ONETEP: Linear-Scaling Density Functional Theory with Tens of Thousands of Atoms

Nicholas Hine

Condensed Matter Theory Group & Thomas Young Centre
Department of Physics & Department of Materials
Imperial College London, United Kingdom

Oak Ridge National Laboratory International Supercomputing Symposium, 13th-15th October 2008

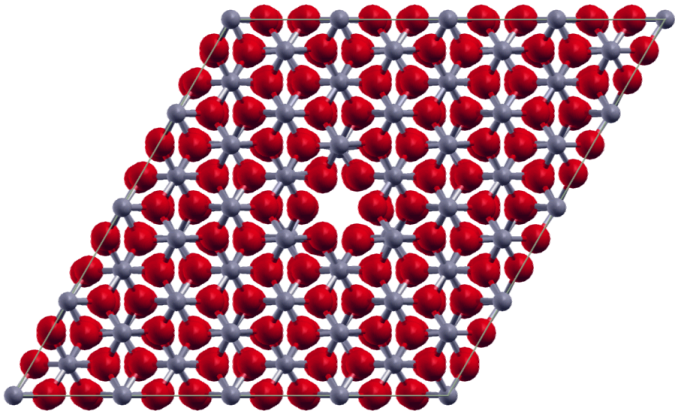
Outline

1. Motivation
2. The ONETEP Code
3. Results

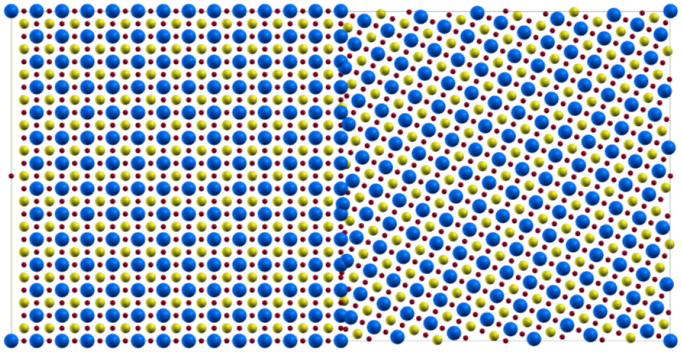
Outline

1. Motivation
2. The ONETEP Code
3. Results

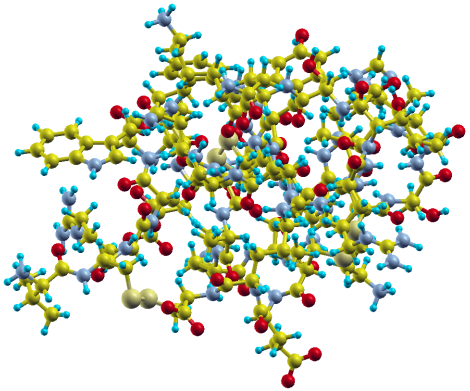
Challenges in Materials Science



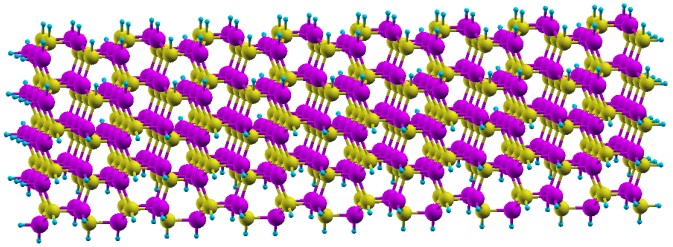
Defects



Interfaces



Biomolecules



Nanostructures

Ab Initio Simulation

Solve Schrödinger's Equation for many-body $\Psi(\mathbf{R})$:

$$\hat{H}\Psi(\mathbf{R}) = E\Psi(\mathbf{R})$$

Density Functional Theory reduces this to solving a system of independent particles interacting only via a mean field:

$$\nabla^2\psi_{n\mathbf{k}}(\mathbf{r}) + V_{\text{eff}}(\mathbf{r})\psi_{n\mathbf{k}}(\mathbf{r}) = \epsilon_{n\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{r})$$

Effective Potential $V_{\text{eff}}(\mathbf{r})$ includes contributions from electron density, ion cores, and any other terms in $\hat{H}$.

