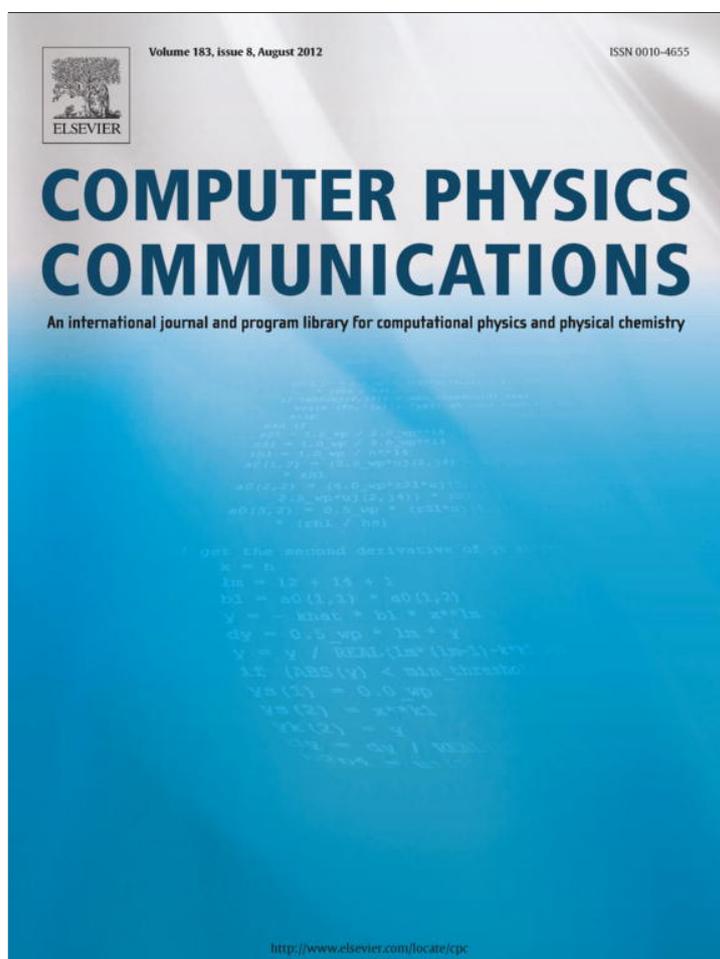


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Introducing k -point parallelism into VASP

Asimina Maniopoulou^a, Erlend R.M. Davidson^{b,c}, Ricardo Grau-Crespo^c, Aron Walsh^d, Ian J. Bush^a, C. Richard A. Catlow^c, Scott M. Woodley^{c,*}

^a Numerical Algorithms Group Ltd, Wilkinson House, Jordan Hill Road, Oxford OX2 8DR, UK

^b London Centre for Nanotechnology, University College London, London WC1E 6BT, UK

^c University College London, Department of Chemistry, 20 Gordon Street, London WC1H 0AJ, UK

^d Centre for Sustainable Chemical Technologies and Department of Chemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

ARTICLE INFO

Article history:

Received 15 November 2011

Accepted 15 March 2012

Available online 17 March 2012

Keywords:

Parallelization

k -Points

Plane waves

DFT

Methods of electronic structure calculations

ABSTRACT

For many years *ab initio* electronic structure calculations based upon density functional theory have been one of the main application areas in high performance computing (HPC). Typically, the Kohn–Sham equations are solved by minimisation of the total energy functional, using a plane wave basis set for valence electrons and pseudopotentials to obviate the representation of core states. One of the best known and widely used software for performing this type of calculation is the Vienna *Ab initio* Simulation Package, VASP, which currently offers a parallelisation strategy based on the distribution of bands and plane wave coefficients over the machine processors. We report here an improved parallelisation strategy that also distributes the k -point sampling workload over different processors, allowing much better scalability for massively parallel computers. As a result, some difficult problems requiring large k -point sampling become tractable in current computing facilities. We showcase three important applications: dielectric function of epitaxially strained indium oxide, solution energies of tetravalent dopants in metallic VO₂, and hydrogen on graphene.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

For many years *ab initio* electronic structure calculations have been one of the main application areas in high performance computing (HPC). Whilst the methods used for those calculations have changed, the approach pioneered by Car and Parrinello in 1985 [1] has been one of the most common to be employed. The approach is based upon density functional theory [2]; the Kohn–Sham [3] equations are solved within a plane wave basis set by minimisation of the total energy functional, with the use of pseudopotentials [4,5] to obviate the representation of core states. A review of the method can be found in Ref. [6]. Such is the importance of these methods that over 25% of all the cycles used on the phase 2b component of HECToR [7], the UK's high-end computing resource, in the period from December 2010–July 2011 were for packages performing total energy pseudopotential calculations.

