



Cite this: DOI: 10.1039/c5cp00437c

# Accurate and efficient linear scaling DFT calculations with universal applicability

 Stephan Mohr,<sup>a</sup> Laura E. Ratcliff,<sup>ab</sup> Luigi Genovese,<sup>\*a</sup> Damien Caliste,<sup>a</sup>  
Paul Boulanger,<sup>a</sup> Stefan Goedecker<sup>c</sup> and Thierry Deutsch<sup>a</sup>

Density functional theory calculations are computationally extremely expensive for systems containing many atoms due to their intrinsic cubic scaling. This fact has led to the development of so-called linear scaling algorithms during the last few decades. In this way it becomes possible to perform *ab initio* calculations for several tens of thousands of atoms within reasonable walltimes. However, even though the use of linear scaling algorithms is physically well justified, their implementation often introduces some small errors. Consequently most implementations offering such a linear complexity either yield only a limited accuracy or, if one wants to go beyond this restriction, require a tedious fine tuning of many parameters. In our linear scaling approach within the BigDFT package, we were able to overcome this restriction. Using an ansatz based on localized support functions expressed in an underlying Daubechies wavelet basis – which offers ideal properties for accurate linear scaling calculations – we obtain an amazingly high accuracy and a universal applicability while still keeping the possibility of simulating large system with linear scaling walltimes requiring only a moderate demand of computing resources. We prove the effectiveness of our method on a wide variety of systems with different boundary conditions, for single-point calculations as well as for geometry optimizations and molecular dynamics.

Received 23rd January 2015,  
Accepted 24th April 2015

DOI: 10.1039/c5cp00437c

www.rsc.org/pccp

## 1. Introduction

The Kohn–Sham formalism of density functional theory (DFT)<sup>1,2</sup> has established itself as one of the most powerful electronic structure methods due to its good balance between accuracy and speed and is thus popular in various fields such as physics, chemistry, biology, and material sciences. Kohn–Sham DFT maps the problem of interacting electrons onto a problem of non-interacting quasi-electrons: given a system containing  $n$  electrons, the  $3n$ -dimensional many-electron wave function is assumed to be given by a single Slater determinant built from  $n$  orthonormal single particle orbitals  $\psi(\mathbf{r})$  which *a priori* extend over the entire system. The drawback of this approach is that it has an inherent cubic scaling due to the orthonormality which is imposed on these so-called Kohn–Sham orbitals. To orthonormalize a set of  $n$  functions it is necessary to calculate the scalar products among all of them, which has the complexity  $\mathcal{O}(n^2)$ . In addition the cost of calculating one scalar product is proportional to the size of the underlying basis,  $m_{\text{basis}}$ . Since both  $n$  and  $m_{\text{basis}}$  are proportional to the total size of the system – typically indicated by the number of

atoms  $N$  – one ends up with a complexity  $\mathcal{O}(N^3)$ . This behavior is common to all programs using a systematic basis set, be it plane waves,<sup>3–5</sup> finite elements<sup>6</sup> or wavelets.<sup>7</sup> Even when not using a systematic basis set, *e.g.* in the case of Gaussians<sup>8</sup> or atomic orbitals,<sup>9</sup> the complexity remains  $\mathcal{O}(N^3)$  due to the matrix diagonalization which is required in a straightforward implementation. Due to this limitation this standard approach is only suitable for systems containing a few hundred or at most a few thousand atoms. As a consequence a number of algorithms which are capable of performing linear scaling calculations have been developed over the past decades, as for example the divide-and-conquer method,<sup>10</sup> the density-matrix minimization approach,<sup>11</sup> or the Fermi operator expansion.<sup>12–14</sup> An exhaustive overview is given in ref. 15.

In order to obtain an algorithm which can scale linearly with respect to the size of the system, one thus has to abandon the concept of the extended Kohn–Sham orbitals  $\psi_i$  and work with other quantities which are strictly localized. One such quantity is the density matrix  $\mathbf{F}(\mathbf{r}, \mathbf{r}') = \sum_i f_i \psi_i(\mathbf{r}) \psi_i(\mathbf{r}')$  with  $f_i$  being the occupation number of orbital  $i$ . For insulators or metals at finite temperature it can be shown that the elements of  $\mathbf{F}(\mathbf{r}, \mathbf{r}')$  decay exponentially for increasing  $|\mathbf{r} - \mathbf{r}'|$ .<sup>16–22</sup> One particular approach is to write the density matrix in separable form by introducing a set of so-called support functions,<sup>23,24</sup> an idea that has already been used in various codes such as ONETEP,<sup>25</sup> Conquest,<sup>26</sup> CP2K<sup>27</sup> and SIESTA.<sup>28</sup> These support functions can

<sup>a</sup> Université Grenoble Alpes, CEA, INAC-SP2M, F-38000 Grenoble, France.  
E-mail: luigi.genovese@cea.fr

<sup>b</sup> Argonne Leadership Computing Facility, Argonne National Laboratory, Illinois 60439, USA

<sup>c</sup> Institut für Physik, Universität Basel, Klingelbergstr. 82, 4056 Basel, Switzerland

also be thought of as a localized basis to directly represent the Kohn–Sham orbitals. Consequently, if one neglects elements of the density matrix which are below a given threshold, the number of non-zero elements scales only linearly with respect to the size of the system, thus paving the way towards an algorithm with such a reduced complexity. Even if there exists an ideal set of support functions – namely the maximally localized Wannier functions<sup>29</sup> – this is of course not known beforehand. There are thus two possibilities: either use a larger number of support functions and hope that the most important features of the Wannier functions can be captured in this way, or to use a smaller number of support functions and optimize them *in situ* (namely within a predefined localization region) to get as close as possible to the Wannier functions. Another important aspect is the choice of the underlying basis which is used to give a numerical representation of the support functions. Ideally, as we would like to efficiently describe localized functions, such a basis set should at the same time feature orthogonality and compact support. This can indeed be offered by Daubechies wavelets,<sup>30</sup> making them an ideal basis set for linear scaling calculations. Moreover, wavelets are able to yield a reasonable precision by keeping the number of computational degrees of freedom relatively small for a systematic approach. In addition the multiresolution properties of wavelets allow for an adaptive resolution and thus to have a finer mesh only in those regions – close to the atoms – where it is actually required.

We have recently implemented a linear scaling code based on Daubechies wavelets, distributed in the BigDFT package,<sup>31</sup> where the support functions are optimized *in situ* and therefore optimally adapted to their chemical environment. Consequently we are in this way able to drastically reduce the size of the subspace in which the density matrix is represented, yielding at the same time very high accuracy. Together with the aforementioned properties of Daubechies wavelets this leads to an optimally reduced number of degrees of freedom and enables us to perform calculations on thousands of atoms while only requiring moderate amounts of memory and CPU time.