Density Functional Theory

Advantages:

- Unbiased (no empirical parameters!)
- Accurate (not quite 'chemical' accuracy, but...)
- Predictive (crystal structures, binding energies, surface reconstructions, etc)

Disadvantages:

- Requires some approximations (exchange-correlation potential, pseudopotentials)
- Slow (iterate to SCF convergence, large basis sets)
- Eigenstates of $\hat{H}$ are *extended* over whole system
  $\Rightarrow$ *CUBIC* scaling with system size!

Density Functional Theory

Advantages:

- Unbiased (no empirical parameters!)
- Accurate (not quite 'chemical' accuracy, but...)
- Predictive (crystal structures, binding energies, surface reconstructions, etc)

Disadvantages:

- Requires some approximations (exchange-correlation potential, pseudopotentials)
- Slow (iterate to SCF convergence, large basis sets)
- Eigenstates of $\hat{H}$ are *extended* over whole system
  $\Rightarrow$ *CUBIC* scaling with system size!

Linear Scaling Density Functional Theory

Consider the single particle density matrix

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{n\ (\mathrm{occ})} \int_{BZ} \psi_{n\mathbf{k}}(\mathbf{r}) f_{n\mathbf{k}} \psi^*_{n\mathbf{k}}(\mathbf{r}') \mathrm{d}\mathbf{k}$$

In an insulator, DM is 'near-sighted': detail in one region does not affect far away regions.

Write density matrix in terms of strictly localized functions ϕ_α and *truncate* $K^{\alpha\beta}$ beyond fixed range:

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

'Nonorthogonal Generalized Wannier Functions' ϕ_α are subspace rotation and unitary transformation of the $\psi_{n\mathbf{k}}$'s.

$K^{\alpha\beta}$ is the generalisation of $f_{n\mathbf{k}}$ — the 'Density Kernel' (Density Matrix in basis of duals of NGWFs)

Linear Scaling Density Functional Theory

Consider the single particle density matrix

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{n\ (\mathrm{occ})} \int_{BZ} \psi_{n\mathbf{k}}(\mathbf{r}) f_{n\mathbf{k}} \psi^*_{n\mathbf{k}}(\mathbf{r}') \mathrm{d}\mathbf{k}$$

In an insulator, DM is 'near-sighted': detail in one region does not affect far away regions.

Write density matrix in terms of strictly localized functions ϕ_α and *truncate* $K^{\alpha\beta}$ beyond fixed range:

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

'Nonorthogonal Generalized Wannier Functions' ϕ_α are subspace rotation and unitary transformation of the $\psi_{n\mathbf{k}}$'s.

$K^{\alpha\beta}$ is the generalisation of $f_{n\mathbf{k}}$ — the 'Density Kernel' (Density Matrix in basis of duals of NGWFs)

Linear Scaling Density Functional Theory

In terms of $K^{\alpha\beta}$ and $\{\phi_\alpha(\mathbf{r})\}$, all required quantities can be obtained by construction and manipulation of *sparse* matrices. Eg for the Hamiltonian:

$$H_{\alpha\beta} = \int \phi_\alpha(\mathbf{r})\hat{H}\phi_\beta(\mathbf{r})\,\mathrm{d}\mathbf{r}$$

and for the density:

$$n(\mathbf{r}) = \phi_\alpha(\mathbf{r})K^{\alpha\beta}\phi_\beta(\mathbf{r})$$

So we can find the total energy of the system from

$$E_{\mathrm{tot}} = \mathrm{Tr}(K^{\alpha\beta}H_{\alpha\beta}) - E_{DC}[n(\mathbf{r})]$$

