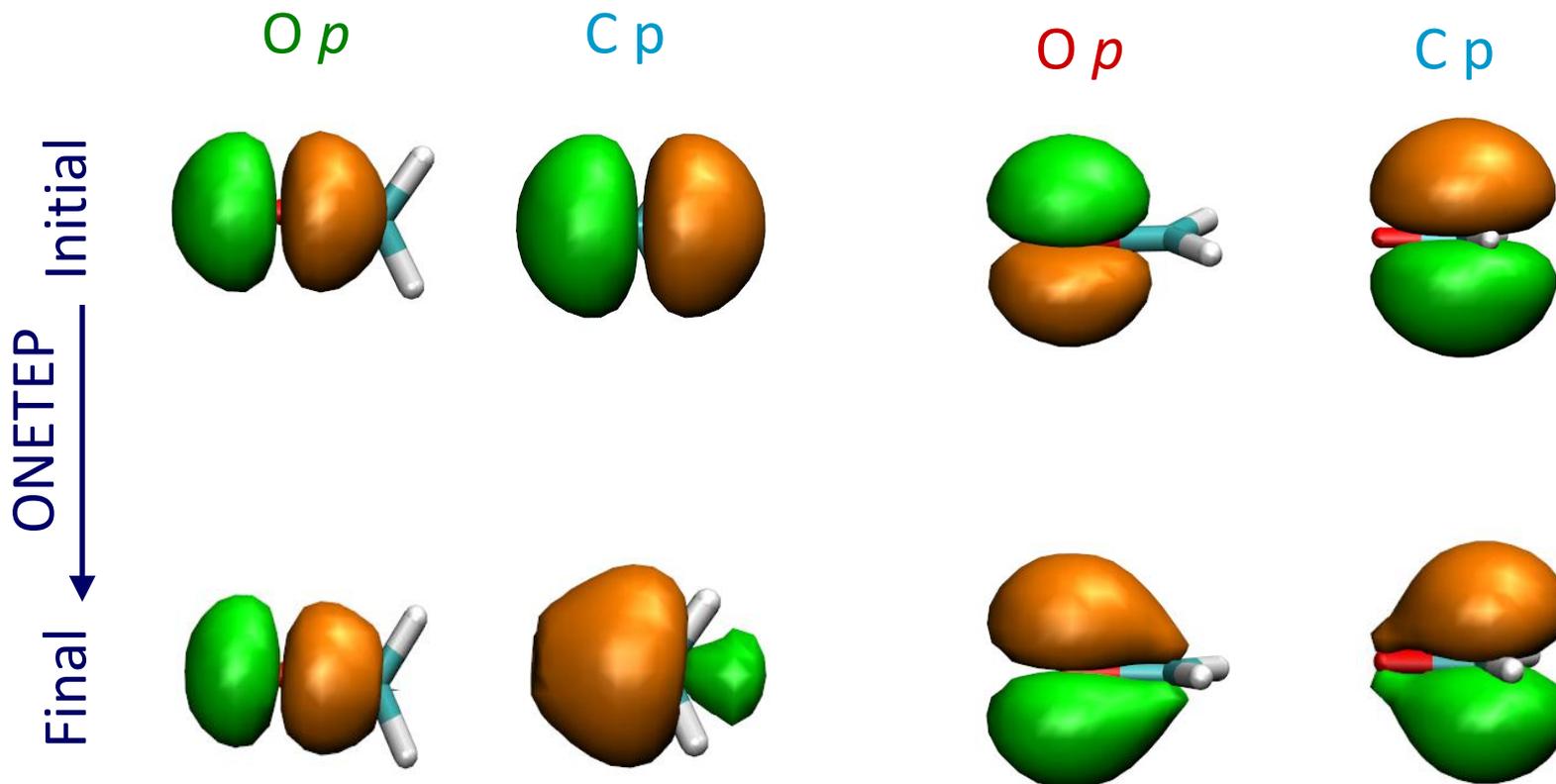


- **Inner loop:** Optimise total (interacting) energy E w.r.t \mathbf{K} for fixed $\{\phi_\alpha\}$ while imposing **idempotency** and **normalisation**

P. D. Haynes, C.-K. Skylaris, A. A. Mostofi and M. C. Payne, *J. Phys. Condens. Matter* **20**, 294207 (2008)

NGWF optimisation

formaldehyde, H_2CO



NGWF optimisation

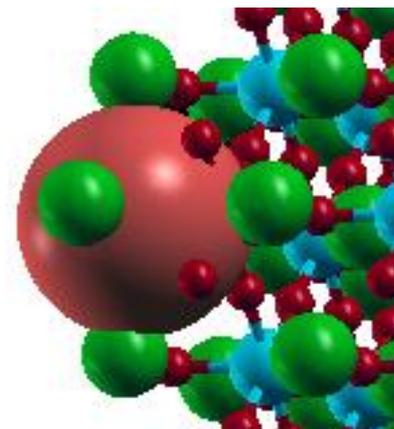
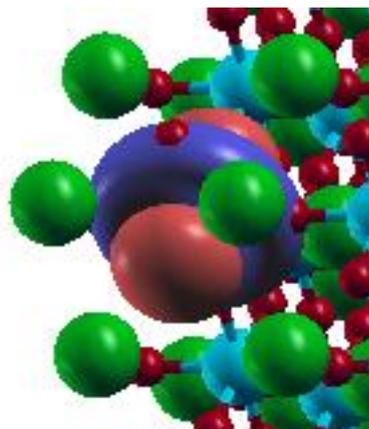
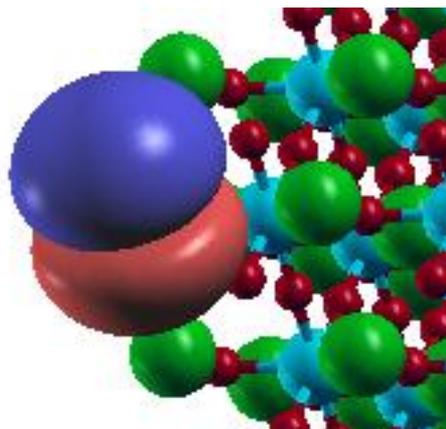
BaTiO₃

