Introducing ONETEP

Part II - Efficient implementation of a parallel code

Chris-Kriton Skylaris

Peter D. Haynes

Arash A. Mostofi

Mike C. Payne

Electronic Structure Discussion Group

Theory of Condensed Matter, Cavendish Laboratory

Cambridge, 19 May 2004
ONETEP: A density-matrix linear-scaling DFT method

Molecular orbitals (MOs) \[ \rho(r, r') = \sum_n f_n \psi_n(r) \psi_n^*(r') \]

Non-orthogonal Generalised Wannier Functions (NGWFs) \[ = \sum_{\alpha \beta} \phi_\alpha(r) K^{\alpha \beta} \phi_\beta^*(r') \]

- Optimise non-orthogonal localised functions \( \{ \phi_\alpha(r) \} \) instead of orthogonal extended wavefunctions \( \{ \psi_n(r) \} \) \{ linear scaling \}

- Aim: to achieve the same accuracy as traditional plane-wave methods
PSINC basis set (=plane waves) for the NGWFs

\[
D(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{G}}^{G_{\text{max}}} e^{i\mathbf{G} \cdot \mathbf{r}}
\]

\[
\phi_\alpha(\mathbf{r}) = \sum_{\text{sim. cell}}^{\text{sim. cell}} D_m(\mathbf{r}) C_{m\alpha}
\]

\[
= \sum_{\text{sim. cell}}^{\text{sim. cell}} D(\mathbf{r} - \mathbf{r}_m) C_{m\alpha}
\]

\[
C_{m\alpha} = 0 \text{ if } m \notin \text{sphere of } \alpha
\]
DFT always computationally demanding – ONETEP O(N) scheme should take full advantage of parallel computers

$$E[n] = E[\{K^{\alpha \beta}\}, \{\phi_{\alpha}(r)\}] = E[\{K^{\alpha \beta}\}, \{C_{m\alpha}\}]$$

ONETEP two-nested-loop CG optimisation scheme

\[
n(r) = \sum_{\alpha \beta} \phi_{\alpha}(r) K^{\alpha \beta} \phi^{\ast}_{\beta}(r)
\]

\[
F[\{C_{m\alpha}\}] = \text{minimise } F[\{C_{m\alpha}\}] \text{ w.r.t. } \{C_{m\alpha}\}
\]
\[
\text{minimise } E[\{K^{\alpha \beta}\}, \{C_{m\alpha}\}] \text{ w.r.t. } \{K^{\alpha \beta}\}
\]
\[
\text{keep the } \{C_{m\alpha}\} \text{ fixed}
\]

\[
\frac{\partial E}{\partial K^{\alpha \beta}} \propto H_{\alpha \beta} = \langle \phi_{\alpha} | \hat{H} | \phi_{\beta} \rangle
\]

\[
\frac{\partial F}{\partial C_{m\alpha}} \propto [\hat{H} \phi_{\beta}](r_{m}) K^{\beta \alpha}
\]
Parallel implementation using the Message Passing Interface (MPI) paradigm – each processor runs its own copy of the program with its own data

Parallelisation of data: Distribution of atoms (and NGWFs) to processors according to a space-filling curve

without SF curve

with SF curve
Effect on sparsity pattern of Hamiltonian matrix

without SF curve  4650 x 4650  with SF curve
“Small” sparse matrices!

BRC4-RAD51 complex
(3000 atoms)

ONETEP NGWFs:
7600 x 7600

1% -> 4.4 MB!

Can scale up to 100 processors (~5000 atoms) without data-parallel matrices!

Atomic orbital “DZP” basis:
27500 x 27500

\[ \rho(r, r') = \sum_{\alpha\beta} \phi_{\alpha}(r) K_{\alpha\beta} \phi^*_{\beta}(r') \]
PPD & FFTbox representation of functions

\[ \phi_\alpha \text{ in PPDs} \]

\[ \phi_\alpha \text{ in FFTbox} \]

- Compact storage in 1D arrays
- Fast communication

- QM operators
- Sums, products, interpolation

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Quantum mechanical operators delocalise to the whole volume of the FFTbox. **Computationally expensive!**

\[ \hat{O} \varphi_\beta \]

- \( \hat{T} \) local
- \( \hat{V} \) non-local
- Interpolation to 2Gmax PSINC basis (fine grid)

Calculation of \( \hat{O} \varphi_\beta \) for each \( \varphi_\alpha \) is linear-scaling but not ideal (large prefactor)!

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
There is a way to reduce the prefactor: Enlarge FFTbox!

