

# ONETEP: Linear-Scaling Density-Functional Theory with Plane-Waves

Published in *Molecular Simulation*, **33** (7), 551 (2007). DOI: 10.1080/08927020600932801.

Arash A. Mostofi\*, Peter D. Haynes\*\*, Chris-Kriton Skylaris†  
and Mike C. Payne\*\*

\* Dept. of Materials Science and Engineering, MIT, Cambridge MA, USA (mostofi@mit.edu)

\*\* Theory of Condensed Matter, Cavendish Laboratory, University of Cambridge, UK

† Physical and Theoretical Chemistry Laboratory, University of Oxford, UK

## ABSTRACT

Conventional methods for atomistic simulations based on density-functional theory (DFT), such as the plane-wave pseudopotential approach, have had an immense impact on the way in which material properties are studied. In spite of this success, the system-size accessible to such techniques is limited because the algorithms scale with the cube of the number of atoms. The quest to bring to bear the predictive power of DFT calculations on ever larger systems has resulted in much recent interest in linear-scaling methods for DFT simulations. To this end we present an overview of ONETEP (Order- $N$  Total Energy Package), our linear-scaling method based on a plane-wave basis set, which is able to achieve the same accuracy and convergence rate as the conventional plane-wave DFT approach. The novel features of our method which result in its success are described and results of calculations on titanium oxide clusters from the ONETEP parallel code are presented.

**Keywords:** linear scaling, order-n, density functional, electronic structure, large-scale simulation

## 1 INTRODUCTION

Density-functional theory (DFT) [1, 2] has made an impact on a broad range of scientific disciplines over the past decade, from condensed matter to the earth sciences to biochemistry. A number of reasons have resulted in the plane-wave (PW) pseudopotential approach [3, 4] becoming one of the methods of choice for simulations based on DFT. For example, with a plane-wave basis set: the kinetic energy operator is diagonal in momentum space; quantities are switched efficiently between real space and momentum space using fast-Fourier transforms; the atomic forces are calculated by straightforward application of the Hellmann-Feynman theorem; the completeness of the basis is controlled systematically with a single parameter.

The resulting simulations require a computational effort which scales as the cube of the system-size [5], which makes the cost of large-scale calculations prohibitive. For this reason there has been much interest in developing methods whose computational cost

scales only linearly with system-size [6] and which enable researchers to bring to bear the predictive power of density-functional calculations on truly nano-scale systems.

At first sight the extended nature of plane-waves makes them unsuitable for representing the localised orbitals of linear-scaling methods. This is why other approaches have focussed on using localised basis sets instead, for example Gaussians [7], numerical atomic orbitals [8, 9], spherical waves [10], B-splines [11] and grids [12, 13].

In spite of this, we have developed ONETEP (Order- $N$  Electronic Total Energy Package) [14, 15], a linear-scaling method based on plane-waves which overcomes the above difficulty and which is able to achieve the same accuracy and convergence rate as traditional  $\mathcal{O}(N^3)$  PW calculations.

The remainder of this paper is organised as follows: in Section 2 the theoretical background to the formulation of linear-scaling methods, and in particular ONETEP, is presented; Section 3 contains results and discussion on titanium oxide clusters designed to represent surfaces.

## 2 THEORETICAL BACKGROUND

In the Kohn-Sham DFT method [1, 2], the many-body Schrödinger equation for a system of  $N$  interacting electrons is mapped onto  $N$  one-body equations describing a system of non-interacting particles, in an effective potential  $V_{\text{eff}}$ , which has the same ground state density as the original interacting problem. For simplicity of notation we consider  $\Gamma$ -point sampling and a spin-unpolarised system:

$$\hat{H}_{\text{KS}}\psi_i(\mathbf{r}) = \left[ -\frac{1}{2}\nabla^2 + V_{\text{eff}}[n](\mathbf{r}) \right] \psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \quad (1)$$

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

where  $\hat{H}_{\text{KS}}$  is the Kohn-Sham effective Hamiltonian and  $i$  is the band index of the Bloch eigenstate  $\psi_i$  with associated eigenvalue  $\varepsilon_i$  and occupancy  $f_i$  (which must be either zero or one at zero temperature). The factor of 2 in the electronic density  $n(\mathbf{r})$  takes account of spin degeneracy. As the effective potential  $V_{\text{eff}}$  is dependent on

the Kohn-Sham eigenstates  $\psi_i$  through the electron density  $n(\mathbf{r})$ , the above system of equations must be solved self-consistently.