In this paper we will present the progress we have recently made for this linear scaling code. Thanks to the properties of Daubechies wavelets, we have been able to identify an algorithm that guarantees a very high accuracy and at the same time provides strongly localized support functions, with the advantage of being automatically adapted to the localization property of the system’s density matrix. Our code offers the ability to perform linear scaling DFT calculations for various boundary conditions, namely either isolated, surfaces or periodic, and at the same time yields an astonishingly high accuracy, exhibits an universal applicability, and requires only a limited amount of computational resources, therefore making advanced simulations like geometry optimization and molecular dynamics very efficient. In particular it offers the advantage that one does not need to do a tedious fine tuning of various parameters, but can straightforwardly get accurate results using a set of default values.

We will show that thanks to our method one can calculate easily energies with an absolute accuracy of the order of 10 meV per atom for a large variety of systems, with forces that are

accurate enough such that geometry optimization and molecular dynamics calculations can be done with the same accuracy of a cubic-scaling approach. From a computational point of view, as the approach scales linearly with respect to system size, it is possible to use the concept of “CPU minutes per atom” to evaluate the computational resources needed for a full DFT calculation. Our code exhibits an excellent level of parallelism that enables effective  $\mathcal{O}(N)$  computations of systems with thousands of atoms with moderate usage of computational resources. We will first present the basic principles of our approach and then demonstrate all of the aforementioned properties by various examples.

## II. Theory and algorithm

The paper in ref. 31 provides a complete description of our algorithm for systems under isolated boundary conditions (BC). Such an approach has now been generalized to periodic and surface BC, following the same guidelines. Here we only recall the main aspect of our method, and refer to the above paper for a detailed presentation.

### A. Overview of the algorithm

The linear scaling version of BigDFT is based on a separable ansatz for the density matrix. Using so-called support functions  $\phi_\alpha(\mathbf{r})$  and a density kernel  $\mathbf{K}$  the density matrix can be written as

$$F(\mathbf{r}, \mathbf{r}') = \sum_{\alpha, \beta} \phi_\alpha(\mathbf{r}) K_{\alpha\beta} \phi_\beta(\mathbf{r}'). \quad (1)$$

Using the orthonormality relations between the support functions and their duals,  $\int \phi_\alpha(\mathbf{r}) \tilde{\phi}_\beta(\mathbf{r}) d\mathbf{r} = \delta_{\alpha\beta}$ , it follows that the elements of the density kernel are given by

$$K_{\alpha\beta} = \iint \tilde{\phi}_\alpha(\mathbf{r}) F(\mathbf{r}, \mathbf{r}') \tilde{\phi}_\beta(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \quad (2)$$

The charge density is given by the diagonal part of the density matrix  $\rho(\mathbf{r}) = F(\mathbf{r}, \mathbf{r})$ , and enters the Kohn–Sham Hamiltonian operator

$$\mathcal{H}[\rho] = -\frac{1}{2}\nabla^2 + \mathcal{V}_{\text{KS}}[\rho] + \mathcal{V}_{\text{PSP}}, \quad (3)$$

where  $\mathcal{V}_{\text{KS}}[\rho](\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \mathcal{V}_{\text{XC}}[\rho](\mathbf{r})$  is the Kohn–Sham potential containing the Hartree potential and the exchange–correlation potential, and  $\mathcal{V}_{\text{PSP}}$  is the pseudopotential representing the ions in the system. In BigDFT the pseudopotentials are norm-conserving GTH-HGH<sup>32</sup> pseudopotentials and their Krack variants,<sup>33</sup> possibly enhanced by a nonlinear core correction.<sup>34</sup>

Using the formalism based on the density matrix  $F(\mathbf{r}, \mathbf{r}')$ , the band structure energy, which is the central quantity to be calculated, is given within the framework of Kohn–Sham DFT by the expression

$$E_{\text{BS}} = \text{tr}(\mathcal{H}[\rho]F) = \text{tr}(\mathbf{KH}), \quad (4)$$

where the elements of the Hamiltonian matrix are defined as

$$H_{\alpha\beta} = \int \phi_\alpha(\mathbf{r}) \mathcal{H}(\mathbf{r}, \mathbf{r}') \phi_\beta(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \quad (5)$$

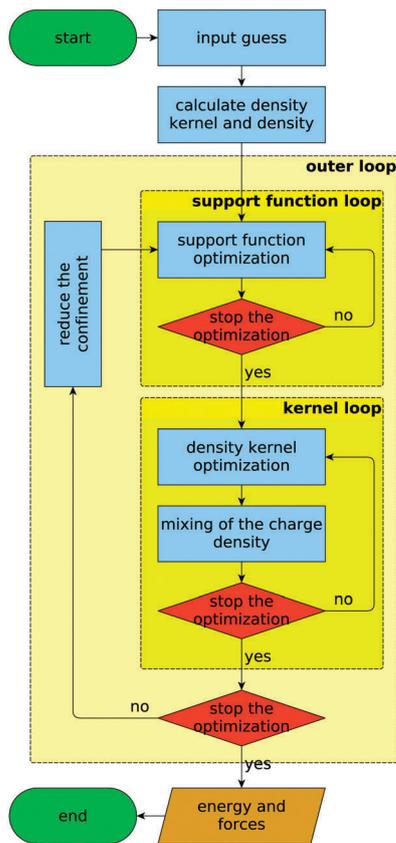


Fig. 1 Flowchart of the linear scaling algorithm. It consists of a double loop structure: in the inner first loop the support functions are optimized with a fixed density kernel, in the second inner loop the density kernel is optimized using a fixed set of support functions. These two loops are then iterated in an outer loop until overall convergence is reached.

The algorithm thus consists of two parts: finding a good set of support functions  $\phi_\alpha$ , which allow one to write the Hamiltonian matrix  $\mathbf{H}$ , and determining the density kernel  $\mathbf{K}$  in the dual (and never explicitly calculated) basis  $\tilde{\phi}_\alpha(\mathbf{r})$ . This second part is done self-consistently, *i.e.* the Hamiltonian is updated in each iteration. The support functions are expanded in terms of an underlying Daubechies wavelet basis, and we again refer to ref. 31 for the details. A flowchart of the algorithm is shown in Fig. 1.

### B. Locality of the support functions

In order to reach linear scaling, the support functions must be localized and the density kernel sparse. In other terms, we need to establish some mapping between the Kohn–Sham degrees of freedom and some localization regions of the system, which are typically centered on the atoms. For the support functions this can be achieved by introducing cutoff radii around the center of each function and setting to zero all Daubechies wavelet components which lie outside of the region; for the density kernel an element  $K_{\alpha\beta}$  is set to zero if the centers of  $\phi_\alpha$  and  $\phi_\beta$  are farther away than the sum of the kernel cutoffs of the regions  $\alpha$  and  $\beta$ . As the kernel cutoff is related to the extension of the dual support functions (see eqn (2)), the sparsity of  $\mathbf{K}$  is governed by the cutoff implicitly used for these dual functions.