Ba *p*

Ti *d*

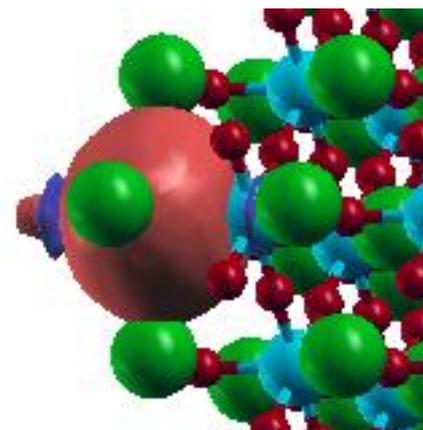
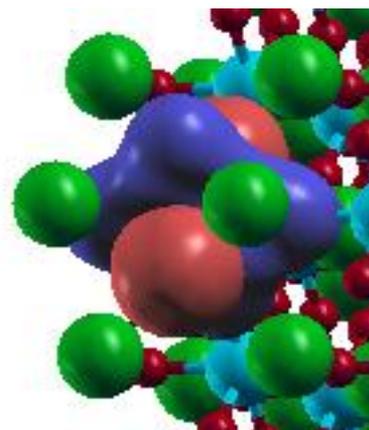
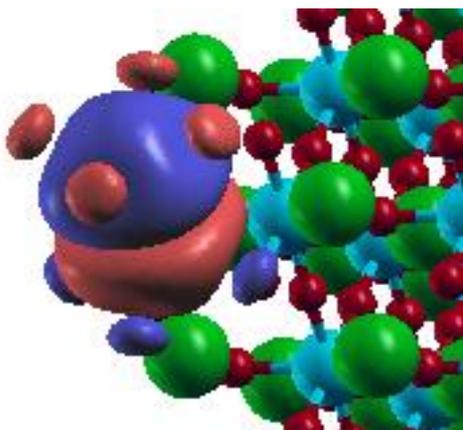
O *s*