The delocalised nature of the Bloch states  $\psi_i$  and the requirement that they are all mutually orthogonal,  $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ , result in the  $N^3$  bottleneck in conventional PW pseudopotential approaches.

An equivalent description of the system may be given in terms of the single-particle density matrix:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i f_i \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}'). \quad (3)$$

The density  $n(\mathbf{r})$  is given by the diagonal elements of the density matrix:  $n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r})$ , and the total energy of the non-interacting system is

$$E_0 = 2 \sum_i f_i \varepsilon_i = \text{Tr}[\hat{\rho} \hat{H}_{\text{KS}}], \quad (4)$$

to which may be applied the usual double-counting corrections for the Hartree and exchange-correlation terms in order to obtain the total energy of the interacting system.

## 2.1 Linear-Scaling Methods

Linear-scaling methods exploit the “nearsightedness” [16, 17] inherent in quantum many-body systems by exploiting the localisation of Wannier functions [18–21] or the density matrix [22, 23]. In practice, this is done by expressing the density matrix in the following separable form [24, 25]:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r}'), \quad (5)$$

where  $\{K^{\alpha\beta}\}$  are the elements of the density kernel [24] and  $\{\phi_\alpha\}$  are a set of non-orthogonal generalised Wannier functions [26] (NGWFs) which span both the occupied space and part of the unoccupied manifold. The orthogonality condition on the Bloch states together with the Pauli exclusion principle for the filling of states are manifested in the idempotency condition on the density matrix,  $\rho^2 = \rho$ .

Linear-scaling is achieved by imposing two spatial cut-offs: first, the NGWFs are truncated such that they are localised within spherical regions of fixed radii  $\{r_\alpha\}$ ; second, elements of the density kernel  $K^{\alpha\beta}$  corresponding to NGWFs whose centers are separated by more than a fixed value  $r_K$  are neglected. Imposing these spatial cut-offs reduces the information content of the density matrix to linear-scaling with system-size. The values of the cut-offs  $\{r_\alpha\}$  and  $r_K$  are treated as variational parameters with respect to which the calculation of physical properties must be converged.

## 2.2 The ONETEP Method

Most linear-scaling methods fall into one of two categories: those in which the total energy functional is minimised with respect to the elements of the density kernel while using a fixed, and usually large, set of localised orbitals [7, 27, 28]; and those, such as ONETEP [14, 15], in which both the density kernel and NGWFs are optimised [11, 13, 14].

Optimising the NGWFs *in situ*, rather than using a fixed set, gives greater accuracy but comes at an extra computational cost, though there is a saving in that a minimal set of NGWFs may be used while maintaining transferability.

In ONETEP minimisation of the total energy functional proceeds as two nested loops: in the inner loop the density kernel is optimised using a combination of the density matrix minimisation method of Li, Nunes and Vanderbilt [29, 30] and the penalty functional method of Haynes and Payne [31]; in the outer loop the NGWFs are optimised and this is discussed below. The procedure is iterated until self-consistency is achieved.

In order to perform the optimisation of the NGWFs it is necessary to expand them in some underlying set of basis functions. In ONETEP these are chosen to be periodic cardinal sine (psinc) functions [32]. They are related to plane-waves by a Fourier transform and therefore, while being localised, they inherit their orthogonality, the systematic control of accuracy with a single parameter, and the ability to use fast Fourier transforms (FFTs) to switch quantities efficiently between real and reciprocal space.

