

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

Nicholas D.M. Hine¹, Arash A. Mostofi¹, Peter D. Haynes¹

1. Blackett Laboratory and Thomas Young Centre, Imperial College London, Exhibition Road, London SW7 2AZ, U.K.

Abstract

We present recent improvements to the ONETEP code. ONETEP is an ab initio electronic structure package for total energy calculations within density-functional theory. Its main distinguishing features are true 'linear scaling', in that the total computational effort scales only linearly with system size, and 'plane-wave' accuracy, in that the convergence of the total energy is systematically improvable with increasing cutoffs. We present recent improvements to the parallel performance of the code, and thus in effect considerable increases in the scope and scale of feasible calculations with ONETEP, especially in solids. On parallel computers comprising large clusters of commodity servers, our recent improvements make calculations of tens of thousands of atoms in a solid feasible even for small numbers of cores (10-100). Efficient scaling with number of atoms is demonstrated for large systems, and a number of new applications are presented.

ONETEP Theory

Traditional Kohn-Sham DFT finds single-electron states $\psi_i(\mathbf{r})$ with energies ϵ_i to solve the Schrödinger Equation for an effective potential $V[n](\mathbf{r})$:

$$\hat{H}\psi_i(\mathbf{r}) = \left[-\frac{\hbar}{2m}\nabla^2 + V[n](\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (1)$$

In linear-scaling DFT, we use the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ rather than the eigenstates. In terms of $\psi_i(\mathbf{r})$ and occupation numbers f_i , this is

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i f_i \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}'), \quad (2)$$

or in terms of a set of *localised* nonorthogonal functions $\phi_\alpha(\mathbf{r})$

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r}'), \quad (3)$$

where the matrix $K^{\alpha\beta}$, the density kernel, is a generalisation of occupation numbers to a nonorthogonal basis.

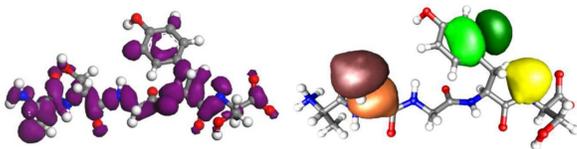


Figure 1: (left) An extended eigenstate for an oligopeptide molecule (right) Example localized NGWFs in the same molecule

Approaches that use eigenstates inevitably scale as $O(N^3)$ with the number of atoms N : the system has $O(N)$ eigenstates, each of size $O(N)$, and each needing to stay orthogonal to $O(N)$ others. Localised-orbital approaches, however, can scale as $O(N)$. In an insulator, the kernel $K^{\alpha\beta}$ can be truncated beyond some cutoff radius R_K , so the matrix is *sparse*. The overlap matrix $S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$ is also sparse for localised $\phi_\alpha(\mathbf{r})$, as are elements of the Hamiltonian matrix $H_{\alpha\beta} = \langle \phi_\alpha | \hat{H} | \phi_\beta \rangle$. With $H_{\alpha\beta}$ and the density $n(\mathbf{r}) = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r})$ we can find the total energy E with $O(N)$ scaling by using

$$E[\{K^{\alpha\beta}\}, \{\phi_\alpha\}] = \sum_{\alpha\beta} K^{\alpha\beta} H_{\beta\alpha} + E_{DC}[n(\mathbf{r})], \quad (4)$$

and simultaneously minimising E with respect to the kernel and the coefficients describing the NGWFs, subject to the constraint that the density kernel remains idempotent and that its trace equals the number of electrons.

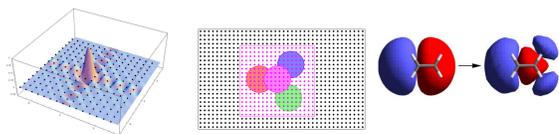


Figure 2: (left) A psinc function (middle) FFT box containing overlapping NGWFs (right) Example of NGWF optimisation of a p -orbital.

ONETEP combines $O(N)$ scaling with 'plane-wave' accuracy, in that the convergence of the total energy is systematically improvable by increasing cutoffs. The localised basis in ONETEP comprises 'Nonorthogonal Generalised Wannier Functions' (NGWFs) expressed in terms of a basis of periodic bandwidth-limited delta functions, or psinc functions, (see Fig 2) strictly localized to spherical regions of radius R_{ϕ_α} . These psinc functions, with coefficients $C_{i,\alpha}$, are centered on the grid points \mathbf{r}_i of a regular grid specified by a plane-wave cutoff energy E_{cut} . The minimisation of the energy occurs via nested loops: The outer loop minimises the energy with respect to the coefficients $C_{i,\alpha}$

$$E_{\text{min}} = \min_{\{C_{i,\alpha}\}} L(\{C_{i,\alpha}\}), \quad (5)$$

while inner loop, performed at fixed $C_{i,\alpha}$, minimizes the energy with respect to the kernel elements $K^{\alpha\beta}$

$$L(\{C_{i,\alpha}\}) = \min_{\{K^{\alpha\beta}\}} E(\{K^{\alpha\beta}\}; \{C_{i,\alpha}\}). \quad (6)$$

Parallel Optimisation

ONETEP was developed from the beginning as a parallel code and its efficient scaling and performance on isolated molecules, nanotubes and similar systems with a high degree of sparsity has been well-documented. Recent work has focused on improving performance in solids, where communications bottlenecks in large systems has previously limited the useful applicability of the code.