Initial

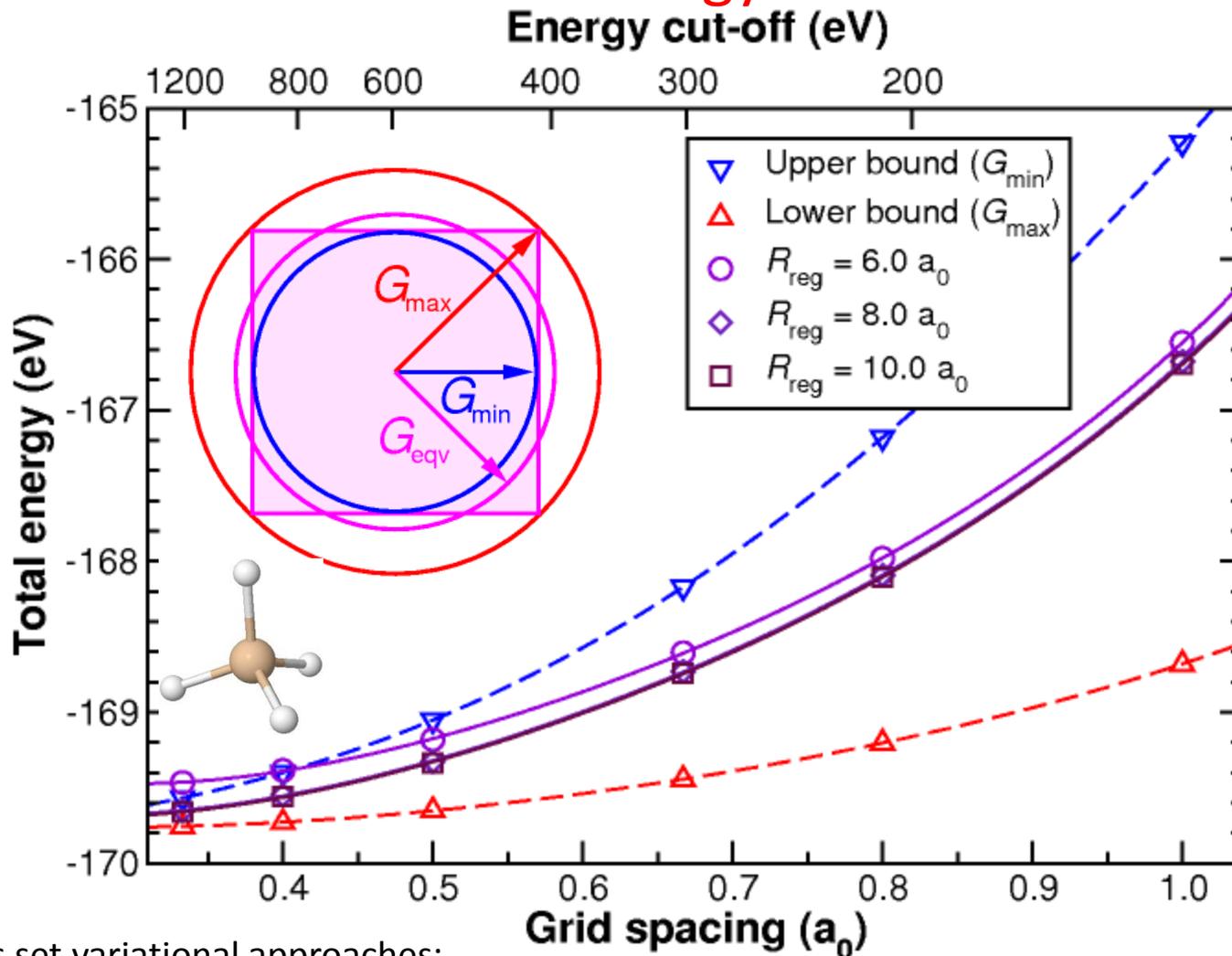


ONETEP

Final



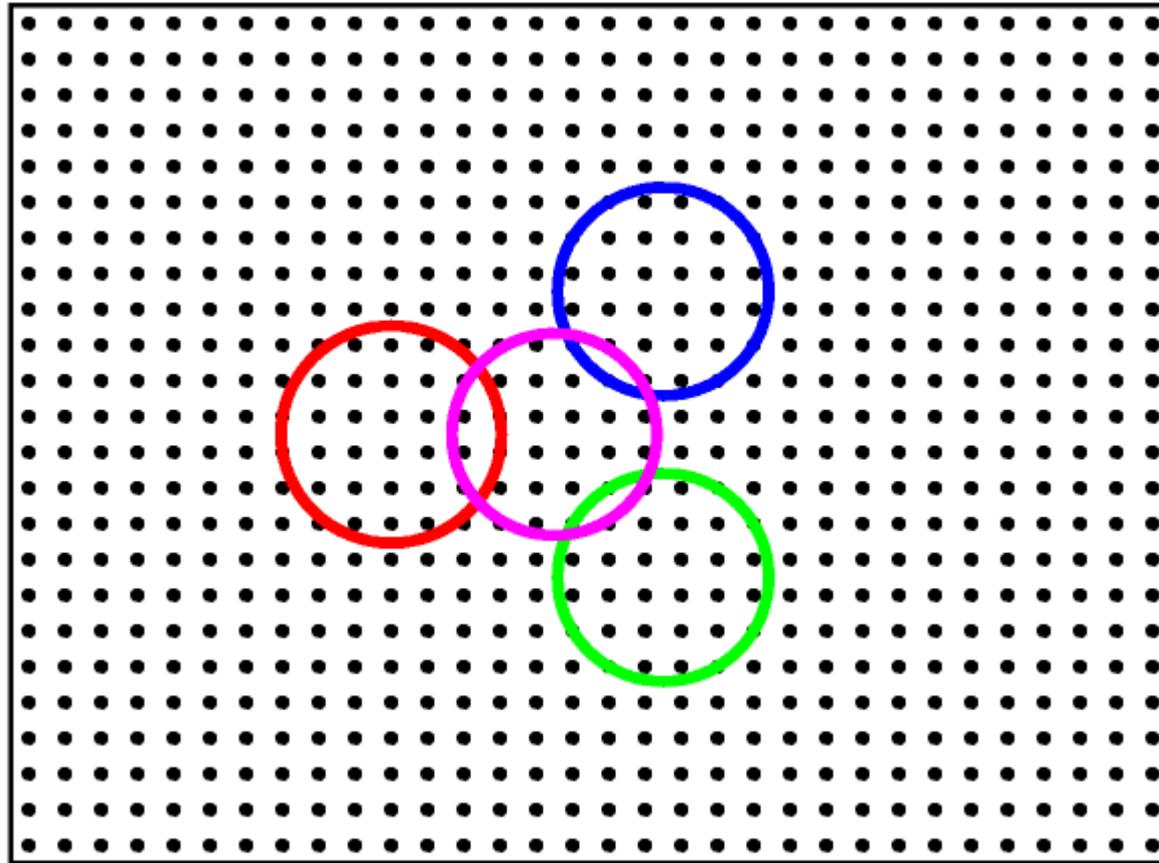
Psinc basis energy cut-off



Basis set variational approaches:

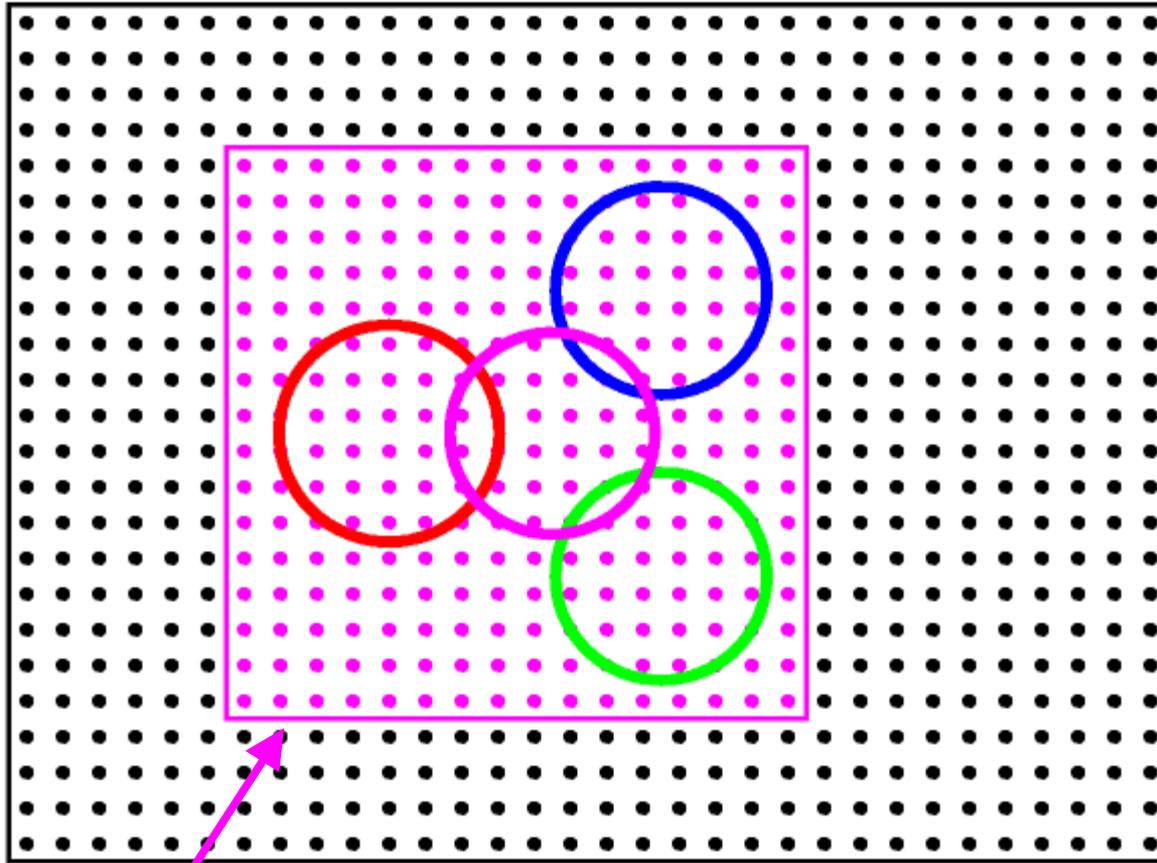
C.-K. Skylaris, O. Dieguez, P. Haynes and M. C. Payne, *Phys. Rev. B* **66**, 073103 (2002).

FFT box technique



simulation cell

FFT box technique

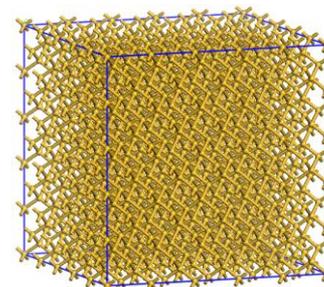
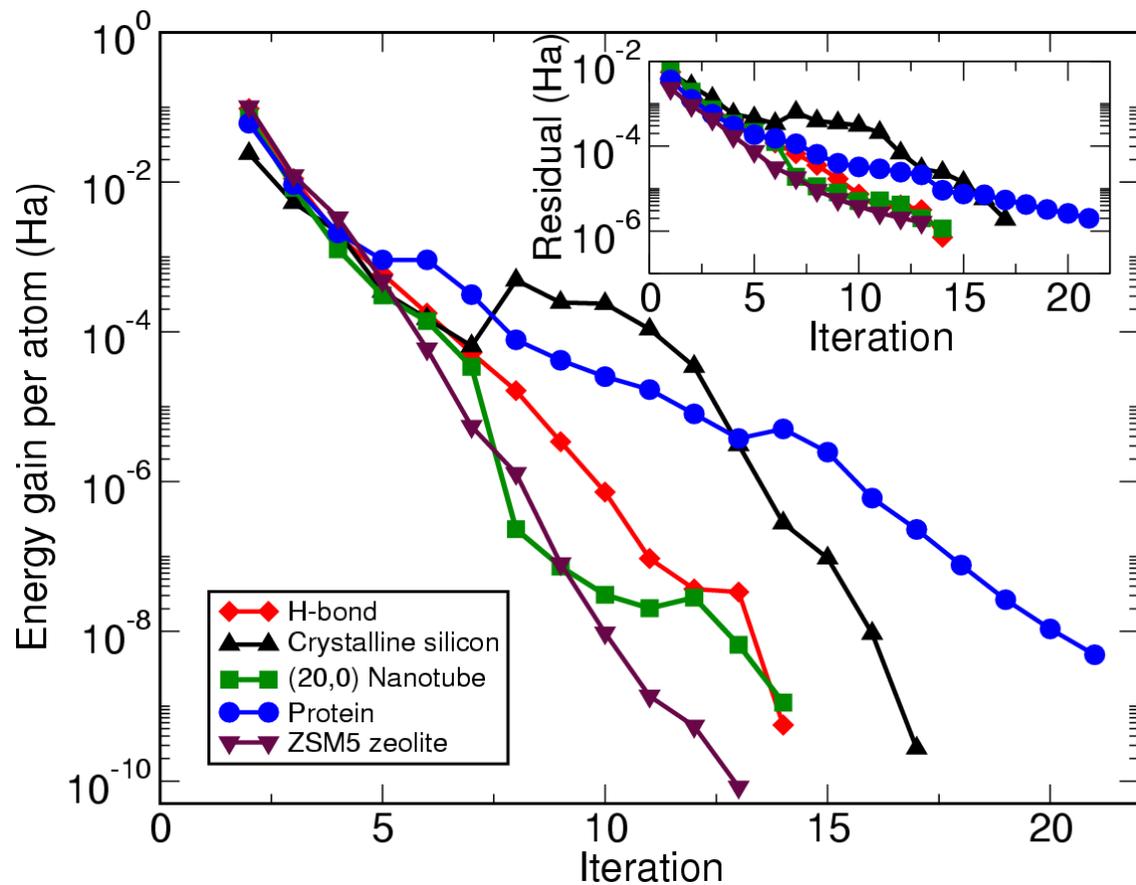
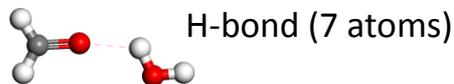