Linear-scaling is obtained by exploiting the strict localisation of the NGWFs by performing FFT operations within a box whose size depends only on the radii of the NGWFs, and is hence independent of system-size. This “FFT box” technique has been demonstrated to be a very accurate approximation [33]. The FFT box also gives an indication of the likely cross-over point, i.e., the system-size at which the linear-scaling method is faster than traditional methods. Both ONETEP and the conventional PW approach invest a large fraction of the total computational effort in performing FFTs. In the case of ONETEP these are nearly all done in the FFT box, whereas conventionally they are all done in the whole simulation cell. The FFT box of ONETEP is typically  $20 \times 20 \times 20 \text{ \AA}^3$  and its dimensions do not vary much with the size or nature of the system. The size of the FFT box corresponds to the volume occupied by a few hundred atoms in a solid, but the cross-over is far more favourable for less densely-packed systems such as biological molecules, nanotubes or systems involving surfaces. Furthermore, with a conventional PW code vacuum requires the same level of description as regions where atomic binding occurs. However in ONETEP, where there are no atoms there are no FFT boxes and

a saving is made. This can reduce the cross-over by an order of magnitude for isolated molecules, clusters and polymers.

Finally, we use a preconditioning scheme [32] that ensures that the number of self-consistent iterations required to reach convergence is independent of system-size, resulting in an overall computational cost for the whole calculation that scales with  $N$ .

### 3 TITANIUM OXIDE CLUSTERS

Within the PW pseudopotential framework, surfaces can be modelled in periodic supercells as infinite slabs with the proviso that one must ensure that there is enough vacuum region perpendicular to the surface to avoid spurious interactions with slabs in neighbouring supercells. As our method uses a psinc basis that is equivalent to plane-waves, this option is also possible with ONETEP. The alternative, however, is to model surfaces as isolated clusters, surrounded by large vacuum regions in all directions. This would be very computationally expensive for a conventional PW code but, because of the strict localisation of the NGWFs, such systems may be studied with ONETEP with little additional cost.

We have performed calculations on clusters ranging in size from ten to 200 atoms. We have used norm-conserving pseudopotentials [34] in separable form [35], a local density approximation [36, 37] for the exchange and correlation functional and an 800 eV cut-off for the psinc basis set. Four NGWFs were associated with each titanium and oxygen atom and the NGWF radii were all set to 3.175 Å.

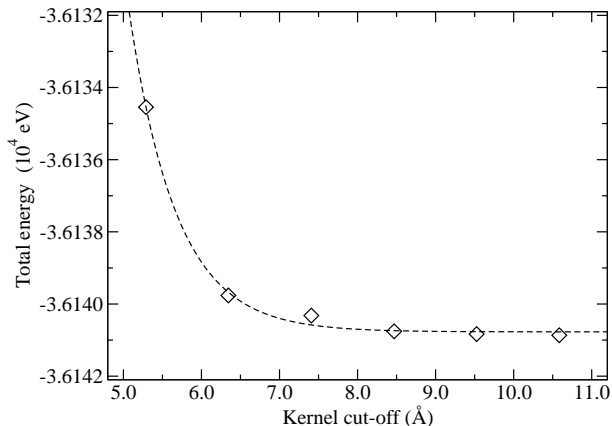


Figure 1: Convergence of the total energy of a  $\text{Ti}_{38}\text{O}_{76}$  cluster with respect to the density kernel cut-off.

Fig. 1 shows how the total energy of a  $\text{Ti}_{38}\text{O}_{76}$  cluster converges as the density kernel cut-off is increased. The trend line shown is a best fit to an exponential decay which reflects the expected behaviour of the density-

matrix. The convergence in practice is not smooth but occurs in jumps as particular matrix elements are included when the cut-off matches the distance between the relevant atoms. In this system, the total energy is converged to  $10^{-3}$  eV per atom for a kernel cut-off of 8 Å.

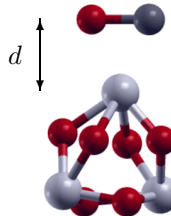


Figure 2: The system used to study the interaction of a CO molecule with a  $\text{Ti}_3\text{O}_6$  cluster.