Outline

1 Motivation

2 The ONETEP Code

3 Results

ONETEP

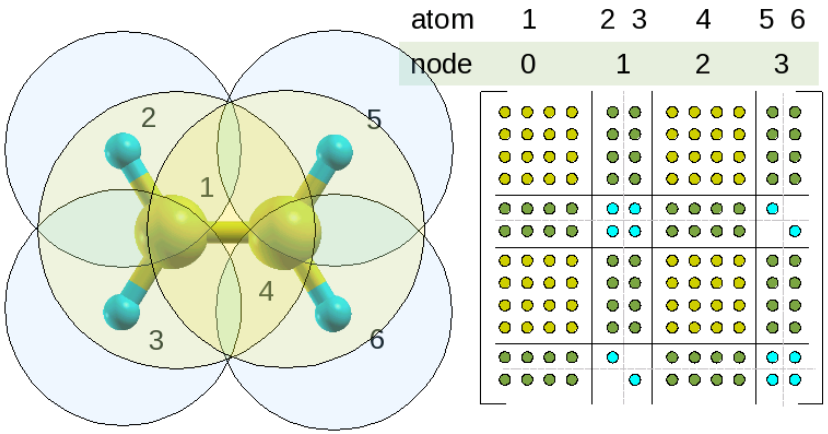
Order-N Electronic Total Energy Package

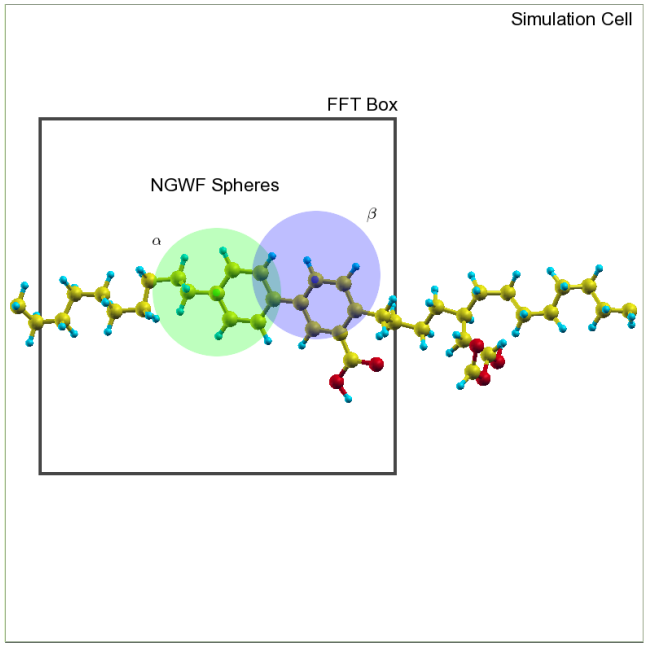
Developers

- Chris-Kriton Skylaris
- Peter Haynes
- Arash Mostofi
- Mike Payne
- Nicholas Hine, Mark Robinson, Quintin Hill, Oswaldo Dieguez, ...

Implementation in ONETEP

NGWFs centered on an atom, each overlapping an N-independent number of nearby others, in blocks of 1,4 or 9 $\rightarrow$ Highly optimised sparse algebra system able to scale to hundreds of CPUs efficiently.