One of the best known and widely used packages for performing this type of calculation is VASP [8–11], the Vienna *Ab initio* Simulation Package. Indeed on HECToR it is the most extensively used package of all; thus maximising its performance is vital for researchers using this, and related, machines. In this paper we will briefly describe our recent work on improving the parallel scalability of the code for certain classes of common prob-

lems and show three important applications that exploit this new scalability: dielectric function of epitaxially strained indium oxide, solution energies of tetravalent dopants in metallic VO₂, and hydrogen on graphene. We have achieved substantial improvements by introducing a new level of parallelism based upon the use of k -point sampling within VASP. Whilst this is common in similar codes, VASP does not currently support it, and we will show that through its use, the scalability of total energy calculations can be markedly improved; which is a particularly important as often the total energy calculation needs to be performed many times during geometry optimisation of the atomic and electronic structure. Typically the total size of the system under study is limited by time constraints, and so parallel scaling of the calculation on moderate sized systems must be advantageous if many cores are to be exploited efficiently.

The remainder of the paper is organised as follows: first the current parallelisation is briefly described; then we introduce our parallelisation strategy over k -points; and finally the initial performance tests followed by three case applications are presented.

2. Methodology: the parallelisation of VASP

Details of VASP parallelisation are covered in some detail elsewhere [12], and here we shall only cover those that are relevant to the current work.

* Corresponding author.

E-mail address: scott.woodley@ucl.ac.uk (S.M. Woodley).

VASP 5 offers parallelisation (and data distribution) over bands and over plane wave coefficients; both may be used together. How this division occurs is controlled by the NPAR tag in one of the required VASP input files (namely, the INCAR file). In particular if there are a total of NPROC cores in the job, each band will be distributed over NPROC/NPAR cores, where NPAR is defined (chosen) by the user. Thus if $NPAR = 1$ all of the bands will be distributed over all processors, while if $NPAR = NPROC$ the coefficients for a given band are all associated with one core.

NPAR reflects a tensioning between conflicting requirements. If bands are distributed across all processors then the communication costs for the parallel three-dimensional Fast Fourier Transforms (FFTs) required by the plane wave pseudopotential method are high, but the cost for linear algebra operations required by the method, such as orthogonalisation and diagonalisation, are relatively low. On the other hand if $NPAR = 1$ the communication cost for the FFTs is non-existent but it is high for the linear algebra operations. Thus NPAR must be chosen carefully to obtain the best performance possible, and importantly it will depend upon the chemical system being studied, the hardware upon which the run is being performed, and the number of cores being used (amongst other possibilities).

Thus the use of NPAR allows a run to either stress the parallel FFT or parallel diagonalisation. Unfortunately neither of these operations scale well on distributed memory parallel architectures [13,14], especially for the moderate size grids and matrices used in many VASP runs.

Parallelisation over bands and over plane waves are not the only possible ways that *ab initio* electronic structure codes can exploit modern HPC resources. Other software can also exploit parallelisation over k -points; examples include CASTEP [15,16] and CRYSTAL [17,18], k -points arise from the translational symmetry of the systems being studied [19]. This symmetry also results in many, but not all, operations at a given k -point being independent from those at another k -point. This naturally allows another level of parallelism, and it has been shown that exploitation of k -point parallelism can greatly increase the scalability – see Ref. [20] for a recent example. More generally, this use of hierarchical parallelism is one of the more common methods that aid scaling to very large numbers of cores [21].

The standard release of VASP is yet, however, to exploit this possibility. Therefore we have modified the code from VASP 5.2.2 to add this extra level of parallelisation. The code is organised so that the cores may be split into a number of groups, and each of these groups performs calculations on a subset of the k -points. The number of such groups is specified by the new KPAR tag, which is set in the INCAR file. Thus if the run uses 10 k -points and KPAR is set to 2 there will be 2 k -point groups each performing calculations on 5 k -points. Similarly if KPAR is set to 5 there will be 5 groups each with 2 k -points. It can therefore be seen that KPAR has an analogous rôle to NPAR mentioned above, except that it applies to k -point parallelism. Currently KPAR is limited to values that divide exactly both the total number of k -points and the total number of cores used by the job. It should be noted that NPAR is also subject to the latter restriction.

The introduction of another level of parallelism through use of the k -points does potentially greatly increase the scalability of the code. However, it should not be viewed as a panacea. Not all operations involve k -points, and thus Amdahl's law [22] effects will place a limit on the scalability that can be achieved. Furthermore, some quantities are not perfectly parallel across k -points; evaluation of the Fermi level being an obvious example. And it should be noted that introduction of k -point parallelism does introduce some extra communication and synchronisation.

Probably the biggest limitation on the use of k -point parallelism is that due to system size. On increasing the size of the unit cell

Table 1

Run times for computing the PBE energy for Na_3AlH_6 , with a total of 60 atoms and 10 k -points (first test system), using the implementation of k -point parallelisation in VASP 5.2.2.

KPAR	Cores	Time (s)	Speed-up
1	64	293	64
2	128	159	118.29
5	320	75	249.61
10	640	47	399.23

used by the calculation, fewer k -points are required in order to converge the calculation (of energy or structure) to a given precision; and in the limit only a single k -point may be sufficient to accurately represent the system. This limits what can be achieved by k -point parallelism; but there are many examples where more than one k -point is required [23–26]. In these cases, parallelisation over k -points is a useful technique as it is more efficient to use the smallest unit cell and investigate different k -points, rather than use a supercell. This is especially true when the total energy calculation is only one part of a larger calculation, for instance in a geometry optimisation.