In contrast to other linear scaling codes employing the same support function approach<sup>25–28</sup> we decided to impose an additional constraint to keep the support functions approximately orthogonal. In other terms, we optimize  $\phi_\alpha$  such that the overlap matrix

$$\mathbf{S}_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle \quad (6)$$

is close to the identity matrix. By choosing the support functions to be orthogonal the dual and non-dual entities are identical and the cutoffs for the support functions and the density kernel matrix elements can consequently be chosen along the same lines, leading to a degree of sparsity for the kernel which is comparable to that of the overlap matrix itself.

In addition, quasi-orthogonal support functions lead to an overlap matrix which is close to the identity and whose inverse can thus be cheaply calculated using polynomial expansions. This property is in particular important for the Fermi operator expansion<sup>12,13</sup> (FOE) procedure that we use to determine the density kernel. The efficiency of this approach relies heavily on the sparsity of the matrix  $\mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2}$ . Even if  $\mathbf{H}$  is very sparse thanks to the cutoff radii of the support functions, this property is much less pronounced for  $\mathbf{S}^{-1/2}$  in the case of non-orthogonal support functions. Thus, if one wants to keep a high degree of sparsity also for the matrix product, it is indispensable to work with a set of orthogonal support functions. For the determination of the density kernel the code offers several possibilities, each one having its particular strengths and thus areas of application. In this paper we always used the FOE method, as it is a method that allows one to perform calculations for very large systems. However, there is no reason not to adopt also other approaches like the Pole Expansion method.<sup>35</sup>

### C. Optimization of the support functions

The support functions are optimized by minimizing a target function. Ideally this target function should lead to strongly localized support functions which at the same time yield a very high accuracy. For the latter property the correct quantity to be minimized is the band structure energy of eqn (4). In order to combine the two properties we define our target function as

$$\Omega = \sum_{\alpha} K_{\alpha\alpha} \langle \phi_\alpha | \mathcal{H}_\alpha | \phi_\alpha \rangle + \sum_{\beta \neq \alpha} K_{\alpha\beta} \langle \phi_\alpha | \mathcal{H} | \phi_\beta \rangle, \quad (7)$$

where  $\mathcal{H}_\alpha$  is the Kohn–Sham Hamiltonian including a confinement, *i.e.*  $\mathcal{H}_\alpha = \mathcal{H} + c_\alpha(\mathbf{r} - \mathbf{R}_\alpha)^4$ , with  $\mathbf{R}_\alpha$  being the center of the localization region. The prefactor for the confinement  $c_\alpha$  is smoothly reduced during the run, as explained in more detail in ref. 31. In this way we have a strong confinement in the beginning, leading to a decent localization, and still obtain a high precision as we correctly minimize the band structure energy in the end. Since the decrease of the confinement is done automatically – taking into account the properties of the system –, this procedure is universally applicable and thus enables us to obtain at the same time localized and accurate support functions without the need for any fine tuning. In addition – as is explained in more detail in Section IIIA of ref. 31 – the adoption of quasi-orthogonal support functions

further justifies the use of the  $\alpha$ -dependent Hamiltonian, as the localized gradient is expressed only in terms of orthogonalized  $\phi_\alpha$ .

In spite of all these benefits, it must be stated that orthogonality and locality are in general two contradicting properties; consequently a strict enforcement of the orthogonality might lead to convergence problems. Therefore we only perform an explicit orthogonalization in the very beginning; in the following the orthogonality is only approximately conserved by the use of a Lagrange multiplier in the support function gradient. This is enough to keep the overlap matrix diagonally dominant – the off-diagonal elements being typically of the order of 0.1 – and thus to maintain the aforementioned benefits.

### III. Assessment of the accuracy

In order to assess the precision of our code, it has to be compared with a reference. We use as such the cubic version of the code,<sup>7</sup> which has been shown to yield very accurate results for a wide range of systems,<sup>34</sup> in either free, periodic or surface BC. Thus, by demonstrating that the linear version of BigDFT is able to reproduce the results of the cubic scaling version, we can be sure to have reached a very high level of precision. All calculations were done using the PBE<sup>36</sup> functional unless otherwise stated.

#### A. Comparison of energies and forces

One of the most important features of our code is the ability to yield a high level of precision without the need for doing a tedious fine tuning of the parameters. As an illustration of this property, we show the energies and forces calculated with both the linear and cubic scaling version for eight systems with rather different electronic structures; they are depicted in Fig. 2. The chosen parameters were very similar for all systems, with the main difference being the various cutoff radii used for the support function regions of the different atom kinds. For the first six systems this choice was solely based on the row in which the element appears in the periodic table: 5.5 bohr for elements of the first row, 6.3 bohr for those of the second row, and 7.0 bohr for those of the third row. The kernel cutoff was set to 9.0 bohr; the code automatically adjusts this value if it is smaller than the cutoff radius of the support function plus 8 times the grid spacing, see ref. 31 for details. The number of support functions per atom we used for these benchmarks was 1 for H, 5 for Pb, 9 for Si, and 4 for all the other elements. These parameters seem to be fairly universal, meaning that any other system can be readily calculated without the need for any adjustments. The last two systems we have included in our test set exhibit periodic and surface boundary conditions, resp.: a periodic hybrid organic–inorganic perovskite structure ( $\text{PbI}_3\text{CNH}_6$ )<sub>64</sub> and a graphene surface with 836 atoms, both calculated at the  $\Gamma$  point. These two systems were calculated using even smaller values for the support function cutoff radii in order to demonstrate the robustness of our algorithm; of course the convergence criteria were chosen less tight. The specific values were 5.0 bohr for H, 5.5 bohr for C and N, 7.5 bohr for Pb, and 8 bohr for I. In addition, as graphene is known to be a system with no gap and is thus to be expected to have a

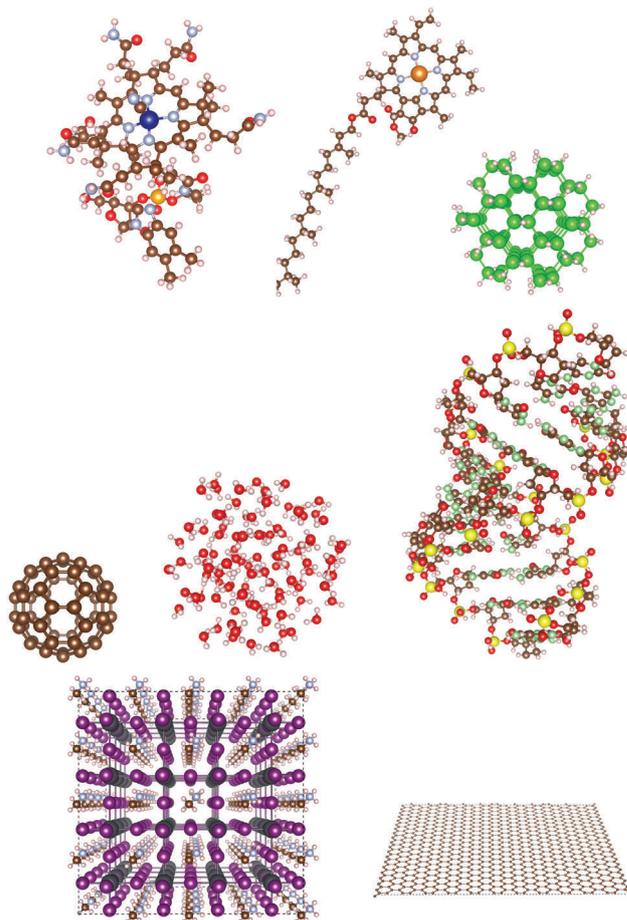
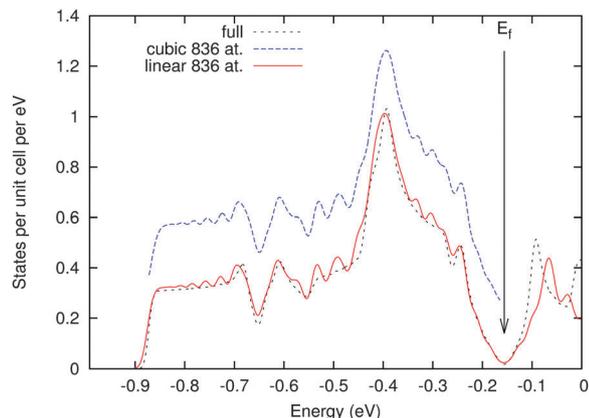


