Order-N first-principles calculations with the CONQUEST code

M. J. Gillan [a,1], D. R. Bowler [a] A. S. Torralba [a] T. Miyazaki [b]

[a] *London Centre for Nanotechnology, UCL, Gordon St., London WC1H 0AH, UK*
[b] *National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*

Abstract

We summarize the principles underlying the CONQUEST code for first-principles modelling of systems containing many thousands of atoms. Very recent technical developments implemented in the code are outlined. We give illustrations of physical systems currently being studied with the code, ranging from biologically important molecules to Ge hut clusters on Si (001). Our studies of hut clusters require structural relaxation of systems of over 20,000 atoms using electronically self-consistent density-functional theory.

Key words: first principle; electronic structure; condensed matter

1. Introduction

For many years, density functional theory (DFT) has formed the basis for almost all modelling of large complex systems containing many atoms (1). There are now many standard codes used by hundreds of research groups worldwide to study a vast range of scientific problems in surface physics and chemistry, matter under extreme conditions, nanoscience, biomaterials and the earth sciences, to name only a few important areas. But it was recognised long ago (2; 3) that standard DFT algorithms scale rather poorly with the number of atoms N, the scaling of computer effort being at best $O(N^2)$, deteriorating to $O(N^3)$ for very large N. This is why it is still challenging to apply these methods to systems of more than ~ 1000 atoms, which falls far short of what is needed for some scientific problems. However, Kohn's near-sightedness principle (4) tells us that it should be possible to do much better than this, and that $O(N)$ (often called linear scaling) performance should be achievable. Serious efforts to develop practical $O(N)$ methods for DFT go back

about 15 years (2; 3; 5; 6; 7; 8; 9; 10; 11) (for a review of early work, see Ref. (12)). The feasibility of $O(N)$ scaling was shown 10 years ago in demonstration DFT calculations with our CONQUEST code (13) on systems of several thousand atoms. With feasibility no longer in doubt, the last few years have seen the emergence of several general purpose $O(N)$ DFT codes, which are now efficient enough to apply to real scientific problems (11; 14; 15; 16; 17). After recalling briefly the principles underlying the CONQUEST code (14), we outline some very recent technical developments in the code, and we illustrate some of the scientific problems that it is being used to study.

With DFT, the near-sightedness principle is expressed by the locality of the Kohn-Sham density matrix $\rho(\mathbf{r}, \mathbf{r}')$, meaning that $\rho(\mathbf{r}, \mathbf{r}') \to 0$ as $|\mathbf{r} - \mathbf{r}'| \to \infty$. But the variational principle of DFT can be formulated in terms of the density matrix: the DFT ground state is obtained by minimizing the total energy E_{tot} with respect to $\rho(\mathbf{r}, \mathbf{r}')$, subject to the 'weak idempotency' condition that the eigenvalues of ρ should all lie between 0 and 1. Linear scaling is then obtained by minimizing E_{tot} with respect to ρ, subject to the constraint that $\rho(\mathbf{r}, \mathbf{r}') = 0$ for

---

[1] e-mail: m.gillan@ucl.ac.uk

$|\mathbf{r} - \mathbf{r}'| > r_c$, where r_c is a chosen cut-off distance. These ideas are implemented in CONQUEST, with the additional constraint that $\rho(\mathbf{r}, \mathbf{r}')$ be 'separable' (the number of its non-zero eigenvalues is finite), so that:

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{i\alpha, j\beta} \phi_{i\alpha}(\mathbf{r}) K_{i\alpha, j\beta} \phi_{j\beta}(\mathbf{r}') . \qquad (1)$$

The functions $\phi_{i\alpha}(\mathbf{r})$, referred to as 'support functions', are chosen to be non-zero only within spherical regions of radius R_{reg} centred on the atoms ($\phi_{i\alpha}$ is the αth support function on atom i). We impose the constraint of weak idempotency of ρ by the 'auxiliary density matrix' (ADM) technique (18). In this technique, K is expressed as $K = 3LSL - 2LSLSL$, where L is the ADM, and S is the overlap matrix of support functions: $S_{i\alpha, j\beta} = \langle \phi_{i\alpha} | \phi_{j\beta} \rangle$. To obtain $O(N)$ operation, a spatial cut-off R_L is imposed so that $L_{i\alpha, j\beta} = 0$ for pairs of atoms whose separation exceeds R_L.