3. Results: initial scaling performance

A number of smaller systems were used during the initial testing of the scalability on the UK national HPC resource, HECToR, before and after implementing the k -point parallelisation. Here we report on three example cases:

1. A supercell of Na_3AlH_6 , with a total of 60 atoms. Each run uses 10 k -points and the PBE exchange-correlation functional [27,28].
2. A unit cell of Litharge (α -PbO); each has only 4 atoms. Here 108 k -points are used, generated by the Monkhorst–Pack method, and again the PBE functional is employed.
3. A unit cell of Litharge (α -PbO), again with only 4 atoms, but this time using 24 k -points and a hybrid functional containing an exact exchange contribution [29], and where phonons are calculated.

All runs have been performed on the phase 2b component of the HECToR system: a large Cray XE6 system managed by UK's national supercomputing service HPCx. The nodes are based upon AMD Magny-Cours processors, and contain 24 cores each clocking at 2.1 GHz. There is 32 Gbytes of memory associated with each node, and inter-node communication is via Cray's Gemini network. More details may be found at the HECToR web site [30]. To ensure a fair comparison of the new k -point parallel code with the original for each run, an optimal value of NPAR has been found. All times reported are total run times, i.e. not just the time for the energy minimisation.

The run times for computing the total energy, based on the PBE functional, for Na_3AlH_6 , are shown in Table 1. This first test system contains 60 atoms in the unit cell and 10 k -points. In these runs the extra parallelism due to k points is exploited over and above a base run on 64 cores of the original code, and speed-ups are measured relative to this run where we assume perfect scaling. The optimal value of NPAR for the all runs was found to be 4. It can be seen that the speed-up obtained for this test case is very good, and remarkably better than that previously obtained – see Fig. 1. Without exploiting the extra level of parallelism over k -points, the optimal values of NPAR were found to be 4, 8, 32 and 32 for 64, 128, 320 and 640 cores, respectively.

Similar behaviour is found for our second test system, computing the PBE total energy for the four-atom unit cell of α -PbO, see Table 2 and Fig. 2. A larger number of 108 k -points facilitate

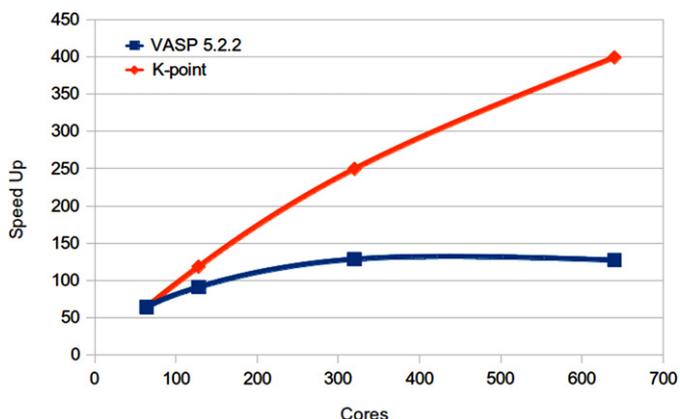


Fig. 1. Speed-up obtained on HECToR before (blue squares) and after (red diamonds) implementation of *k*-point parallelisation within VASP 5.2.2 for computing the PBE energy for Na₃AlH₆, with a 60-atom unit cell and 10 *k*-points.

Table 2

Run times for computing the PBE total energy for α -PbO with a unit cell of 4 atoms and a Monkhorst–Pack set of 108 *k*-points (second test system) and using the implementation of *k*-point parallelisation in VASP 5.2.2.

KPAR	Cores	Time (s)	Speed-up
1	144	2141	144
2	288	1098	280.82
3	432	766	402.42
4	576	608	507.38
6	865	410	752.00
9	1296	279	1106.78
12	1728	216	1429.07
18	2592	150	2053.15
27	3888	106	2920.85
36	5184	88	3521.04
54	7776	65	4719.49
108	15,552	41	7510.26

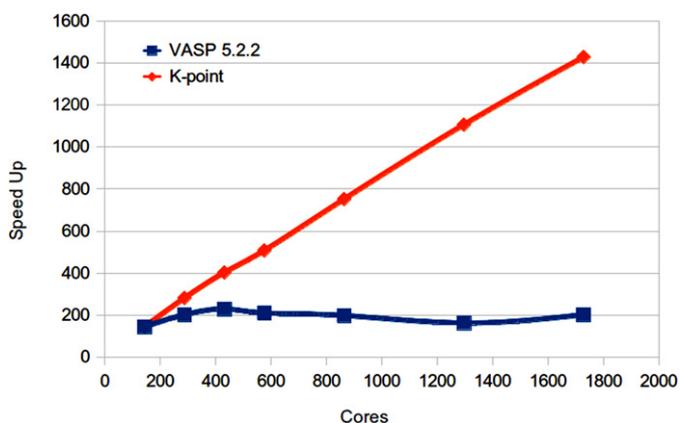


Fig. 2. Speed-up obtained on HECToR before (blue squares) and after (red diamonds) implementation of *k*-point parallelisation within VASP 5.2.2 for computing the PBE energy for α -PbO, with a 4-atom unit cell and 108 *k*-points.

the exploitation of *k*-point parallelism. In fact excellent scaling is observed up to around 4000 cores, and even above that notable increases in speed are seen all the way up to 15,552 cores, whereas without the level of parallelism over *k*-points there is no significant reduction in CPU time when using more cores. Note that the optimal values of NPAR for the latter runs were found to be 18, 18, 18, 36, 36 and 36 for 144, 288, 432, 576, 864 and 1728 cores, respectively.