Fig. 2 Illustration of the eight systems used for the comparison of the linear and cubic energies and forces: vitamin B<sub>12</sub>, chlorophyll, Si<sub>87</sub>H<sub>76</sub>, C<sub>60</sub>, (H<sub>2</sub>O)<sub>100</sub>, DNA, (PbI<sub>3</sub>CNH<sub>6</sub>)<sub>64</sub>, graphene.

density kernel which decays more slowly, the cutoff for the latter was slightly increased to 10 bohr. To illustrate the reliability of our approach even for smaller cutoff radii, we show in Fig. 3 the Density of States obtained from the eigenvalues of the matrix  $\mathbf{H}$  of eqn (5) in the support function basis against the fully unconstrained calculation given by the cubic code for C<sub>836</sub> at the  $\Gamma$  point. As can be seen, ground-states features can be described accurately by a basis set with only 4 degrees of freedom per atom. In other terms, we have shown that, for such cases, delocalized degrees of freedom like the components of the last occupied orbital of graphene can be efficiently described by localized support functions; in spite of the fact that the systems are not extremely large, our calculations lead to matrices whose degree of sparsity is already between and 84% and 95%. A similar behavior can be observed for the perovskite structure, where the sparsity of the matrices lies in the range from 68% to 90%.

The results for all eight systems are summarized in Table 1. As can be seen from the first 6 examples, it is possible to get with these input parameters a fairly constant energy offset between the linear and cubic versions of about 10 meV per atom. The difference of the force norm is typically of the order



**Fig. 3** Normalized density of states (DoS) for the  $C_{836}$  graphene structure calculated both with the linear scaling code (solid red curve) and with the cubic code (blue dashed curve), shifted to show the complete resemblance. A Gaussian smearing of 10 meV is applied. For the cubic code, only the states below the Fermi level  $E_f$  are calculated. For comparison, the dotted curve shows the DoS for a fully-converged  $k$ -point calculation made with the BigDFT code using 4 atoms in the (orthorhombic) unit cell and a Monkhorst-Pack mesh of  $36 \times 36$   $k$ -points, corresponding to 361 irreducible points.

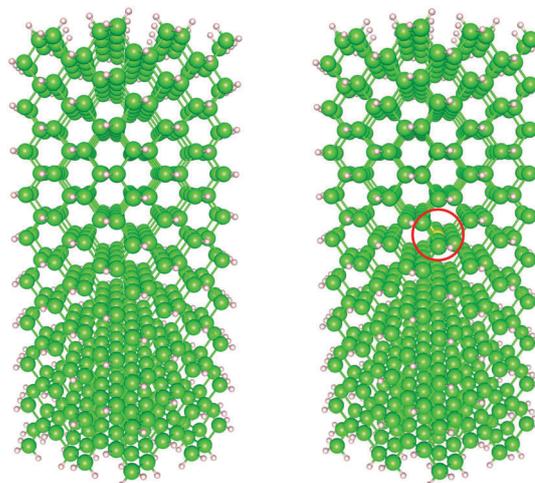
of  $10^{-3}$  hartree per bohr. The last two examples, which have been calculated using the smaller cutoff radii, show that this parameter is the only one that affects the obtained precision. We emphasize that these levels of precision, largely sufficient to treat the great majority of systems at the DFT level, were obtained without any fine tuning, using the same set of input parameters. Last but not least, it is worth noting that the support functions are optimized such that Pulay forces<sup>37</sup> are absent in our approach,<sup>31</sup> and the evaluation of the forces is therefore straightforward.

## B. Calculations of energy differences

Absolute energies are always somewhat arbitrary as they are only defined up to an additive constant. Therefore energy differences are more meaningful as this ambiguity vanishes. For the linear scaling version there is in addition the benefit that the offset caused by the finite cutoff radii cancels to a large degree and it is thus possible to get very close to the exact result.

The absolute energy error of 10 meV per atom mentioned in the previous section can thus be considered as an upper bound; in fact it is however possible to obtain a much higher accuracy than one might think at first sight.

As an example, we calculated the energy difference between a hydrogen-passivated silicon nanowire in its pristine configuration and another one which contains a defect. As a defect we chose a substitutional atom; one of the silicon atoms was replaced by phosphorus. These systems have been used as a case study to determine the binding energy of impurities in semiconductors using charged DFT calculations, see ref. 38. The two configurations are depicted in Fig. 4. Apart from the fact that silicon is a rather delicate system, this defect is even more challenging as it requires the comparison of two systems with a different number of electrons. The calculations were performed with the analogous parameters as those used for the benchmarks in Section IIIA, again demonstrating the universal character of the approach. From the results in Table 2 it becomes clear that energy differences can indeed be calculated



**Fig. 4** The two silicon wires (consisting of 660 atoms) which are used to benchmark the calculation of energy differences. Whereas the one on the left side is a pure wire, there is a defect in the one on the right side with one silicon atom having been replaced by a phosphorus atom. The red circle highlights this substitutional atom.