FFT box

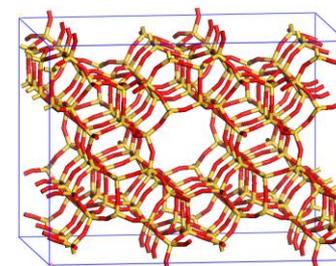
C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, C. J. Pickard & M. C. Payne, *Comp. Phys. Comm.* **140**, 315 (2001)

A. A. Mostofi, C.-K. Skylaris, P. D. Haynes & M. C. Payne, *Comp. Phys. Comm.* **147**, 788 (2002)

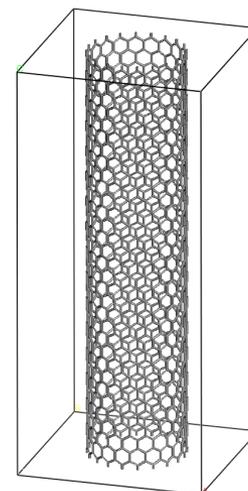
True linear scaling



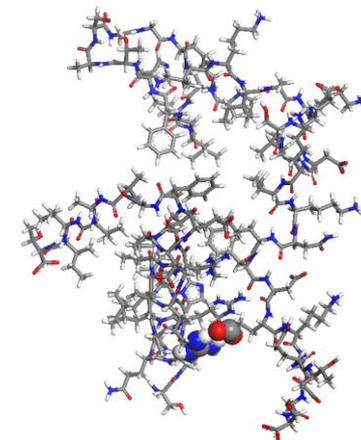
Crystalline silicon
(1000 atoms)



ZSM5 zeolite
(576 atoms)



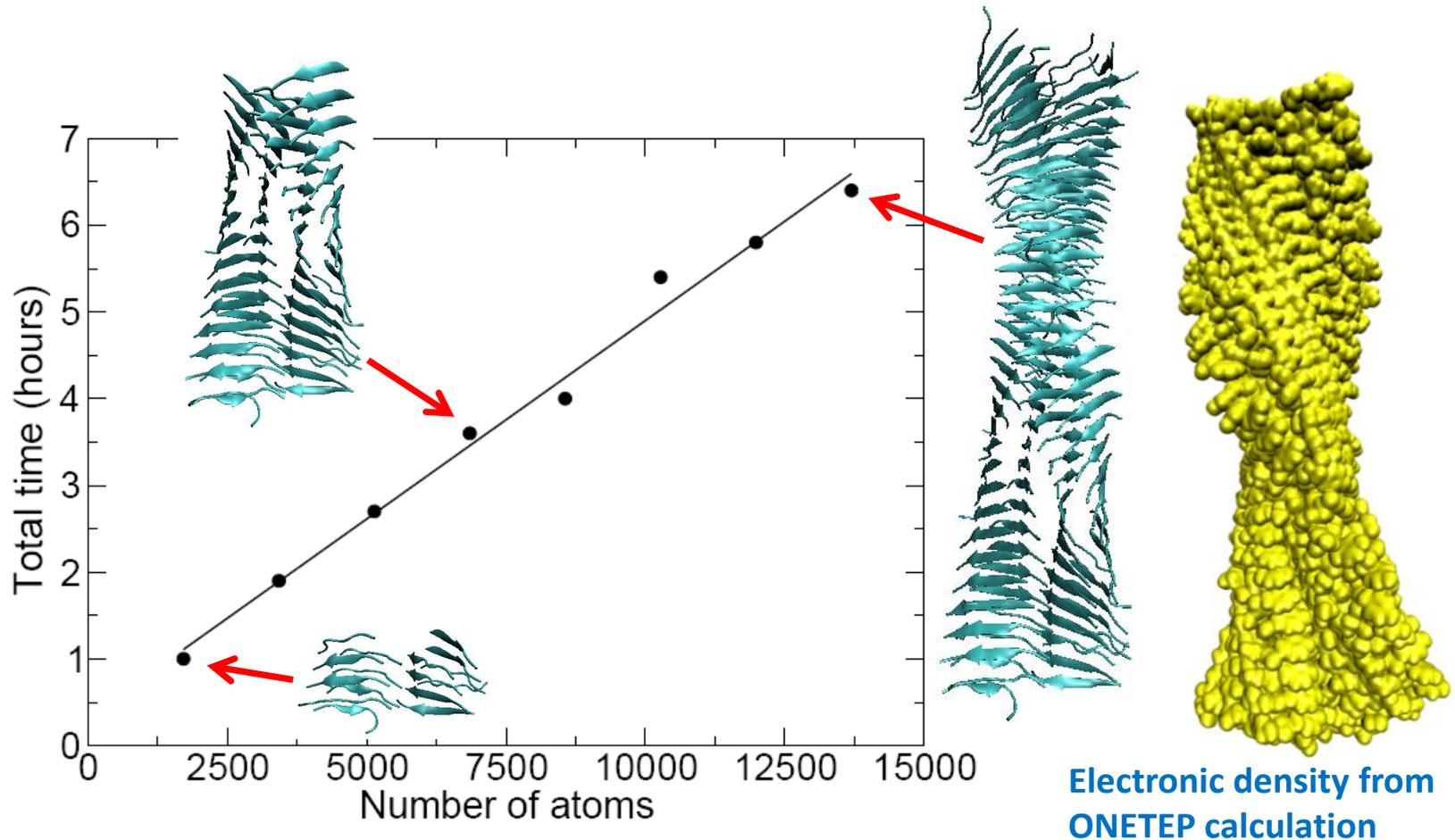
(20,0) Nanotube
(1280 atoms)