If the functional used includes a non-local Hartree–Fock exact exchange term, which in itself is expensive to compute, it also couples together different *k*-points. Thus, additional communications

Table 3

Run times for computing the total energy using hybrid functional for the α -PbO system with a unit cell of 4 atoms and a Monkhorst–Pack set of 108 *k*-points (third test system) and using the implementation of *k*-point parallelisation in VASP 5.2.2.

KPAR	Cores	Time (s)	Speed-up
1	256	14,674	256
2	512	7579	495.69
3	768	4979	764.44
4	1024	3790	991.19
6	1536	2552	1471.69
8	2048	1944	1932.04
12	3072	1400	2683.73
24	6144	1053	3567.13

between sub-groups of cores is required. However, as can be seen from Table 3, the scaling is still extremely good up to 3000 cores. Comparison of the scaling of our *k*-point implementation with the original version of the code was, unfortunately, not possible as we were unable to run the original on more than 256 cores. However it is unlikely that its scaling will be as good as that seen for the *k*-point code, and further illustrates how the implementation of *k*-point parallelism allows the efficient exploitation of many more cores.

4. Results

4.1. Dielectric function of epitaxially strained indium oxide

The calculation of optical properties is one of the most slowly converging with respect to *k*-point sampling in the Brillouin zone [31,32], which is a particular issue for semiconducting materials with low carrier effective masses (high band dispersion), where valence to conduction band separations vary greatly across *k*-space. Therefore, ensuring that the calculated optical properties are converged with respect to *k*-points is essential. Here, we are concerned with indium oxide, which is primarily used as a transparent conducting oxide in optoelectronic devices. Through a combination of computational and experimental studies, we have developed a deep understanding of the materials structural [33], electronic [34] and optical [35] properties. Of recent interest has been the change in optical response for ultra-thin epitaxially strained (111) In₂O₃ grown on a Y-stabilised ZrO₂ substrate [36]. We have modelled the system using a (111) oriented supercell, which is strained in the *ab* plane (120-atom unit cell, 1056 valence electrons, 1200 bands). The dielectric function is calculated within the independent-particle approximation (based on the longitudinal approximation [31,37]) using the PBE exchange–correlation functional [27]. We can assess the convergence through the change in the high-frequency dielectric constant (ϵ_∞) with respect to *k*-point sampling. On the phase 2b HECToR system, using a standard compilation of VASP 5.2, the maximum-point density possible within the queue limit of twelve hours is $2 \times 2 \times 2$ using 96 cores, while using *k*-point parallelism this can be increased to $6 \times 6 \times 6$ and scale up to 1536 cores. As shown in Table 4, sampling at the gamma point only results in a very poor result ($\epsilon_\infty = 8.58$), and the value converges for denser *k*-point meshes towards a value of $\epsilon_\infty = 5.78$ (in the hexagonal *ab* plane). The anisotropy is also slow to converge: for coarser *k*-point meshes, the in-plane value is smaller than the out-of-plane, which is reversed for the more dense meshes. This case demonstrates very well the utility of *k*-point parallelism in VASP for calculating the optical properties of solids.

4.2. Solution energies of tetravalent dopants in metallic VO₂

Vanadium oxide VO₂ undergoes a transition from a monoclinic structure to a rutile-like tetragonal structure when heated above

Table 4

Calculated high-frequency dielectric constant of epitaxially strained indium oxide (PBE functional) as a function of k -point sampling of the first Brillouin zone.

High-frequency dielectric constant	k -point grid (Gamma centred)	Irreducible k -point grids
8.576 ($\epsilon_{\infty}^{\perp}$), 8.729 ($\epsilon_{\infty}^{\parallel}$)	1 × 1 × 1	1
5.927 ($\epsilon_{\infty}^{\perp}$), 5.996 ($\epsilon_{\infty}^{\parallel}$)	2 × 2 × 2	8
5.782 ($\epsilon_{\infty}^{\perp}$), 5.755 ($\epsilon_{\infty}^{\parallel}$)	4 × 4 × 4	36
5.781 ($\epsilon_{\infty}^{\perp}$), 5.743 ($\epsilon_{\infty}^{\parallel}$)	6 × 6 × 6	112

~ 340 K [38,39]. While the low temperature phase is a semiconductor, the high temperature phase is metallic, which is the basis for the potential application of VO₂ as a smart window coating: in hot weather, the coating is metallic and will reflect most of the infrared radiation, keeping the interior cool, while still letting the visible light pass through; conversely, in cold weather the semi-conducting film will allow the heat-carrying infrared radiation to enter the building, reducing the need for heating. The transition temperature of pure VO₂ is clearly too high for room temperature regulation, but it can be decreased by doping with certain elements, e.g. tungsten [40–42]. There is also considerable interest in finding other stable dopants, which besides switching the transition temperature, are simultaneously capable of improving the colour and optical behaviour of the VO₂ films [43,44].