An important feature of CONQUEST is that it can be run in several ways. For very accurate calculations, the total energy should be minimized with respect to free variation of the support functions $\phi_{i\alpha}(\mathbf{r})$, and we achieve this by representing the $\phi_{i\alpha}$ in terms of a basis set of cubic splines. However, for more rapid calculations, where lower accuracy suffices, the $\phi_{i\alpha}$ can be represented in terms of a fixed basis set of atomic-like orbitals. For even greater speed, electronic self-consistency can be dispensed with by using the Harris-Foulkes approximation to the DFT total energy (19; 20). The number of support functions $\phi_{i\alpha}$ is generally chosen to be different on different species of atoms. For example, if we model a biomolecule, it sensible to choose the number of support functions to be equal to four on first-row elements such as C, O or N, but to be equal to one on H. CONQUEST provides for this flexibility. In all cases, we can choose either to use the LNV $O(N)$ procedure, or alternatively to diagonalise the Hamiltonian in the basis of $\phi_{i\alpha}$. With diagonalisation, we revert to $O(N^3)$ scaling, but this may still be more efficient for systems of a few hundred atoms or less. Whatever level of precision is chosen, the forces on the atoms are calculated analytically as the exact derivative of the total energy, so that structural relaxation and molecular dynamics can readily be performed. Efficient parallelization of the code has been implemented, following the principles summarized in Refs. (13; 14; 21), and the practical parallel scaling has been shown to be excellent on a variety of platforms. An important issue in achieving good parallel efficiency is the grouping of atoms in compact "bundles" and the of integration-grid points in compact "domains".

2. Recent technical progress

Although the $O(N)$ capabilities of CONQUEST were well established some time ago, the range of systems to which it could be applied was rather limited. Our intention now is to release the code under a GNU General Public License during 2007, and to this end we have enhanced its functionality and ease of use. We have now standardized the pseudopotentials used in the code on the Troullier-Martins form (23). The reason for this choice is that these pseudopotentials are used in a number of plane-wave/pseudopotential codes, such as ABINIT (24). This makes it rather convenient to cross-check CONQUEST results against standard plane-wave calculations. If one chooses to use PAO basis sets in CONQUEST, it is necessary to generate the PAO's using the standard pseudopotentials. The code for doing this has been adapted from the PLATO code (25; 26)

Any practical DFT code needs to be able to use a range of available exchange-correlation functionals. To make this possible, we have recently implemented the PBE form of generalized-gradient approximation (GGA) (27), with PW92 parameterization (28) for the local part. The gradient calculations are done following the scheme of White and Bird (29), which is formally exact on a grid, and involves the computational of only four Fast Fourier Transforms (FFT's). The linearity of the scheme preserves $O(N)$ operation. In order to keep the ability of the code to perform structural relaxation with non-self-consistent Harris-Foulkes calculations, the original computation of forces had to be adapted to the newly implemented GGA functional. As we will report elsewhere (30), we are able to maintain the condition that the forces are exact derivatives of the total energy, and the number of FFT's remains equal to four.

We mentioned above that the division of atoms and grid points into compact groups is important in achieving good parallel efficiency. The way this was done in early versions of the code is described in Ref. (13). However, those methods turned out to be inefficient for problems in which a significant part of the system consists of empty space – a common situation when dealing with surface problems. We have now been able to develop more sophisticated proce-

dures, which significantly improve the efficiency.