**Table 1** Comparison of the energies and forces for different systems (both relaxed and unrelaxed), calculated using the linear and cubic versions. For the first six systems, the linear energies have an offset of about 10 meV per atom with respect to the cubic version, and the deviations of the linear forces from their cubic counterparts is of the order of  $10^{-3}$  hartree per bohr. The last two systems were calculated using smaller cutoff radii, resulting in a larger offset in the energy. Apart from this difference all systems were calculated using similar parameters

	Cubic		Linear		Difference	
	Energy hartree	Force norm hartree per bohr	Energy hartree	Force norm hartree per bohr	Energy meV per atom	Force norm hartree per bohr
Vitamin B <sub>12</sub>	-926.78	$2.13 \times 10^{-3}$	-926.71	$1.93 \times 10^{-2}$	11.43	$1.72 \times 10^{-2}$
Chlorophyll	-476.70	$3.05 \times 10^{-3}$	-476.64	$1.24 \times 10^{-2}$	12.33	$9.38 \times 10^{-3}$
Si <sub>87</sub> H <sub>76</sub>	-386.79	$7.80 \times 10^{-4}$	-386.69	$1.01 \times 10^{-2}$	17.20	$9.27 \times 10^{-3}$
C <sub>60</sub>	-341.06	$2.69 \times 10^{-4}$	-341.02	$7.36 \times 10^{-3}$	17.23	$7.09 \times 10^{-3}$
(H <sub>2</sub> O) <sub>100</sub>	-1722.99	$5.23 \times 10^{-1}$	-1722.87	$5.26 \times 10^{-1}$	10.89	$2.80 \times 10^{-3}$
DNA	-4483.12	$5.60 \times 10^{-1}$	-4482.84	$5.63 \times 10^{-1}$	10.69	$2.40 \times 10^{-3}$
(PbI <sub>3</sub> CNH <sub>6</sub> ) <sub>64</sub>	-3766.07	$1.40 \times 10^{-1}$	-3765.47	$1.66 \times 10^{-1}$	21.06	$2.58 \times 10^{-2}$
Graphene (836 atoms)	-5430.13	$1.52 \times 10^{-2}$	-5428.91	$4.39 \times 10^{-3}$	39.61	$1.08 \times 10^{-2}$

**Table 2** Energy differences between the pure and the impure Si wire, for both the linear and the cubic version. The energy offset between the linear and the cubic version cancels to a large degree and thus yields a very accurate result for the energy difference

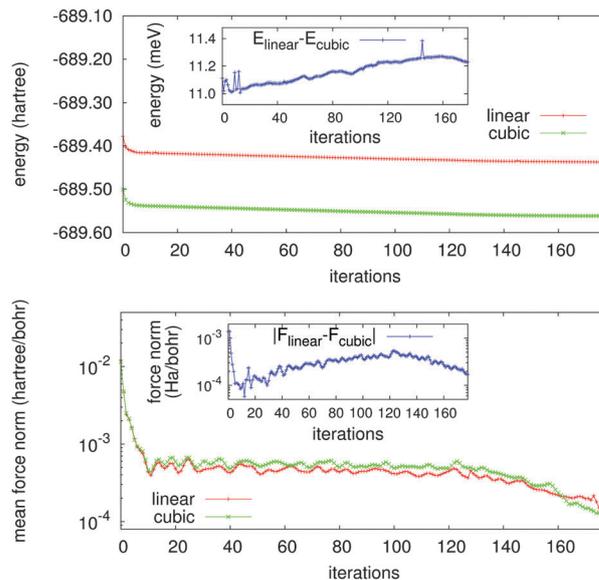
	Pure (eV)	Impurity (eV)	Difference (eV)
Cubic	-53179.7148	-53251.5162	71.8014
Linear	-53168.4248	-53240.3371	71.9123
Difference	11.2900	11.1791	0.1109
Relative difference	—	—	0.15%

with a very high precision. Whereas the offset between the linear and cubic version in the absolute energy is about 11.2 eV – which is about 17 meV per atom, in agreement with the results of Section III A –, the discrepancy between the two versions becomes as little as 0.11 eV for the energy difference between the pure and the impure wire. Comparing this mismatch with the reference value of 71.8 eV one gets a relative error of only 0.15%.

### C. Geometry optimizations and molecular dynamic

A good test to check at the same time the accuracy of the energies and the forces is to perform a geometry optimization for both the linear and the cubic version, starting from the same non-equilibrated structure. If the forces calculated by the linear version are accurate enough, this should lead to identical trajectories and thus to a parallel evolution of the energies and the forces. As an example we took an alkane consisting of 302 atoms, which is a pretty floppy system and requires many iterations to converge. The results are shown in Fig. 5. As can be seen, the offset in the energy between the linear and the cubic version remains more or less constant throughout the entire geometry optimization, and is again of the order of 10 meV per atom. In addition, the forces are pretty much identical, with variations of the order of only  $10^{-4}$  hartree per bohr. This shows that the two trajectories performed by the linear and the cubic versions are identical, demonstrating the high precision of the forces calculated by the linear version. Again it must be stressed that these results were obtained with the same standard set of parameters which had been used in the previous examples.

Another test to check at the same time the accuracy of the energy and the forces is to perform a molecular dynamics (MD) trajectory for both the linear and the cubic version, starting from the same initial configuration. As an example we took a  $C_{60}$  fullerene surrounded by a small shell of 92 water molecules, as depicted in Fig. 6. The calculations were done with similar parameters to the geometry optimization, with the main difference being the cutoff radii which were set to smaller values; for the support functions we used 4.5 bohr for H and 4.8 bohr for C and O, and the cutoff for the density kernel was set to 8.0 bohr for H and 8.1 bohr for C and O. Thanks to these small values, we get – in spite of the fact that the system is not very large – matrices which already exhibit a decent degree of sparsity, namely between 54% and 88%. Still, as can be seen from the results in Fig. 7, these cutoffs are sufficient to get a similar trajectory as with the cubic version.

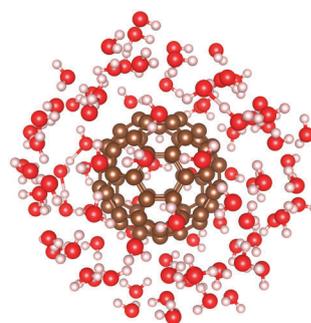


**Fig. 5** Evaluation of the energy (upper panel) and the forces (lower panel) as a function of the number of iterations in the geometry optimization. For the energy, the offset remains rather constant throughout the entire run, leading to two parallel curves; as can be seen from the inset, the offset only varies within less than 1 meV per atom. The curves for the forces are more or less superposed, with a difference of the order of  $10^{-4}$  hartree per bohr. The large plateau in the middle where the force remains constant is due to the relaxation of the elastic wave inside the alkane chain: an adjustment of the bond lengths can only start at the two endpoints of the alkane and then slowly propagate like a wave towards the center of the molecule.

### D. Energetic ordering

The statement that our approach is universally applicable does not mean that it can blindly be applied to any system. There are situations where the “noise” introduced by the finite cutoff radii is larger than the “signal” one is looking for; in this case the results of our approach – as of any method relying on such a truncation – must be interpreted with care. By universally applicable we rather mean that we can straightforwardly – *i.e.* without the need of doing a tedious fine tuning of the input parameters and the basis set – apply our method to those cases where the “signal to noise ratio” is large enough even when finite cutoff radii are used.