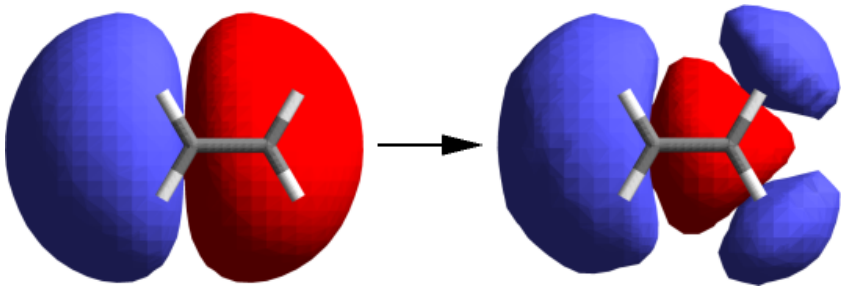


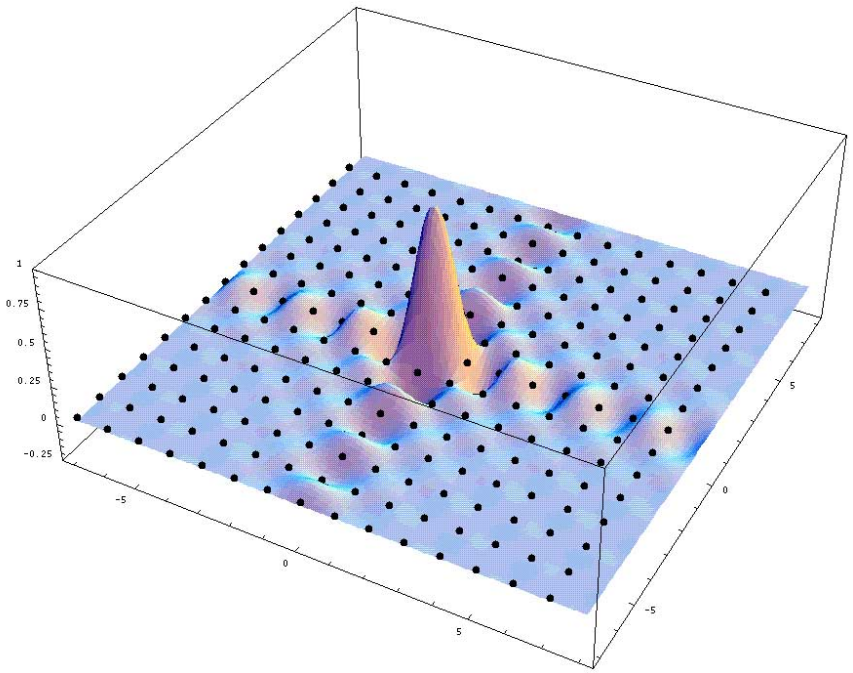
Localised nature of NGWFs allows kinetic energy & other operators to be calculated inside a moving 'FFT box' cut out of the simulation cell.

Implementation in ONETEP

Two loop strategy:

- Outer loop: Minimise E wrt NGWFs for optimal basis
- Inner loop: Optimise Kernel with fixed NGWFs.





Psinc functions: all the advantages of plane-waves... but localized.

ONETEP Architecture

Computational Resources for Total Energy of N atom system

- $O(N)$ scaling of Total Computational Effort
- $O(N)$ scaling of Total Memory Required

Parallelisation over N_p parallel processors

- $O(1/N_p)$ scaling of Total Time of Calculation
- $O(1/N_p)$ scaling of Memory per Processor
- Efficient Load Balancing
- Efficient Communications

ONETEP Architecture

Computational Resources for Total Energy of N atom system

- $O(N)$ scaling of Total Computational Effort
- $O(N)$ scaling of Total Memory Required

Parallelisation over N_p parallel processors

- $O(1/N_p)$ scaling of Total Time of Calculation
- $O(1/N_p)$ scaling of Memory per Processor
- Efficient Load Balancing
- Efficient Communications

Main Challenges

- Sparse Algebra system able to scale from 100s to 10000s of atoms, and 1 to 1000s of CPUs, with filling fractions from 0.01% to 100%.
- Efficient calculation of operator integrals requiring NGWF communication.
- Efficient initialisation (minimise $O(N^2)$ routines during startup)
- Data distribution over CPUs means efficient algorithms for deposition and extraction to and from whole-cell data are vital.