In applications of nano-devices, a crucial physical effect is often the transport of electrons, and the exchange of energy between the ionic and electronic sub-systems. These are effects that are not included in conventional first-principles molecular dynamics (m.d.) techniques, which explicitly or implicitly enforce the Born-Oppenheimer approximation, that the electronic sub-system adiabatically follows the motion of the ions. Recently, an important extension of m.d. has been developed, known as "correlated electron-ion dynamics" (CEID) (31), in which the quantum spread of the ions is included *via* a small-amplitude moment expansion. With CEID, it is possible to make direct numerical simulations of, for example, inelastic current-voltage spectroscopy in atomic wires. An ambition for the future is to implement CEID within the CONQUEST code, and we are currently formulating the strategies needed to do this.

3. Scientific applications

In the immediate future, we expect the most important applications of CONQUEST to be in the area of biomolecular systems and nano-systems (there are, of course, close links between the two types of systems). In all cases, it is clearly essential to build up experience with $O(N)$ methods, starting with relatively small systems, where we can cross-check against the results of more standard codes. For nano-systems, we started this learning process with simple tests on semiconductor surfaces, as reported recently (22), and we are now making exploratory calculations on much larger and more complex systems. For biomolecules, we are still at the stage of tests on systems of a few hundred atoms.

In preparation for large-scale calculations on DNA systems, we a currently making extensive tests with CONQUEST on single DNA bases and on DNA base pairs, and comparing with results obtained with other codes, including SIESTA, VASP and GAUSSIAN. As expected, we find excellent agreement for the equilibrium bond lengths of covalently bonded atoms. Results of these tests will be published in the near future (33)

We are also performing tests on the important enzyme dihydrofolate reductase (DHFR), whose function in living organisms is to catalyze the reduction of dihydrofolate to produce tetrahydrofolate. The latter is an important molecule in metabolism. In particular, it is an essential cofactor in one-carbon transfer reactions. As a consequence, DHFR, which is the only enzyme that synthesizes it, has receive much attention, for example as a target for anti-malarial drugs. Although the specific substrate for DHFR is dihydrofolate (DHF), in some species the enzyme also catalyzes, very inefficiently and less specifically, the reduction of folate, a precursor of DHF.

The reasons why DHFR is specific for DHF remain unclear. LDA DFT calculations of the active site suggested that enzyme-induced polarization of the substrates may be a cause for the preference, at least in the *Escherichia coli* enzyme. Indeed, one study (34) found large electron density differences (EDD) between the density of DHF when bound to the enzyme with respect to that in vacuum. However, results from MP2 calculations, although qualitatively supportive for a role of polarization, are less conclusive (35; 36).

All existing studies used a point-charge model for the bulk protein, restricting the quantum mechanical (QM) calculations to a few atoms at the active site. Hence, the quantitative discrepances between different studies may be due to that limitation of the models, rather than to the different QM methods employed. Since DHFR is a relatively small protein (159 amino acids in *Escherichia coli*, or about 3000 atoms), we decided to assess such posibility by using Conquest to perform LDA DFT calculations in extended models of the active site, with the ultimate goal of including the whole of the protein. Thus, we did not model bulk protein in any way, since its effect was expected to become obvious as the size of the model increased.

Our preliminary results on portions of the protein of up to 300 atoms show that indeed larger models are quantitatively closer to MP2 results than to the original LDA calculations. We found larger polarization on DHF than on folate, and only DHF displayed polarization on the bond susceptible of hydrogenation, consistent with the observed specificity (see Fig. 1). Furthermore, calculations on different conformations of the protein agree with experimental evidence regarding the mechanism. In particular, the presence of some amino acids of the so-called Met20 loop seems to be essential for catalysis, as represented by polarization on the hydrogenable bond of the substrate. Moreover, those amino acids must be occluding the active site for polarization to be observed, as expected in the proposed mechanism (37).
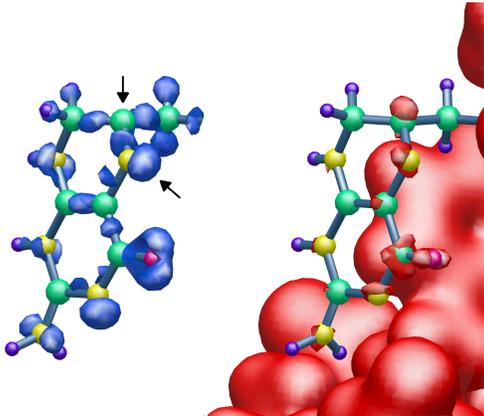
Fig. 1. Electronic density difference plot for dihydrofolate (DHF) upon binding to the enzyme dihydrofolate reductase (±0.001 electron/bohr). Charge deficiency (respect to the density of DHF in vacuum) is shown in blue; charge excess is red. The enzyme induces polarization on N5 and C6 atoms (marked with arrows), and electronic density withdrawal from the bond linking them. These effects are consistent with the catalyzed reaction, namely, protonation of N and hydride transfer to the bond, and are much weaker for the very inefficient reduction of folate, a secondary substrate (not shown).