We consider here the stabilities of different tetravalent dopants (M = Ti, Zr and Hf) in rutile VO₂ in the dilute limit, with respect to the formation of a separate oxide phase MO₂, which can be assessed by evaluating the solution energies:

$$W = E[V_{n-1}MO_{2n}] - \frac{n-1}{n}E[V_nO_{2n}] - \frac{1}{m}E[M_mO_{2m}]$$

where n and m represent the number of formula units in the simulation cells for VO₂ and MO₂, respectively, and the total (free electronic) energies E are calculated at the DFT level with periodic boundary conditions, using VASP and the gradient-corrected PBE functional [27]. In order to avoid interactions of the dopant ion with its periodic images, a relatively large 2 × 2 × 3 supercell was chosen for the simulation of pure and doped VO₂. This system contains 24 formula units and has a minimum dopant–dopant separation of about 10 Å along the crystal axes. Furthermore, since this phase is metallic, a dense k -point mesh is required for integrations in the reciprocal space. Using the k -point parallelism in VASP, we tested 4 × 4 × 4 (18 irreducible k -points) and 8 × 8 × 8 (75 irreducible k -points) meshes for pure VO₂. The latter calculation was performed over 360 processors on the HECToR phase 2b system, with a speed-up factor of approximately 13 compared to the standard version running on 24 processors (ideal scaling would correspond to a speed-up factor of 15). Since we found a difference of less than 1 meV in the total energy of the supercell calculated using the two different meshes, the lower density (4 × 4 × 4) mesh was employed for the doped cells. The reduction in symmetry introduced by the dopant increases the number of irreducible k -points from 18 to 21, and therefore the calculations of the doped cell were performed on 168 processors, with KPAR = 7 groups (of 3 k -points each) running in parallel, on 24 processors each. All geometries were optimised until the forces on the ions were all less than 0.01 eV/Å.

The calculated solution energies of Ti, Zr and Hf in VO₂, with respect to tetragonal (rutile) TiO₂ and monoclinic (baddeleyite) ZrO₂ and HfO₂ are shown in Table 5. We also show the relaxed inter-atomic distances within the first coordination sphere of the dopants. Of the three dopants considered, Ti⁴⁺ is the one that incorporates more easily in the VO₂ structure, which is not surprising considering that it has similar radius (0.60 Å) to V⁴⁺ (0.58 Å) and that it forms an oxide with the same structure. The reversal in the trend down the group between Zr⁴⁺ and Hf⁴⁺ is explained

Table 5

Solution energies and local geometry of tetravalent dopants in rutile VO₂ – for comparison, the equilibrium V–O distances in pure VO₂ are 2 × 1.95 Å and 4 × 1.91 Å.

Dopant	W (eV)	$d(M-O)$ (Å)
Ti ⁴⁺	0.28	2 × 2.01, 4 × 1.95
Zr ⁴⁺	1.23	2 × 2.10, 4 × 2.07
Hf ⁴⁺	1.03	2 × 2.08, 4 × 2.04

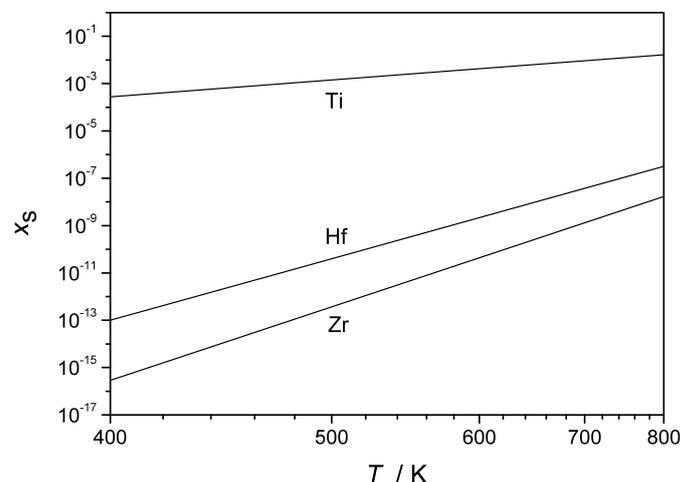


Fig. 3. Equilibrium solubility limits of M = Ti, Hf and Zr in V_{1-x}M_xO₂ with respect to separation into MO₂ phases.

by the lanthanide contraction effect [45], which leads to Hf–O distances shorter than expected. The calculated solution energies can be employed to estimate the equilibrium solubility limits of the dopants [46,47], that is, the maximum x in V_{1-x}M_xO₂ that is thermodynamically stable with respect to MO₂ phase separation, as $x_s = \exp(-W/k_B T)$, where k_B is Boltzmann's constant and T is the temperature (Fig. 3). Of course, it is possible experimentally to dope beyond the equilibrium limit using metastable synthesis routes. However, the further from the equilibrium the more difficult it is in practice to achieve the desired dopant concentrations, and therefore it can be expected that, while doping VO₂ with Ti is relatively easy (Ti-doped VO₂ has been studied experimentally by several authors, for example, see Ref. [48]), Hf and Zr doping will be more problematic due to the formation of competing oxide phases. This factor should be taken into account when considering the use of these dopants as a route to influence the luminous transmittance of VO₂ in the visible wavelength range [49].