Protein (988
atoms)

Skylaris, Haynes, Mostofi & Payne, *J. Phys.: Condens. Matter* **17**, 5757 (2005)

Linear-scaling: Amyloid fibrils



Structures of the amyloid fibril kindly provided by the authors of
J. T. Berryman, S. E. Radford and S. A. Harris, Biophysical Journal, 97 1 (2009)

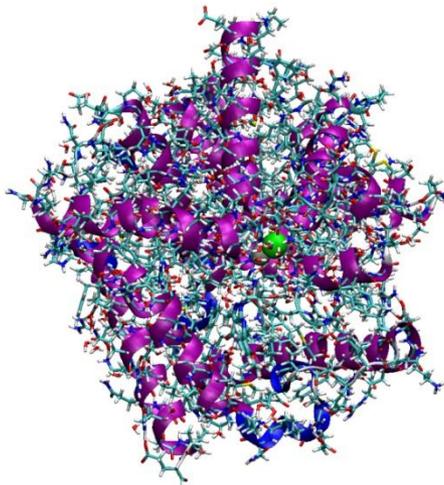
Parallel implementation

- Message Passing Interface (MPI), Fortran 2003 and standard mathematical libraries

Skylaris, Haynes, Mostofi & Payne, *Phys. Stat. Sol. (b)* **243**, 973 (2006)

Hine, Haynes, Mostofi, Skylaris & Payne, *Comput. Phys. Comm.* **180**, 1041 (2009)

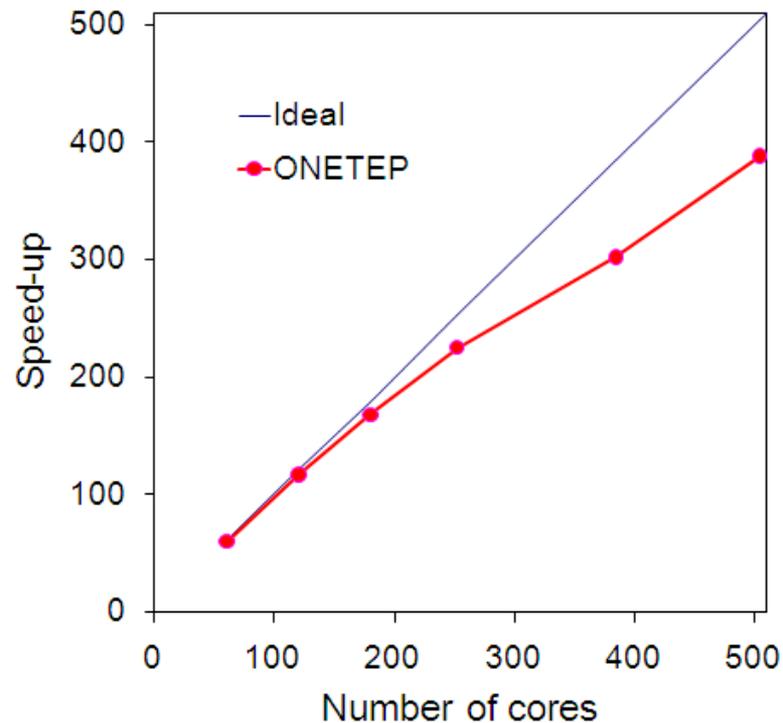
- Example: Speed-up for PDE5 protein (5756 atoms)



- Porting on GPU cards

K. Wilkinson and C.-K. Skylaris, *J Comp. Chem.*

Accepted for publication (2013)



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 2 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
```

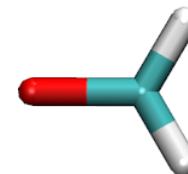
```
O    O    8    4    6.5
C    C    6    4    6.5
H    H    1    1    6.5
```