As an example for a system where such a set of input parameters is not able to reproduce the correct results of the



**Fig. 6**  $C_{60}$  fullerene surrounded by a shell of 92 randomly arranged water molecules.

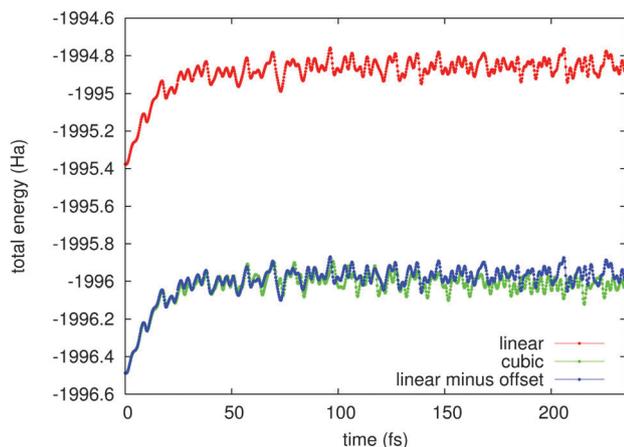


Fig. 7 Total (*i.e.* kinetic plus potential) energy of a short MD trajectory for the system depicted in Fig. 6, for both the linear and the cubic version. The time step was 0.24 fs, and a thermostat was used to keep the averaged temperature of the system at 300 K. For the blue curve, the energy offset between the two versions at the initial configuration was subtracted from the linear curve in order to allow for a better comparison.

cubic version we present the energetic ordering for the 12 energetically lowest structures<sup>39</sup> of  $B_{12}C_{48}$ , depicted in Fig. 8. Such boron–carbon fullerenes are a very delicate system; previous studies<sup>40,41</sup> have even led to different conclusions. In order to successfully describe the energetic ordering of these configurations it is indispensable to catch energy differences – *i.e.* signals – of the order of 1 meV per atom, which can be easily obtained with the cubic version of BigDFT used in ref. 39. The energy spectra for these 12 configurations are shown in Fig. 9, for three different setups: the cubic version using the PBE functional, the linear version using as well the PBE functional, and the cubic version using the LDA<sup>42</sup> functional. As can be seen, there are – even if the main features of the system are well captured – notable differences between the cubic and the linear version: whereas the global minimum configuration is the same, there has been some reordering of other local minima; in particular the energetic gap between the lowest state and the first local minimum is wrong. However the third panel demonstrates that these boron–carbon fullerenes are indeed an extremely delicate system. As can be seen, the energetic ordering is also considerably modified if the cubic version is used with another functional: using LDA instead of PBE, the energy levels of the lowest configuration and the first local minimum remain identical, but the higher levels are completely jumbled. Due to this very high sensitivity it is thus not surprising that the linear version is not able

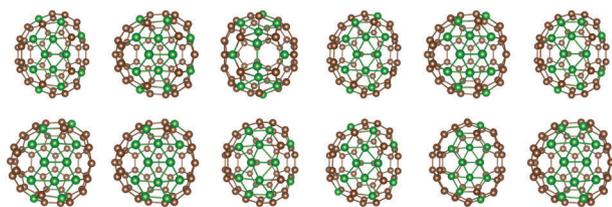


Fig. 8 The 12 energetically lowest configurations for the fullerenes  $B_{12}C_{48}$ .

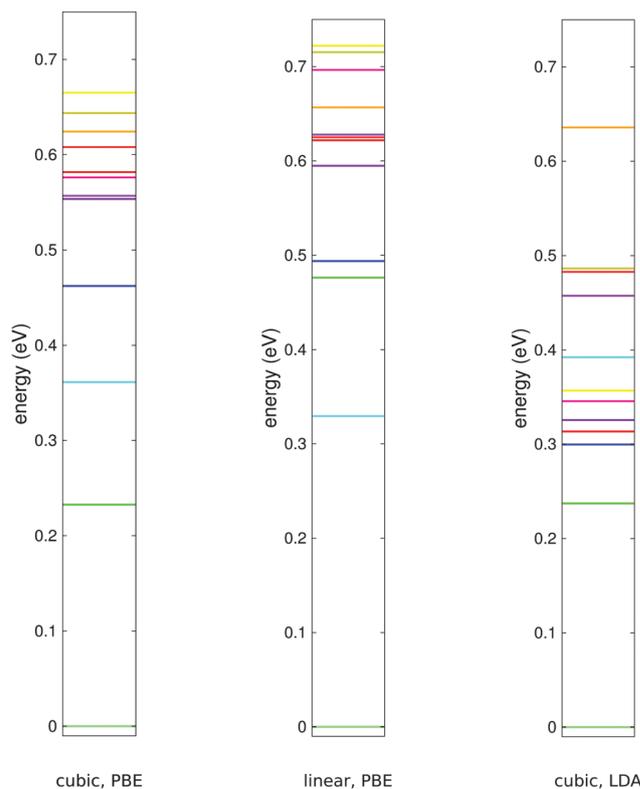
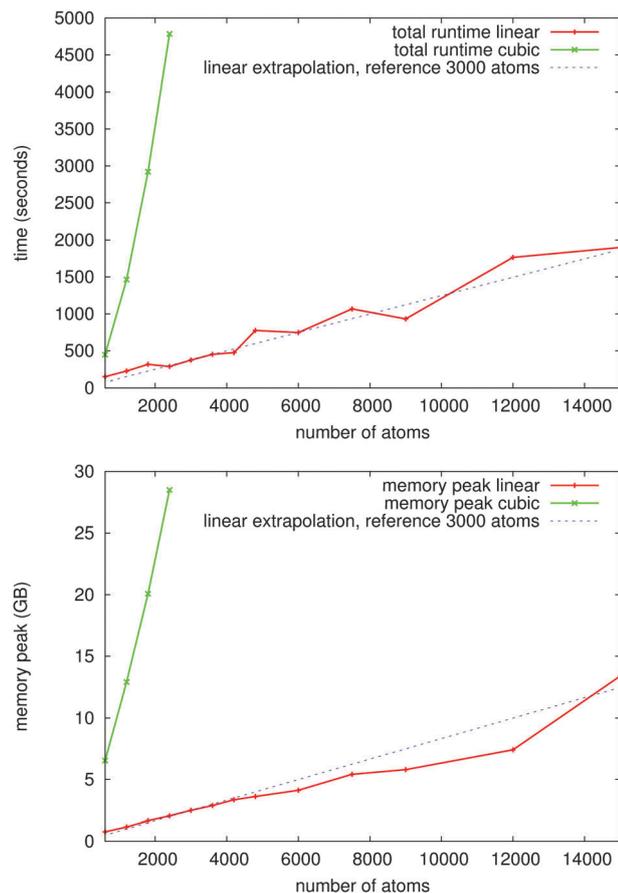


Fig. 9 Energetic ordering of the lowest configurations of  $B_{12}C_{48}$ . From left to right: cubic version with PBE, linear version with PBE, cubic version with LDA. All energy levels have been shifted such that the global minimum is always at zero. The colors correspond always to the same level, demonstrating that the energetic ordering is not always the same.

to satisfactorily reproduce the results of the cubic version, as the noise introduced by the finite cutoff radii is higher than the signal we look for.