Matrix Algebra

One time-limiting component of ONETEP calculations is sparse matrix algebra, especially during kernel optimisation. As Fig 3 shows, the pattern of filling of the sparse matrices representing $S_{\alpha\beta}$, $H_{\alpha\beta}$, and $K^{\alpha\beta}$ can be highly structured, allowing considerable optimisation of the communication and computation patterns. Recent improvements include:

- Hierarchical sparsity: division of matrix into node-node 'segments' and atom-atom 'blocks': dense storage used when filling exceeds threshold η .
- Efficient dense matrix algebra used on the level of segments or segments as appropriate.
- Reduced total comms volume by communicating only those blocks or segments of multipliers contributing to matrix product.

Combined, these developments have dramatically improved both the speed and scaling (with system size and number of parallel processors) of matrix algebra.

NGWF Operations

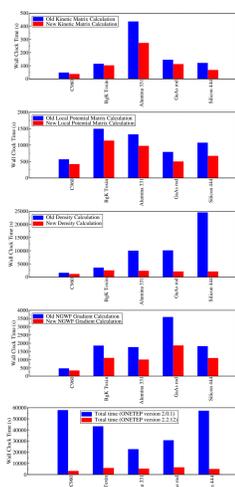


Figure 6: Timings for row sums operations on 16 nodes, for a range of systems (C Nanotube, Organic BgK toxin, Al_2O_3 Crystal, GaAs Nanorod, Si Crystal).

Parallel Scaling

Combined with other parallel optimisations, these improvements have resulted in very considerable decreases in the total computation time:

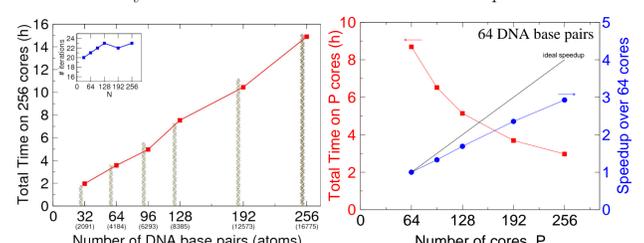


Figure 7: (left) Scaling with system size for DNA strands: clear linear scaling of the total time is observed (right) Scaling with number of cores on which the calculation is run — efficient speedups are obtained up to at least 256 cores.

Recent Applications

GaAs / ZnO Nanorods

The properties of III-V and II-VI semiconductor nanorods in the Wurtzite structure, including polar bonding and lack of inversion symmetry, are of interest to a variety of fields, including spintronics and photovoltaics. Linear-scaling DFT with ONETEP enables simulation of realistic systems of thousands of atoms, elucidating the the complex interplay between bonding and long-range electrostatic effects required to model these systems.

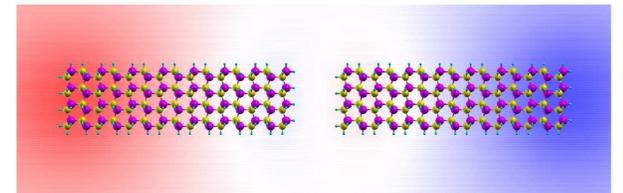


Figure 8: Electrostatic Potential of a pair of GaAs Nanorods.

Defects in Ceramics

Understanding defects and defect clusters in crystalline materials is a tough challenge for electronic structure methods due to the requirement of embedding, large systems in periodic hosts, with significant finite size errors if the supercell used is too small. In metal oxide systems, the presence of oxygen requires a dense grid when using plane-waves and norm-conserving pseudopotentials, making the calculations even tougher. The system sizes available to linear-scaling DFT make calculations on supercells of thousands or even tens of thousands of atoms feasible. This allows investigation of properties of charged defects, including defect formation energies, migration barriers, defect concentrations and diffusion coefficients. Grain boundaries and dislocations are also possible.

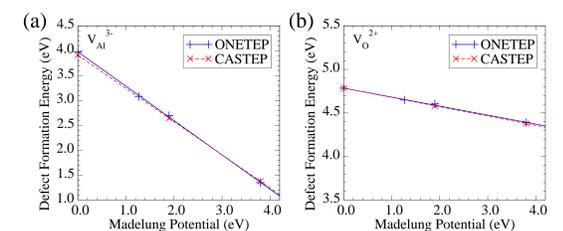


Figure 9: Formation energies of Aluminium and Oxygen vacancies in Al_2O_3 as a function of system size, showing slow convergence and requiring large system sizes for accurate results.

Strongly Correlated Systems

We have recently completed an implementation of the DFT+U method within ONETEP, which allows treatment of systems where so called "strong-correlation" effects are important. This is the case wherever there are electrons occupying localised states in tightly-bound d - or f -subshells which are partially occupied.

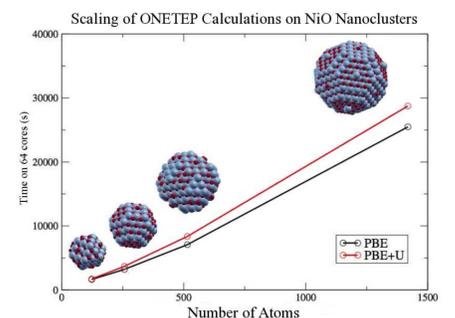


Figure 8: Scaling of DFT+U calculations on NiO nanoclusters. The method incurs minimal overhead compared to a standard calculation.

Protein-Ligand Binding Energies

ONETEP is also ideal for quantum-mechanical investigation of biological systems. Recent investigations have included binding of protein fragments with nanoparticles.

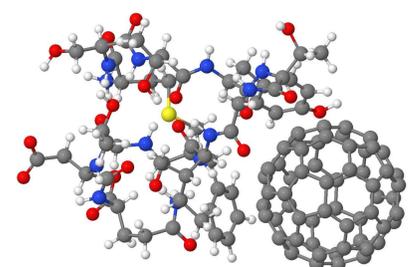


Figure 8: C60 fullerene binding with peptides from a protein involved in formation of amyloid fibrils.