In order to assess the accuracy of ONETEP in determining physical properties such as binding energies, the interaction of a carbon monoxide molecule with a small  $\text{Ti}_3\text{O}_6$  cluster was studied. The results are compared with calculations on the same system performed with the CASTEP code [38], a conventional cubic-scaling PW pseudopotential DFT code. Equivalent basis set cut-offs and identical pseudopotentials were used. The only difference between the calculations was the use of a  $12 \times 12 \times 12 \text{ \AA}^3$  cell for CASTEP but a  $26 \times 26 \times 26 \text{ \AA}^3$  cell for ONETEP. Table 1 shows the energies obtained from two configurations illustrated in Fig. 2. The difference in the total energies calculated by ONETEP and CASTEP is about 0.4 eV per atom but is mostly accounted for by the variational restriction originating from the truncation radii chosen for the NGWFs. When the energy differences between two configurations are compared the agreement between ONETEP and CASTEP is excellent: better than 1 meV per atom.

Table 1: The total energy  $E$  for a CO molecule located at a distance  $d$  from a  $\text{Ti}_3\text{O}_6$  cluster, calculated with ONETEP and CASTEP.  $\Delta E$  is the energy difference between the two configurations.

	$E$ (eV)		$\Delta E$ (eV)
	$d = 2.000 \text{ \AA}$	$d = 2.252 \text{ \AA}$	
CASTEP	-3419.694	-3419.355	0.339
ONETEP	-3415.708	-3415.361	0.347

Finally the linear-scaling of ONETEP is demonstrated in Fig. 3. The time per iteration (on a Sun Fire V40z server with four 2.2 GHz single-core Opteron CPUs) scales linearly with system-size in accord with results on other systems [14]. The number of iterations fluctuated by up to 20% but did not change systematically

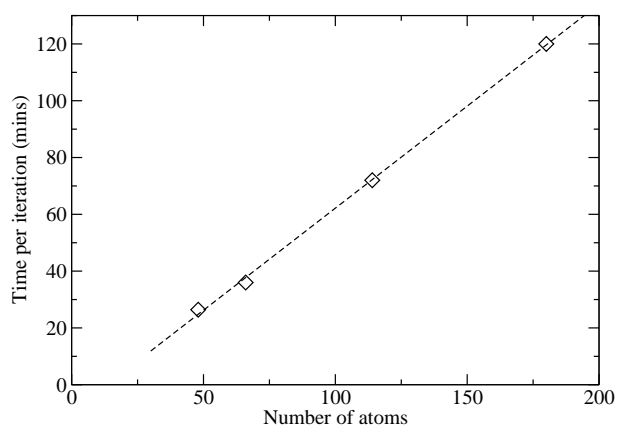


Figure 3: Scaling of computational time per iteration against number of atoms in the  $\text{TiO}_2$  cluster.

with system-size.

## ACKNOWLEDGEMENTS

AAM acknowledges Christ's College, Cambridge for a Research Fellowship and the Royal Society for an equipment grant. PDH and C-KS acknowledge the support of Royal Society University Research Fellowships.