Outline

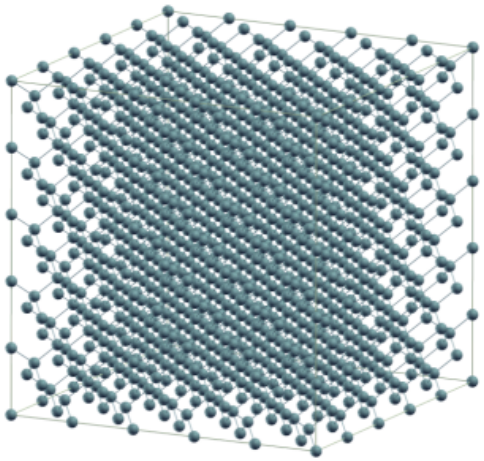
1 Motivation

2 The ONETEP Code

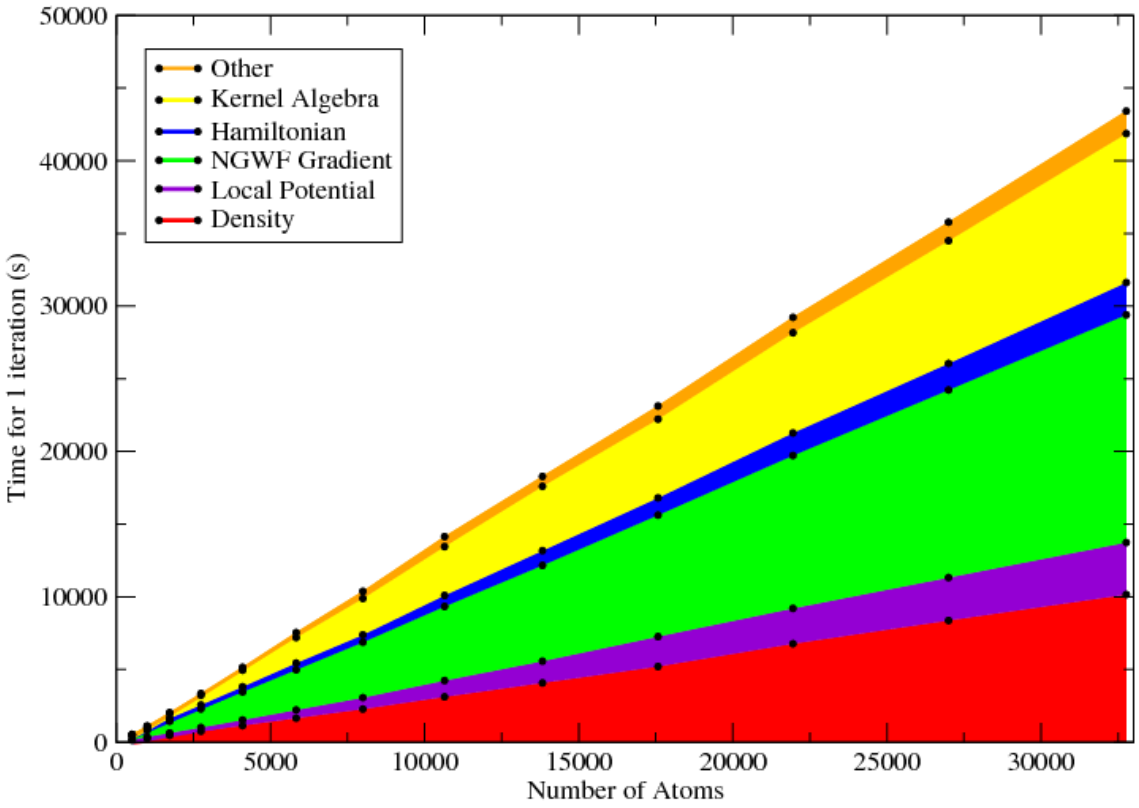
3 Results

Scaling with System Size

Crystalline Silicon, $M \times M \times M \times 8$ atom cell.