## IV. Scaling with system size

The goal of this work was to obtain a code which at the same time yields a very high precision and scales linearly. The first property has already been shown in detail in Section III. To demonstrate the second property we performed single point calculations for randomly generated water droplets of various sizes, for both the linear and the cubic version. The runs were performed in parallel, using in total 6400 cores. The results for the total runtime and the memory peak are shown in Fig. 10. As can be seen both quantities clearly exhibit a strict linear scaling. The cubic version, on the other hand, shows a much steeper increase and does not permit calculations beyond about 2000 atoms. Due to the spherical geometry the degree of sparsity of the matrices is rather low and it is therefore more difficult for the linear scaling version to exploit this property. Nevertheless, even for the smallest droplet used in this benchmark the linear scaling version is faster than the cubic one. Since calculations on even smaller droplets are not meaningful any more due to the large number of cores used, we had to extrapolate the curves of the runtime to get an estimate for the

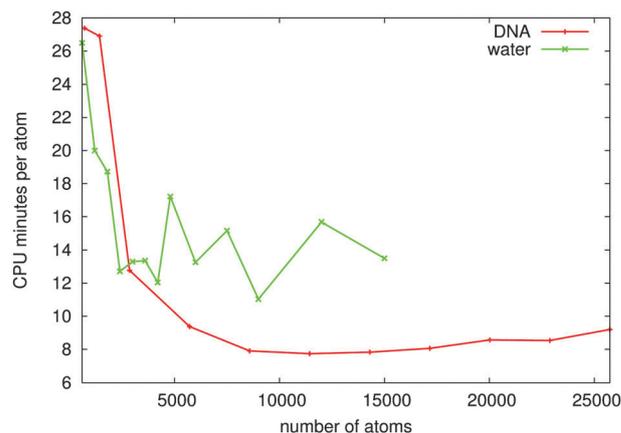


**Fig. 10** Total runtime for one single point calculation (upper panel) and memory peak per MPI task (lower panel) for water droplets of various sizes, ranging from 600 to 15 000 atoms. The linear scaling version indeed exhibits linear scaling, as indicated by the linear extrapolations. The small deviations are mainly caused by a slightly different number of iterations required to reach convergence.

crossover point; the result is about 200 atoms, which is in a range still easily accessible by the cubic version of BigDFT for production runs (see *e.g.* ref. 43).

Moreover it becomes clear that the computational resources consumed by the linear version are only moderate. For a complete single point calculation for 10 000 atoms only a few thousand CPU hours are required, which is a rather small amount considering the capacities of current supercomputers. This demonstrates that we were able to drastically reduce the number of degrees of freedom while keeping a very high level of accuracy.

For a truly linear scaling regime it is also interesting to look at the “CPU minutes per atom” which are needed to perform a fully self-consistent calculation. This quantity can be considered as a (complicated and unknown) functional of the input file. We show such values in Fig. 11 for DNA fragments and the water droplets used for the aforementioned benchmark. As is to be expected for a linear scaling code, this quantity remains constant starting from a certain critical size where the linear scaling parts of the code start to dominate. Moreover one can easily see how the geometry of the system influences the prefactor,



**Fig. 11** CPU minutes per atom needed for a fully self-consistent calculation for DNA fragments and water droplets of different sizes. As expected, this quantity remains constant starting from a certain size, with the prefactor mainly influenced by the geometry of the system to be calculated.

resulting in higher values for the water droplets than for the DNA fragments.

## V. Parallelization

### A. Parallel efficiency

Even if the linear scaling version of BigDFT requires only moderate resources, it is indispensable to have an efficient parallelization scheme in order to keep the runtimes short and thus to have the possibility of performing advanced calculations such as geometry optimizations within a reasonable time frame. To this end we parallelized our code with a mixed distributed/shared memory parallelization scheme using MPI and OpenMP. It is worth noting that the shared memory parallelization is not just an additional speedup. Reducing the number of MPI tasks and in turn increasing the number of OpenMP threads helps to improve the MPI load balancing and to reduce communication overhead; moreover it can substantially reduce the memory peak per compute node and thus help in situations where this resource is critical.

Reaching an efficient parallelization for a linear scaling code is not easy. First of all we have a small number of degrees of freedom and thus little workload that can be shared among the cores. Moreover there are also load balancing problems which can arise due to the different sizes and surroundings of the localization regions. Some solutions to deal with these problems have already been described in a previous publication;<sup>31</sup> in Appendix A we now present a novel approach to improve the handling of the sparse matrices. In spite of the aforementioned obstacles we are able to reach a degree of parallelism which is more than sufficient to efficiently perform large calculations. As an illustration we show the scaling for a fully self-consistent calculation of a DNA fragment consisting of 14 300 atoms. The number of MPI tasks ranged from 160 to 3200, each one being split into 8 OpenMP threads; thus in total we have a range of 1280 to 25 600 cores. These are similar values as used by other linear scaling DFT codes.<sup>44</sup> The results for the runtime and the

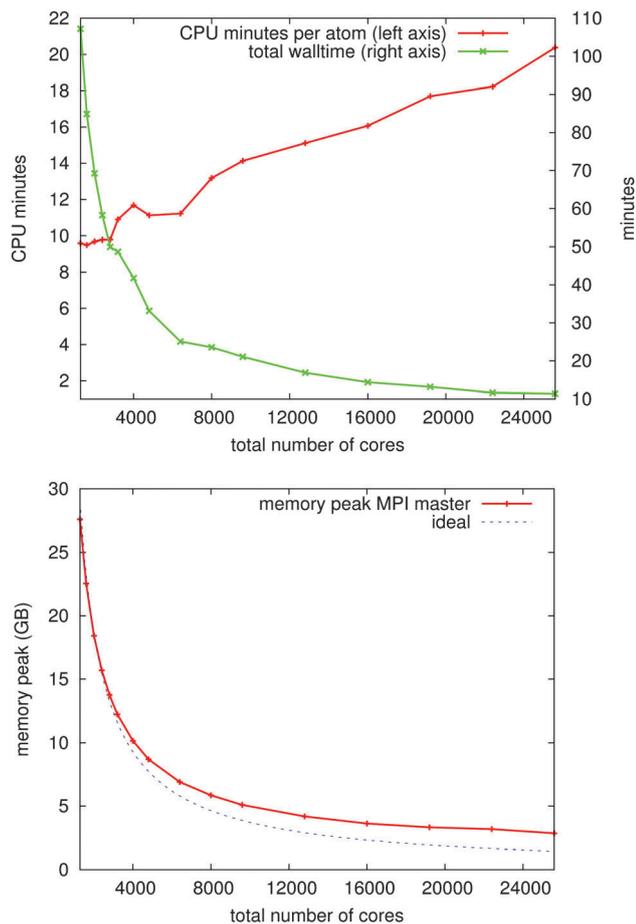


Fig. 12 CPU minutes (defined as walltime times number of cores) per atom and total walltime (upper panel) and memory peak (lower panel) as a function of the total number of cores for a DNA fragment consisting of 14 300 atoms. The runs were all performed using 8 OpenMP threads for each MPI task, thus only the latter number was varied.

memory peak are shown in Fig. 12. For the runtime we show the CPU minutes per atom as well as the total walltime.