Turning now to the application of CONQUEST to nano-systems, we summarise our recent progress in investigating the three-dimensional (3D) structures formed when Ge is deposited on the Si (001) surface. The Ge/Si (001) has been extensively studied, because it is a prototypical example of hetero-epitaxial Stranski-Krastanov growth. When Ge atoms are deposited on Si (001), growth initially occurs layer by layer, up to a critical thickness of about three monolayers (ML). Strain due to the lattice mismatch is relieved by the formation of regularly spaced rows of dimer vacancies in the two-dimensional (2D) structure, resulting in the $2 \times N$ structure. Deposition of further Ge leads to another strain-relief structure, 3D pyramid-like structures known as "hut clusters" (38). Recently, we have studied this transition from 2D to 3D structures, using CONQUEST.

Usually, the stability of 3D structures is governed by (i) the lowering of strain energy in the clusters and the underlying substrate, and (ii) the energy increase arising from the formation of facets. Theoretical approaches used so far have used continuum elasticity theory to described the strain energy, with DFT being used only for the surface energies (39; 40). For the Ge/Si system, the four facets of the hut cluster are well established to be {105} surfaces, and the structure of these surfaces has recently been clarified by DFT calculations (41; 42). Note that the typ-

ical side-length of hut clusters is about 150 Å, and deposition of additional Ge leads to the formation of other 3D structures called "domes", having steeper facets. Interestingly and importantly, the DFT calculations show that the strained Ge (105) surface is more stable han strained Ge (001). This means that the surface energy may actually *stabilise* the structure. If the surface contribution to the overall energy is small, contributions from the edges where the facets meet each other and the wetting layer may also affect the stability of the 3D structure. In addition, as the area of the facets of the experimentally observed Ge hut cluster is not large, the evaluation of the surface part itself is doubtful. For these reasons, the validity of previous theoretical approaches is uncertain, especially for small hut clusters. To overcome these problems, we are using CONQUEST to model the entire hut cluster, together with the wetting layer and the Si substrate.

In preparation for CONQUEST calculations on the full system, we first performed DFT calculations on the Ge (105) surface, including test calculations also on the unstrained and strained Ge bulk. Since the size of this system is relatively small, we can employ diagonalisation in this case. We have clarified the accuracy of the various DFT methods explained above for the unstrained and strained Ge systems. We have also confirmed that full DFT calculations performed with CONQUEST using cubic-spline basis sets are accurate enough for the study of the strained Ge (105) surface. The conditions need for $O(N)$ calculations to achieve good accuracy for this system have also been established.