## REFERENCES

- [1] Hohenberg, P. and Kohn, W. *Phys. Rev.* **136**(3B), B864 (1964).
- [2] Kohn, W. and Sham, L. J. *Phys. Rev.* **140**(4A), A1133 (1965).
- [3] Ihm, J., Zunger, A., and Cohen, M. L. *J. Phys. C* **12**(21), 4409–4422 (1979).
- [4] Denteneer, P. J. H. and van Haeringen, W. *J. Phys. C* **18**(21), 4127–4142 (1985).
- [5] Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., and Joannopoulos, J. D. *Rev. Mod. Phys.* **64**(4), 1045 (1992).
- [6] Goedecker, S. *Rev. Mod. Phys.* **71**(4), 1085 (1999).
- [7] Scuseria, G. E. *J. Phys. Chem. A* **103**(25), 4782–4790 (1999).
- [8] Junquera, J., Paz, O., Sánchez-Portal, D., and Artacho, E. *Phys. Rev. B* **64**(23), 235111 (2001).
- [9] Anglada, E., Soler, J. M., Junquera, J., and Artacho, E. *Phys. Rev. B* **66**(20), 205101 (2002).
- [10] Haynes, P. D. and Payne, M. C. *Comput. Phys. Commun.* **102**(1-3), 17 (1997).
- [11] Hernández, E., Gillan, M. J., and Goringe, C. M. *Phys. Rev. B* **55**(20), 13485 (1997).
- [12] Beck, T. L. *Rev. Mod. Phys.* **72**(4), 1041–1080 (2000).
- [13] Fattbert, J.-L. and Bernholc, J. *Phys. Rev. B* **62**(3), 1713 (2000).
- [14] Skylaris, C.-K., Haynes, P. D., Mostofi, A. A., and Payne, M. C. *J. Chem. Phys.* **122**, 084119 (2005).
- [15] Skylaris, C.-K., Haynes, P. D., Mostofi, A. A., and Payne, M. C. *J. Phys.: Condens. Matter* **17**, 5757 (2005).
- [16] Kohn, W. *Phys. Rev. Lett.* **76**(17), 3168 (1996).
- [17] Prodan, E. and Kohn, W. *Proc. Natl. Acad. Sci.* **102**(33), 11635 (2005).
- [18] Kohn, W. *Phys. Rev.* **115**(4), 809 (1959).
- [19] des Cloizeaux, J. *Phys. Rev.* **135**(3A), A698–A707 (1964).
- [20] Nencieu, G. *Commun. Math. Phys.* **91**(1), 81–85 (1983).
- [21] He, L. and Vanderbilt, D. *Phys. Rev. Lett.* **86**(23), 5341 (2001).
- [22] des Cloizeaux, J. *Phys. Rev.* **135**(3A), A685–A697 (1964).
- [23] Ismail-Beigi, S. and Arias, T. A. *Phys. Rev. Lett.* **82**(10), 2127 (1999).
- [24] McWeeny, R. *Rev. Mod. Phys.* **32**(2), 335–369 (1960).
- [25] Hernández, E. and Gillan, M. J. *Phys. Rev. B* **51**(15), 10157–10160 (1995).
- [26] Skylaris, C.-K., Mostofi, A. A., Haynes, P. D., Diéguez, O., and Payne, M. C. *Phys. Rev. B* **66**, 035119 (2002).
- [27] Soler, J. M., Artacho, E., Gale, J. D., García, A., Junquera, P., Ordejón, P., and Sánchez-Portal, D. *J. Phys.: Condens. Matter* **14**(11), 2745–2780 (2002).
- [28] Challacombe, M. *J. Chem. Phys.* **110**(5), 2332 (1999).
- [29] Li, X.-P., Nunes, R. W., and Vanderbilt, D. *Phys. Rev. B* **47**(16), 10891 (1993).
- [30] Nunes, R. W. and Vanderbilt, D. *Phys. Rev. B* **50**, 17611 (1994).
- [31] Haynes, P. D. and Payne, M. C. *Phys. Rev. B* **59**(19), 12173 (1999).
- [32] Mostofi, A. A., Haynes, P. D., Skylaris, C.-K., and Payne, M. C. *J. Chem. Phys.* **119**(17), 8842–8848 (2003).
- [33] Mostofi, A. A., Skylaris, C.-K., Haynes, P. D., and Payne, M. C. *Comput. Phys. Commun.* **147**, 788 (2002).
- [34] Hamann, D. R., Schlüter, M., and Chiang, C. *Phys. Rev. Lett.* **43**(20), 1494 (1979).
- [35] Kleinman, L. and Bylander, D. M. *Phys. Rev. Lett.* **48**(20), 1425 (1982).
- [36] Ceperley, D. M. and Alder, B. J. *Phys. Rev. Lett.* **45**(7), 566 (1980).
- [37] Perdew, J. P. and Zunger, A. *Phys. Rev. B* **23**(10), 5048 (1981).
- [38] Segall, M. D., Lindan, P. J. D., Probert, M. J., Pickard, C. J., Hasnip, P. J., Clark, S. J., and Payne, M. C. *J. Phys.: Condens. Matter* **14**(11), 2717 (2002).