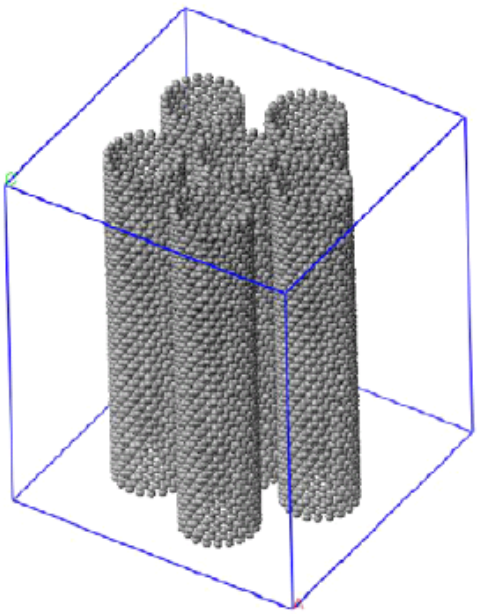


8 x (5 x 5 x 5) = 1000 atoms

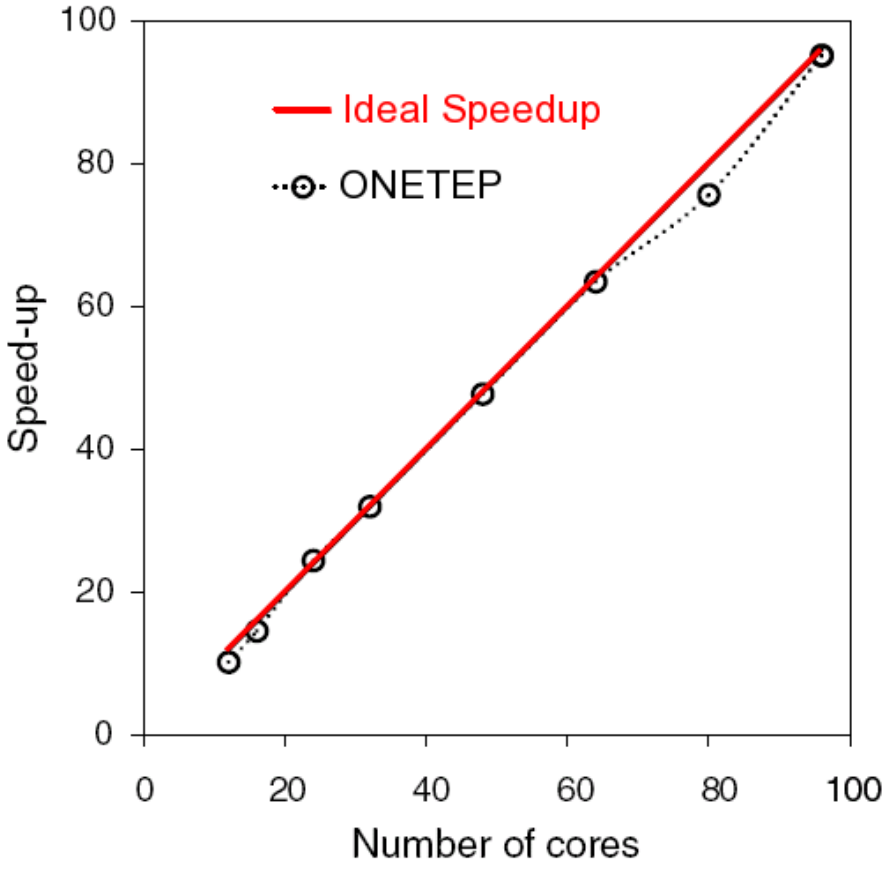


Scaling with Number of Cores

Carbon Nanotubes



9600 atoms



Applications

- Single-Point Energy
- Geometry Optimization
- Molecular Dynamics
- Transition State Search

- Catalysis
- Materials Design
- Surface Reconstructions
- Nanostructure Simulations
- Many More!

Summary

- Linear-Scaling Density Functional Theory with 'Plane-Wave' Accuracy scaleable to tens of thousands of atoms on thousands of cores.
- Highly Efficient Parallelization makes possibility of enormous *ab Initio* simulations available to commodity clusters
- Development work ongoing to improve parallel efficiency and add new features.