4.3. Scaling of hydrogen on graphene with parallel k -points

Graphene has generated a huge interest from the materials science community; for example, each year there are several thousand papers mentioning graphene in the abstract. As a material, graphene is very strong (some 100 times stronger than steel) [50]. Graphene is also a semi-metal (a semiconductor with zero band gap) due to the touching of the conduction and valence bands at the Fermi level. To be useful in semiconductor electronics, a band gap is required, which can be created by chemical doping. Adsorption of hydrogen opens the band gap and, in particular, graphane, which is fully hydrogenated graphene, has been shown theoretically and experimentally to have a finite band gap. Hydrogen at graphene is also used as an important model system for formation of molecular hydrogen in the interstellar medium of space.

DFT is frequently used to study the physical and electronic properties of graphene. However, the presence of a complex electronic structure and a Dirac point at the Fermi level mean that accurate calculations frequently require the sampling of large

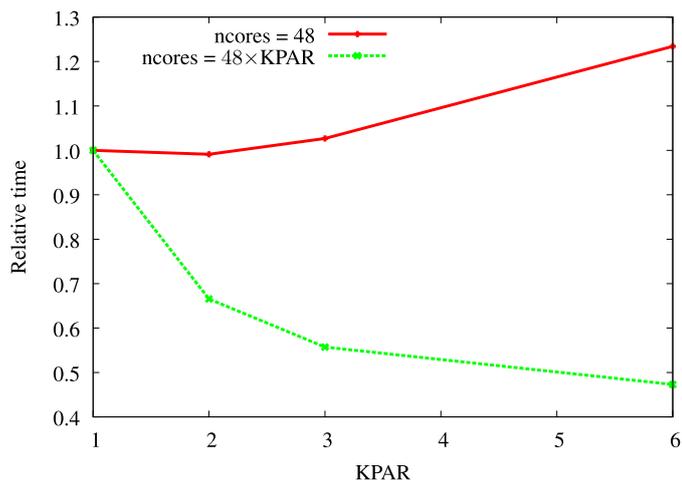


Fig. 4. The effect of KPAR on wall-clock time for a $p(3 \times 3)$ sheet of graphene, where the relative time is (wall-clock time)/(wall-clock time with KPAR = 1); KPAR has values of 1 (no parallelism over k -points), 2, 3 and 6 and where a total of 6 k -points was used in the calculations. The red (solid) line is for the number of MPI processes kept constant, and KPAR is changed. The green (dashed) line is for the number of MPI processes scaled with the value of KPAR. In both cases the NPAR parameter is set to 2.

number of k -points in the Brillouin zone. Furthermore, a large (usually 15 Å) vacuum, required in plane wave codes to isolate the periodic images of graphene in the direction normal to the surface plane, makes calculations relatively expensive for larger supercells. By conducting DFT calculations with varying cell sizes and k -points of a hydrogen adatom chemisorbed on graphene, Hammer et al. [51] found that the supercell size can be kept small as long as enough k -points are used.

We use a $p(3 \times 3)$ cell which contains 18 carbon atoms and a vacuum in the z -direction of 15 Å. Sampling of reciprocal space was done using a $5 \times 5 \times 1$ Monkhorst–Pack grid which is considered well converged, since it gives a binding energy for a chemisorbed hydrogen atom within 20 meV of the $p(7 \times 7)$ cell [51]. We also used a $13 \times 13 \times 1$ k -points mesh, which gives a binding energy within 10 meV of the $p(7 \times 7)$ cell. This results in 5 irreducible k -points for the $5 \times 5 \times 1$ mesh, and 21 for the $13 \times 13 \times 1$ mesh; however to allow us to consider more values of KPAR we add in an extra k -point in each case. We made use of the PBE exchange–correlation functional throughout [27]. Projector augmented wave (PAW) potentials were used [52], and the wavefunctions of the valence electrons were expanded in terms of a plane wave basis set. We used 48 cores with an NPAR of 2, which we found to be the maximum number of parallel processes we could use while still achieving acceptable scaling. We will report on two types of efficiency for this system: the effect on scaling when using the k -points parallelism without increasing the number of CPUs, and the scaling when the number of CPUs is increased in line with the number of k -point parallel groups (KPAR).

The relative times (i.e. the wall-clock time relative to the time for KPAR = 1) for both cases are shown in Figs. 4 and 5. When the number of cores is kept constant (solid red line) at KPAR of 2 the wall-clock time decreases slightly; however larger values decrease the parallel efficiency. It should be noted that changing KPAR, but keeping the number of cores constant will probably require that NPAR is changed to achieve optimal results, which, however, goes beyond the scope of the current work. When the number of cores is scaled relative to KPAR (dashed green lines of Figs. 4 and 5) the wall-clock time is considerably reduced. At KPAR = 2 scaling is very good; however, at higher KPAR values the scaling efficiency is poor, but the time-to-science wall-clock time of the code is still reduced.