Using these results, we have performed $O(N)$ DFT calculations on the entire Ge/Si (001) hut clusters. At the non-self-consistent level, we have performed structual optimisation on systems of different sizes. The largest system treated so far, shown in Fig. 2, contains ~ 23000 atoms, and we found that structure optimisation is robust even for such large systems. We have examined three structural models of the Ge hut cluster having different facet or edge structures, and we have compared their energies with those of the $2 \times N$ reconstructions with $N = 4, 6$ and 8. The results, to be reported in detail elsewhere (43; 44), show that the 2D structure is more stable for small coverages of Ge atoms, but the 3D hut structure becomes more stable when the coverage exceeds 2.6 monolayers, in agreement with experimental observation.
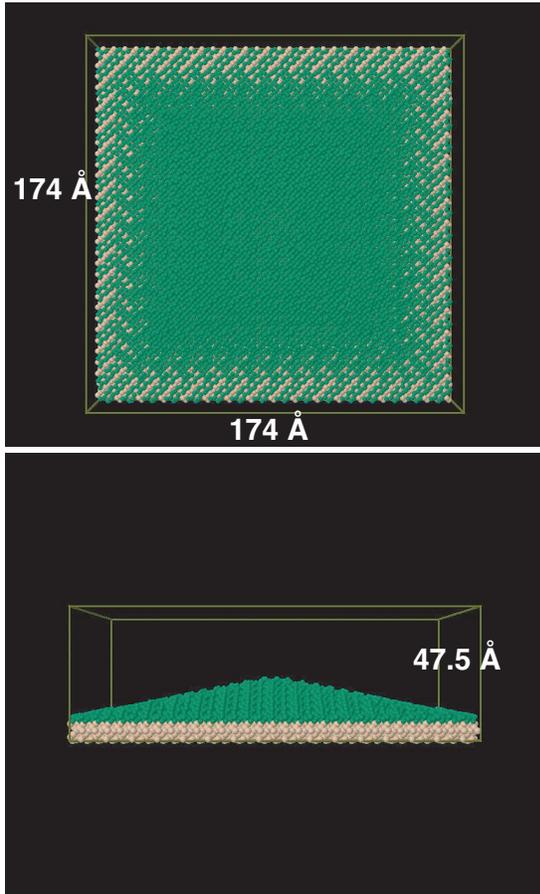
Fig. 2. Atomic geometry of largest Ge/Si (001) hut cluster used for structural relaxation with CONQUEST DFT calculations. Upper and lower panels shows plan and side views, respectively. Pink and green spheres represent Si and Ge atoms. Dimensions of periodically repeated cell in surface plane and normal to surface are marked.

4. Summary

The basic principles underlying contemporary $O(N)$ DFT were established over 10 years ago. However, the realisation of these principles in practical codes has required the solution of a large number of technical problems concerning basis sets, the enforcement of linear scaling in the calculation of the ground-state density matrix, etc. Some of these problems admit of more than one solution, and the codes that have appeared so far, including CONQUEST (14), SIESTA (15), ONETEP (16), and OPEN-MX (17), differ in important ways. We have tried to show here how the CONQUEST code has now passed beyond the stage of feasibility studies, and can now be applied to real scientific problems concerning biomolecular and nano-scale systems.

Comparisons with the results of standard codes for relatively small systems of a few hundred atoms are demonstrating the reliability of the methods. At the same time, it is clear that structural relaxation at different levels of precision, using both self-consistent and non-self-consistent calculations, is becoming a practical proposition for systems containing more than 20,000 atoms.

References

[1] R. M. Martin, *Electronic Structure* (Cambridge University Press, Cambridge, 2004).

[2] S. Baroni and P. Giannozzi, *Europhys. Lett.*, **17** (1991) 547

[3] G. Galli and M. Parrinello, *Phys. Rev. Lett.*, **69** (1992) 3547

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

[5] F. Mauri, G. Galli and R. Car, *Phys. Rev. B*, **47** (1993) 9973

[6] P. Ordejón, D. A. Drabold, M. P. Grumbach and R. M. Martin, *Phys. Rev. B*, **48** (1993) 14646

[7] P. Ordejón, D. A. Drabold, R. M. Martin and M. P. Grumbach, *Phys. Rev. B*, **51** (1995) 1456

[8] J. Kim, F. Mauri and G. Galli, *Phys. Rev. B*, **52** (1995) 1640

[9] E. Hernández and M. J. Gillan, *Phys. Rev. B*, **51** (1995) 10157

[10] E. Hernández and M. J. Gillan, *Phys. Rev. B*, **53** (1996) 7147

[11] P. Ordejón, E. Artacho and J. M. Soler, *Phys. Rev. B*, **53**, 10441 (1996)