```
%endblock species
```

```
%block species_pot
```

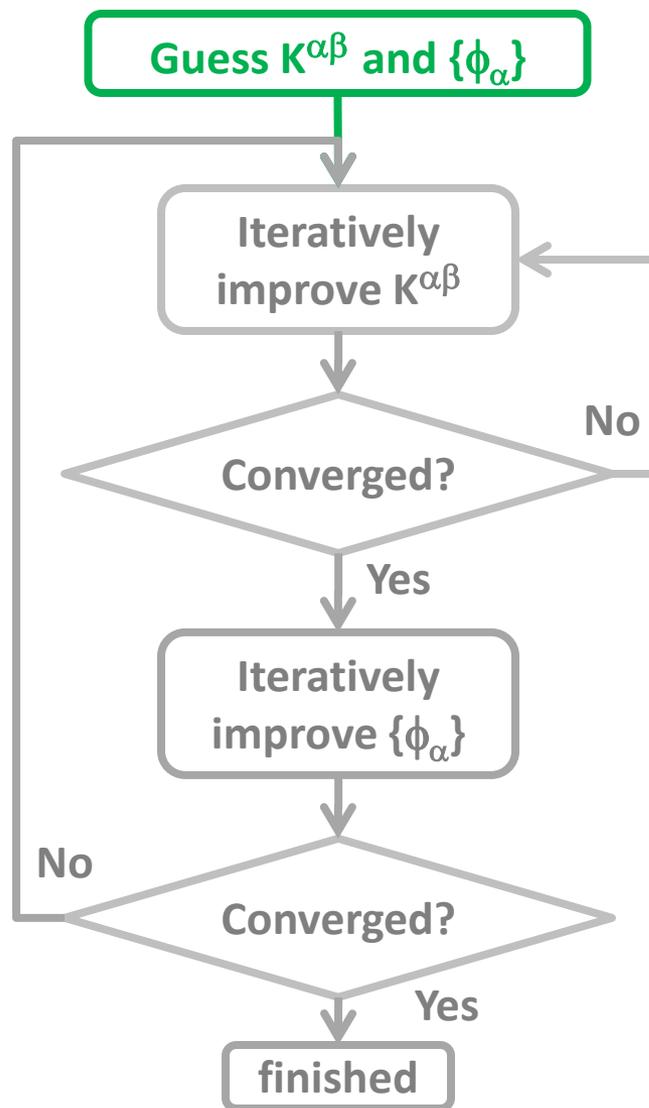
```
O    oxygen.recpot
C    carbon.recpot
H    hydrogen.recpot
```

```
%endblock species_pot
```



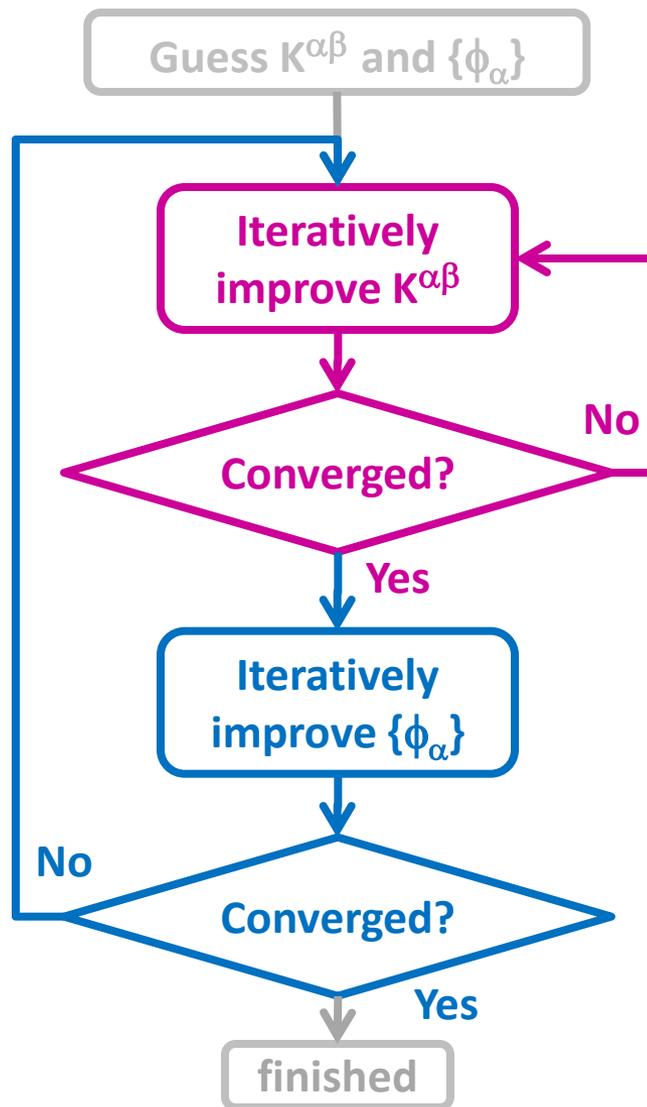
ONETEP calculation outline

- Initialisation phase:
 - Construct initial NGWFs (STOs or PAOs)
 - Construct initial charge density (atomic superposition) and effective potential
 - Construct initial Hamiltonian
 - Obtain initial (non-self-consistent) density kernel using canonical purification
 - Refine initial density kernel (self-consistently) using penalty functional



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



Example output file: formaldehyde