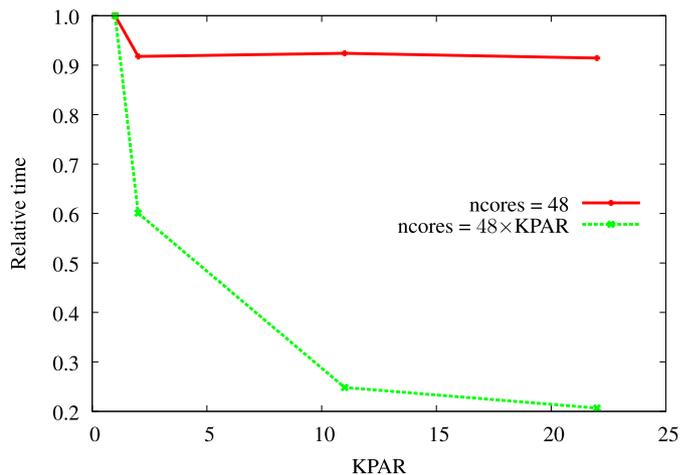


Fig. 5. The effect of KPAR on wall-clock time for a $p(3 \times 3)$ sheet of graphene, where the relative time is (wall-clock time)/(wall-clock time with KPAR = 1), KPAR has values of 1 (no parallelism over k -points), 2, 11 and 22 and where a total of 22 k -points are used in the calculations. The red (solid) line is for the number of MPI processes kept constant, and KPAR is changed. The green (dashed) line is for the number of MPI processes scaled with the value of KPAR. In both cases the NPAR parameter is set to 2.

There are two ways to use the k -points parallelism: either to scale an existing calculation to more cores, reducing the time for the calculation (time-to-science), or to improve the efficiency of a calculation to keep the overall cost low while still getting good performance, as long as NPAR is also adjusted. In this model k -point parallelism provides a means to scale beyond 48 cores, when other modes of parallelism (e.g. over bands and g -vectors) have been exhausted. This approach is particularly useful in situations where a job cannot finish before the maximum permitted run time for a job on a managed computer, and importantly for the increase in massively parallel computers.

5. Conclusions

We have introduced k -point parallelism into VASP 5. Initial results suggest that for small to medium size systems the improvement to scaling over many processors, in particular for our test cases on HECToR, allows many more cores to be exploited efficiently and a marked reduction in times to solution. This improvement will allow modelling of many important systems, and three topical examples have been presented including the optical properties of transparent conducting oxide in optoelectronic devices, the solution energies of metallic systems and the electronic properties of graphene. The implementation allows for efficient usage of massively parallel supercomputers.

Acknowledgements

Via our membership of the UK's HPC Materials Chemistry Consortium, which is funded by EPSRC (EP/F067496), this work made use of the facilities of HECToR, the UK's national high-performance computing service, which is provided by UoE HPCx Ltd at the University of Edinburgh, Cray Inc. and NAG Ltd, and funded by the Office of Science and Technology through EPSRC's High End Computing Programme. This work has been funded under the EPSRC supported distributed CSE programme with the HECToR project. R.G.C. acknowledges support from EPSRC grant EP/J001775/1. The authors would also like to thank A. Turner, D. Alfé and especially Profs J. Hafner, G. Kresse and the VASP team for helpful input and for making the code available to us.