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

[13] C. M. Goringe, E. Hernández, M. J. Gillan and I. J. Bush, *Comput. Phys. Commun.*, **102** (1997) 1

[14] D. R. Bowler, T. Miyazaki and M. J. Gillan, *J. Phys: Condens. Matter*, **14** (2002) 2781

[15] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón and D. Sánchez-Portal, *J. Phys. Condens. Matter*, **14** (2002) 2745

[16] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *J. Chem. Phys.*, **122** (2005) 084119

[17] T. Ozaki and H. Kino, *Phys. Rev. B*, **72** (2005) 045121

[18] X.-P. Li, R. W. Nunes and D. Vanderbilt, *Phys. Rev. B*, **47** (1993) 10891

[19] J. Harris, *Phys. Rev. B*, **31** (1985) 1770

[20] W. M. C. Foulkes and R. Haydock, *Phys. Rev. B*, **39** (1989) 12520

[21] D. R. Bowler, R. Choudhury, M. J. Gillan and T. Miyazaki, phys. stat. sol.(b), **243** (2006) 989

[22] T. Miyazaki, D. R. Bowler, R. Choudhury and M. J. Gillan, *J. Chem. Phys.*, **121** (2004) 6186

[23] N. Troullier and J. L. Martins, *Phys. Rev. B*, **43** (1991) 1993

[24] X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah. F. Jollet, M. Torrent, A. Roy, M. Mikami, P. Ghosez, J.-Y. Raty and D. C. Allan, *Comput. Mater. Sci.*, **25** (2002) 478

[25] A. P. Horsfield, *Phys. Rev. B*, **56** (1997) 6594

[26] S. D. Kenny, A. P. Horsfield and H. Fujitani, *Phys. Rev. B*, **62** (2000) 4899

[27] J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, **77** (1996) 3865

[28] J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).

[29] J. A. White and D. M. Bird, *Phys. Rev. B*, **50** (1994) 4954

[30] A. S. Torralba, D. R. Bowler and M. J. Gillan, in preparation

[31] A. P. Horsfield, D. R. Bowler, A. J. Fisher, T. N. Todorov and C. G. Sanchez, *J. Phys. Condens. Matter*, **17** (2005) 4793

[32] G. Kresse and J. Furthmüller, *Phys. Rev. B*, **54** (1996) 1169

[33] T. Ohtsuka, T. Miyazaki, T. Ohno, D. R. Bowler and M. J. Gillan, in preparation.

[34] J. Bajorath, J. Kraut, Z. Li, D. H. Kitson, and A. T. Hagler, *Proc. Natl. Acad. Sci. USA* **88**, 6423 (1991).

[35] S. P. Greatbanks, J. E. Gready, A. C. Limaye, and A. P. Rendell, *Proteins* **37**, 157 (1999).

[36] M. Garcia-Vicola, D. G. Truhlar, and J. Gao, *J. Mol. Biol.* **327**, 549 (2003).

[37] D. H. J. Schnell, Jason R. and P. E. Wright, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 119 (2004).

[38] Y. W. Mo, D. E. Savage, B. G. Schwartzentruber and M. G. Lagally, *Phys. Rev. Lett.*, **65** (1990) 1020

[39] O. E. Shklyaev, M. J. Beck, M. Asta, M. J. Miksis and P. W. Voorhees, *Phys. Rev. Lett.*, **94** (2005) 176102

[40] G.-H. Lu and F. Liu, *Phys. Rev. Lett.*, **94** (2005) 176103

[41] Y. Fujikawa, K. Akiyama, T. Nagao, T. Sakurai, M. G. Lagally, T. Hashimoto, Y. Morikawa and K. Terakura, *Phys. Rev. Lett.*, **88** (2002) 176101

[42] P. Raiteri, D. B. Migas, L. Miglio, A. Rastelli and H. von Känel, *Phys. Rev. Lett.*, **88** (2002) 256103

[43] T. Miyazaki, D. R. Bowler, R. Choudhury and M. J. Gillan, in preparation.

[44] T. Miyazaki, D. R. Bowler, M. J. Gillan and T. Ohno, in preparation.