```
+-----+
|
|          ##### #          # ##### ##### ##### #####
|          #      # ##      # #          #      #      #
|          #      # # #      # #          #      #      #
|          #      # # #      # #####      ##### #####
|          #      # #      # # #          #      #      #
|          #      # #      ## #          #      #      #
|          ##### #          # #####      #      ##### #
|
|          Linear-Scaling Ab Initio Total Energy Program
|
|          Release for academic collaborators of ODG
|
|
|
|
|          Version 3.5.2.9
|
+-----+
|
|          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, P. J. Cherry, F. Corsetti, N. Corsini,
|          O. Dieguez, S. M. M. Dubois, H. H. Helal, Q. O. Hill,
|          D. D. O`Regan, C. J. Pickard, L. P. Lee, M. I. J. Probert,
|          L. Ratcliff, M. Robinson, A. Ruiz Serrano, J. S. Spencer,
|          G. Teobaldi, K. A. Wilkinson and T. J. Zuehlsdorff
|
|
|
|
|          Copyright (c) 2004-2013
|
```

Example output file: formaldehyde

- Parallel strategy, calculation parameters

```
Reading parameters from file "h2co.dat" ..... done
Running on      2 processors.
There are      2 MPI processes running on the same node as the root process.
Each MPI process is using:
    1 threads for simulation cell FFTs.
    1 threads for parallel FFT box operations.
    1 threads for loops over batched FFT box operations.
    1 threads in other parallel regions.
```

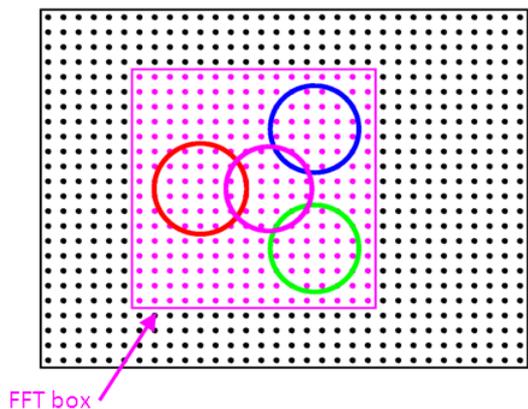
```
Reading geometry and species blocks from file "h2co.dat" ...
  <species_atomic_set> block not found: NGWF initialisation set to SOLVE
... done
```

```
-----
----- RUN-TIME PARAMETERS -----
-----
task                : SINGLEPOINT *! Type of calculation (String)
cutoffenergy        : 22.0493965347193      hartree *! Plane wave kineti
kernelcutoff        : 1000.000000000000      bohr *! Density kernel radiu
xcfunctional         : LDA *! Exchange-correlation functional (String)
```

Example output file: formaldehyde

```
----- Atom counting information -----  
Symbol      Natoms      Nngwfs      Nprojs  
O            1            4            1  
C            1            4            1  
H            2            2            0  
.....  
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

- End of SCF calculation

```
.....  
| *** NGWF optimisation converged *** |  
| RMS NGWF gradient = 0.00000060044935 |  
| Criteria satisfied: |  
| -> RMS NGWF gradient lower than set threshold. |  
=====
```

```
=====
```

```
----- ENERGY COMPONENTS (Eh) -----  
| Kinetic : 14.91654726777243 |  
| Pseudopotential (local) : -75.55457414479615 |  
| Pseudopotential (non-local): 3.08869996279848 |  
| Hartree : 29.55310839834672 |  
| Exchange-correlation : -5.51704179513331 |  
| Ewald : 10.90369328705708 |  
| Total : -22.60956702395475 |  
-----  
Integrated density : 11.99999999999890  
=====
```

```
<<<<< CALCULATION SUMMARY >>>>>
```

ITER	RMS GRADIENT	TOTAL ENERGY	step	Epredicted
1	0.00339839407455	-22.36004521507221	0.577465	-22.56562648226837
2	0.00109741818065	-22.56841844240503	0.775342	-22.59768200552023
3	0.00052360108010	-22.59898597853532	0.778386	-22.60588073836902
4	0.00030634807668	-22.60616555840943	0.779132	-22.60851978307256
5	0.00017102918908	-22.60861455112408	0.617621	-22.60917144516267
6	0.00010980625916	-22.60917266853535	0.829103	-22.60948153377710
7	0.00006521343432	-22.60948683504158	0.578949	-22.60956387837020
8	0.00000603632378	-22.60956697055239	0.021300	-22.60956699447697
9	0.00000390750401	-22.60956702082759	0.000047	-22.60956702085166
10	0.00000388098078	-22.60956702088541	0.003009	-22.60956702376849
11	0.00000060044935	-22.60956702395475	<-- CG	

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 approach)
- Hartree-Fock exchange and hybrid functionals (coming soon)
- 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 orbitals

- Instead of NGWFs construct and use SZ, SZP, DZ, DZP, etc atomic orbital basis sets

Summary of Functionality

Electronic properties

- Density of states, local density of states
- Atomic charges
- Dipole 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

More information

- www.onetep.org
- *J. Chem. Phys.* **122**, 084119 (2005)
- Scientific highlight of the month:
 - Ψ_k Newsletter **72**, December 2005
 - <http://psi-k.dl.ac.uk/>