References

- [1] R. Car, M. Parrinello, *Phys. Rev. Lett.* 55 (1985) 2471.
- [2] P. Hohenberg, W. Kohn, *Phys. Rev.* 136 (1964) B864.
- [3] W. Kohn, L.J. Sham, *Phys. Rev.* 140 (1965) A1133.
- [4] J.C. Phillips, *Phys. Rev.* 112 (1958) 685.
- [5] M.L. Cohen, V. Heine, *Solid State Phys.* 24 (1970) 37.
- [6] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, J.D. Joannopoulos, *Rev. Mod. Phys.* 64 (1992) 1045.
- [7] <http://www.hector.ac.uk>.
- [8] G. Kresse, J. Hafner, *Phys. Rev. B* 47 (1993) 558.
- [9] G. Kresse, J. Hafner, *Phys. Rev. B* 49 (1994) 14251.
- [10] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* 6 (1996) 15.
- [11] G. Kresse, J. Furthmüller, *Phys. Rev. B* 54 (1996) 11169.
- [12] <http://cms.mpi.univie.ac.at/vasp/vasp/vasp.html>.
- [13] http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0303.pdf.
- [14] http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0510.pdf.
- [15] M.D. Segall, P.L.D. Lindan, M.J. Probert, C.J. Pickard, P.J. Hasnip, S.J. Clark, M.C. Payne, *J. Phys.: Cond. Matt.* 14 (11) (2002) 2717–2743.
- [16] S.J. Clark, M.D. Segall, C.J. Pickard, P.J. Hasnip, M.J. Probert, K. Refson, M.C. Payne, *Z. Kristallogr.* 220 (5–6) (2005) 567–570.
- [17] R. Dovesi, R. Orlando, B. Civalleri, R. Roetti, V.R. Saunders, C.M. Zicovich-Wilson, *Z. Kristallogr.* 220 (2005) 571.
- [18] R. Dovesi, V.R. Saunders, R. Roetti, R. Orlando, C.M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N.M. Harrison, I.J. Bush, P. D'Arco, M. Llunell, *CRYSTAL09*, in: *CRYSTAL09 User's Manual*, University of Torino, Torino, 2009.
- [19] N.W. Ashcroft, N.D. Mermin, *Solid State Physics*, Holt Saunders, Philadelphia, 1976.
- [20] I.J. Bush, S. Tomić, B.G. Searle, G. Mallia, C.L. Bailey, B. Montanari, L. Bernasconi, N.M. Harrison, *Proc. Roy. Soc. A* 467 (2011) 2112–2126.
- [21] I.J. Bush, *ACM SIGPLAN Fortran Forum* 29 (3) (2011).
- [22] G. Amdahl, in: *AFIPS Conference Proceedings*, vol. 30, 1967, pp. 483–485.
- [23] R. Grau-Crespo, A.Y. Al-Baitai, I. Saadoune, N.H. De Leeuw, *J. Phys.: Cond. Matter* 22 (2010) 255401.
- [24] H.R. Chauke, P. Murovhi, P.E. Ngoepe, N.H. de Leeuw, R. Grau-Crespo, *J. Phys. Chem. C* 114 (2010) 15403–15409.
- [25] R. Grau-Crespo, K.C. Smith, T.S. Fisher, N.H. de Leeuw, U.V. Waghmare, *Phys. Rev. B* 80 (2009) 174117.
- [26] R. Grau-Crespo, N. Cruz-Hernández, J.F. Sanz, N.H. de Leeuw, *J. Mater. Chem.* 19 (2009) 710–717.
- [27] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865.
- [28] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 78 (1997) 1396.
- [29] A.D. Becke, *Density functional thermochemistry. III. The role of exact exchange*, *J. Chem. Phys.* 98 (1993) 5648.
- [30] <http://www.hector.ac.uk/service/hardware/>.
- [31] M. Gajdos, K. Hummer, G. Kresse, J. Furthmüller, F. Bechstedt, *Phys. Rev. B* 73 (2006) 045112.
- [32] J. Paier, M. Marsman, G. Kresse, *Phys. Rev. B* 78 (2008) 121201.
- [33] A. Walsh, C.R.A. Catlow, A.A. Sokol, S.M. Woodley, *Chem. Mater.* 21 (2009) 4962.
- [34] C. Korber, et al., *Phys. Rev. B* 81 (2010) 165207.
- [35] A. Walsh, et al., *Phys. Rev. Lett.* 100 (2008) 167402.
- [36] A. Walsh, C.R.A. Catlow, K.H.L. Zhang, R.G. Egdell, *Phys. Rev. B* 83 (2011) 161202.
- [37] B. Adolph, J. Furthmüller, F. Bechstedt, *Phys. Rev. B* 63 (2001) 125108.
- [38] F.J. Morin, *Phys. Rev. Lett.* 3 (1959) 34–36.
- [39] S.M. Woodley, *Chem. Phys. Lett.* 453 (4–6) (2008) 167–172.
- [40] C. Tang, P. Georgopoulos, M.E. Fine, J.B. Cohen, M. Nygren, G.S. Knapp, A. Aldred, *Phys. Rev. B* 31 (1985) 1000–1011.
- [41] T.D. Manning, I.P. Parkin, *J. Mater. Chem.* 14 (2004) 2554–2559.
- [42] M. Netsianda, P.E. Ngoepe, C. Richard, S.M. Woodley, *Chem. Mater.* 20 (5) (2008) 1764–1772.
- [43] N.R. Mlyuka, G.A. Niklasson, C.G. Granqvist, *Appl. Phys. Lett.* 95 (2009) 171909.
- [44] W. Burkhardt, T. Christmann, S. Franke, W. Kriegseis, D. Meister, B.K. Meyer, W. Niessner, D. Schalch, A. Scharmann, *Thin Solid Films* 402 (2002) 226–231.
- [45] F.A. Cotton, G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., Wiley–Interscience, New York, 1988, xxi+1145 pp.
- [46] R. Grau-Crespo, N.H. De Leeuw, S. Hamad, U.V. Waghmare, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 467 (2011) 1925–1938.
- [47] K.C. Smith, T.S. Fisher, U.V. Waghmare, R. Grau-Crespo, *Phys. Rev. B* 82 (2010) 134109.
- [48] J. Du, Y. Gao, H. Luo, L. Kang, Z. Zhang, Z. Chen, C. Cao, *Solar Energy Materials and Solar Cells* 95 (2011) 469–475.
- [49] C.G. Granqvist, G.A. Niklasson, N.R. Mlyuka, *Thermochromic material and fabrication thereof*, WO/2010/038202, 2010.
- [50] C. Lee, X.D. Wei, J.W. Kysar, J. Hone, *Science* 321 (2008) 5887.
- [51] Ž. Šljivančanin, E. Rauls, L. Hornekær, W. Xu, F. Besenbacher, B. Hammer, *J. Chem. Phys.* 131 (2009) 084706.
- [52] P.E. Blöchl, *Phys. Rev. B* 50 (1994) 17953.