Smallest FFT box ensuring:

• Hermiticity

• Uniform representation for each operator

• But: $\varphi_\beta$ needs to be re-centred for every single $\varphi_\alpha$ which overlaps with it!

Same advantages plus:

• $\varphi_\beta$ does not need to be re-centred. $\hat{O}\varphi_\beta$ needs to be calculated only once!
Efficient calculation of integrals $<\phi_\alpha | \hat{O} | \phi_\beta >$

$\phi_\alpha$ is *never* placed in FFTbox!

$<\phi_\alpha | \hat{O} | \phi_\beta > = \begin{array}{c}
\text{dot product of (small) 1D arrays}
\end{array}$

Extract **only** PPDs in common with $\phi_\alpha$

Keep $\hat{O} \phi_\beta$ FFTbox in memory and re-use for $<\phi_\beta | \hat{O} | \phi_\beta >$, $<\phi_\gamma | \hat{O} | \phi_\beta >$, etc.

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Efficient calculation of the charge density, $n(r)$

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

$$n(r) = \sum_\alpha n(r; \alpha)$$

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

**Only two interpolations per $n(r;\alpha)$, independent of num $\beta$!**

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Efficient calculation of the NGWF gradient

\[
\frac{\partial E}{\partial C_{m\alpha}} = \hat{H} \phi_{\beta}(r_m) K^{\beta\alpha} + C_{m\beta} Q^{\beta\alpha}
\]

Only one application of \(\hat{H}\) per \(\phi_{\alpha}\)!

(extracted only PPDs belonging to \(\phi_{\alpha}\))

“shave” values outside region of \(\phi_{\alpha}\)

(restricted to region of \(\phi_{\alpha}\) - suitable for updating \(\phi_{\alpha}\) in conjugate gradients optimisation)
Linear-scaling with small prefactor

$\sim N^3$

$\sim N$

Number of atoms

Calculation time (hours)

ONETEP

CASTEP

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Communication model

Functions on different processors which overlap need to be communicated during computation

• Use only **point-to-point** communication

• Keep functions to be sent in buffers and interleave communication with computation by using **non-blocking sends**

• Only use **PPD representation** during communication

• Keep in memory **batches of Ôφβ** to minimise communication

• **Send only if** there is an overlap with current batch of functions of receiving processor

• Work with **processor-processor blocks** of functions/matrices – do $N_{\text{processor}}$ supersteps
Communication model **simplest case – evaluation of integrals**: outline from the viewpoint of processor X

1. Initialise my NGWF batch
2. Apply $\hat{O}$ to my batch

I will be sending to $Y$ & receiving from $Z$

- **Send** my $\phi_\theta$ if needed by $Y$
- **Receive** $\phi_\eta$ from $Z$ if overlap with any member of my batch
- \( \phi_\eta \)

store batch integrals in sparse form

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Speedup with increasing number of processors

- **BRC4 (570 atoms)**
- **BN DWNT (1200 atoms)**
- **Ideal**
True linear-scaling: Number of iterations small, independent of N

Gly

Gly50

Gly200

Nanotubes

![Graphs showing energy gain and residual norm for different molecules and structures.](image)

Energy gain / atom (Ha)

Iteration

Residual norm (Ha)

0 5 10 15 20 25

Gly (10 atoms)
Gly50 (353 atoms)
Gly200 (1403 atoms)
Nanotubes (516 atoms)
Chemical accuracy, no Basis Set Superposition Error

Basis sets, number of localised functions

CASTEP
Plane-waves 1292 eV

ONETEP
Plane-waves 1292 eV
  H: 1 NGWF
  O: 4 NGWFs

NWCHEM
cp-pVTZ+diffuse
  H: 25 contracted Gaussians
  O: 55 contracted Gaussians

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Simulation cell size cost is $O(0)$ in number of atoms. Can do plane-wave calculations in huge simulation cells!

ONETEP calculations of carbon nanotube tip in uniform external electric field (C.-K. Skylaris & G. Csányi et al.)

Local potential

Charge density on local potential iso-surface near Fermi level

Chris-Kriton Skylaris. ESDG 19/5/2004. TCM group, Cavendish Laboratory, University of Cambridge.
Conclusions

• ONETEP – linear-scaling total energy code – takes full advantage of parallel computers

• Accuracy equivalent to conventional cubic-scaling Plane-wave / Gaussian DFT codes

• Low prefactor – breakeven with cubic-scaling codes in the region of a few hundred atoms

• Work in progress: data-parallelisation of simulation cell – even larger simulation cells

• A whole new level of large scale first principles simulations from condensed matter physics to biology now possible - see next week’s instalment which will be brought to you by Peter D. Haynes