As can be seen from the CPU minutes per atom remaining almost constant the speedup is very good up to a few thousand cores. On the other hand this quantity increases for very large numbers of cores, revealing that the speedup becomes only moderate in this range. However this does not mean that the code is poorly parallelized. The reason is rather that, despite of the considerable size of the system, the number of degrees of freedom is so small that using too many cores simply results in a very small number of operations to be executed by a single core and thus to a poor ratio of computation and communication. Moreover it is more difficult to reach an efficient load balancing in such a situation, resulting in many cores being idle most of the time. Consequently this degradation in the parallel speedup will be shifted towards larger values when the system size is increased.

Nevertheless it is worth noting that even for the largest number of cores there is still some speedup, as can be seen from the total walltime decreasing steadily. Even in this range, the CPU time per atom remains of the same order as the one

needed at the crossover point with the cubic code. As calculations with the cubic BigDFT code are already accessible in this range (as pointed out in the previous section), production runs of very large systems become thus feasible by linearly scaling the computing resources needed at the crossover point. The memory parallelization, shown in the lower panel, does not suffer from any degradation, and we come close to a perfect scaling. In summary these results show that our code has an excellent level of parallelism as long as one keeps a good balance between the size of the system and the computational resources that are utilized.

## VI. Conclusion

Using a set of strictly localized and quasi-orthogonal support functions which are adapted *in situ* to their chemical environment we were able to develop a code which can perform accurate DFT calculations with a strict linear scaling with respect to the size of the system. Thanks to the compact support property of the underlying Daubechies wavelets, this reduction of the complexity with respect to the traditional cubic approach does not come at the cost of a loss of precision. Indeed, we are able to get an excellent level of accuracy with a set of standard input parameters and can thus get rid of the tedious fine tuning which is often needed for other  $\mathcal{O}(N)$  approaches. We believe this would considerably simplify the usage of this code by end-users, as this fine-tuning has so far restricted the usage of  $\mathcal{O}(N)$  approaches to a community of specialists.

Moreover we were able to drastically cut down the degrees of freedom needed for an accurate calculation, thus reducing the prefactor of the scaling law and consequently considerably diminishing the amount of computational resources required. Even for systems containing 10 000 atoms a complete and highly precise single point calculation can be done using only of the order of one thousand CPU hours. Furthermore the linear scaling behavior of the program motivates the notion of “CPU minutes per atom”, making it easy to estimate the computational resources needed for the simulation of a given system of any size. Together with the efficient parallelization scheme this low consumption of computational resources also paves the way towards more expensive tasks such as geometry optimizations and dynamics of large systems.

## Appendix A: parallelization of the sparse matrices

Some strategies to obtain an efficient parallelization have already been outlined in ref. 31. Here we present a new concept which we developed to reach an efficient handling of the sparse matrices.

The problem of the transposed approach<sup>31</sup> for calculating the overlap matrix (and similarly the Hamiltonian matrix) is that it requires a global reduction operation at the end. This can pose a bottleneck both from the viewpoint of runtime and memory. Moreover this global reduction is wasteful due to the

locality which is inherent to the linear scaling approach. Each MPI task only needs to know a small portion of the entire matrices and a global reduction is thus not necessary. To circumvent this issue, the MPI tasks are regrouped in taskgroups, sharing a portion of the entire matrix – obviously this portion is chosen such that it contains those parts of the global matrix which are actually needed by each individual task. In addition the taskgroups are defined such that each MPI task belongs to at most two taskgroups. This has two advantages: first of all each MPI task only needs to hold a copy of a part of the global matrix, thus reducing the memory requirements, and secondly the reduction only needs to be performed within a taskgroup. Thanks to the use of non-blocking collective MPI routines a task can participate in two reduction operations at the same time without the risk of serializing the communication scheduling.

The concept is visualized in Fig. 13 for a toy system consisting of 8 MPI tasks. The entire sparse matrix is given by a one-dimensional array which only contains the non-zero entries. The matrix subparts which are needed by each MPI task are indicated by the colored boxes; obviously, the union of all these subparts yields the entire matrix. The MPI tasks are initially regrouped in 3 taskgroups: taskgroup I is responsible for the part needed by the tasks 1–2, taskgroup II for those needed by the tasks 3–6, and taskgroup III for those needed by the tasks 7–8. The taskgroups are then enlarged in order to guarantee that the reduction is done correctly. Therefore taskgroup I includes the tasks 1–4, taskgroup II includes the tasks 2–8, and taskgroup III includes the tasks 5–8.

Obviously the toy system is too small to get a real benefit from the taskgroups. However for large systems there can be a substantial gain. As an example we show in Table 3 the timings obtained with and without the matrix taskgroups for two specific operations. As can be seen, the calculation of the overlap and Hamiltonian matrix (both computation and communication) can be accelerated by more than a factor of 3, demonstrating that actually most of the time was spent in the reduction and not in the computation itself. The other operation, which is required at the end of the FOE procedure to build up the density kernel out of the partial results calculated by

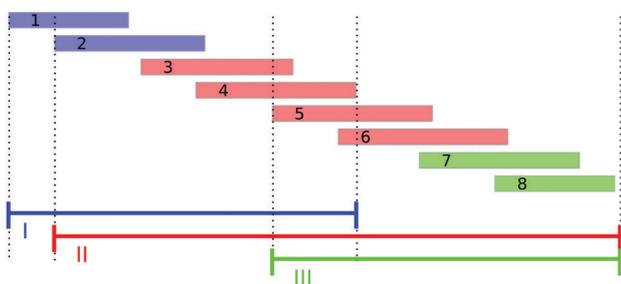


Fig. 13 Illustration of the concept of the matrix taskgroups for a toy system. There are 8 MPI tasks which are regrouped in 3 taskgroups. Each taskgroup is initially defined so as to form a matrix subset, and subsequently extended to include all the processes which share data of the initial subset. Care is taken in not including each MPI task in more than two different taskgroups.

Table 3 Speedups offered by the matrix taskgroups for two specific operations

	Overlap calculation seconds	Gather/compress seconds
Without taskgroups	88.9	266.9
With taskgroups	25.2	16.4

each task and is almost entirely communication based, can even benefit from a speedup of 16.

## Acknowledgements

S.M. acknowledges a postdoctoral fellowship from the CEA Nanosciences horizontal aid programme (project BigPOL). T.D. and L.G. are indebted to the French National Research Agency for funding under Contract No. ANR-2010-COSI-005-01 NEWCASTLE. Computing time has been provided by the “Tur-ing” national GENCI-IDRIS supercomputing center under contract No. i2014096905 and the Swiss CSCS project s499.

## References

- 1 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864.
- 2 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133.
- 3 X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, P. Ghosez, M. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. Verstraete, G. Zerah and J. Zwanziger, *Comput. Phys. Commun.*, 2009, **180**, 2582.
- 4 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.
- 5 M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark and M. C. Payne, *J. Phys.: Condens. Matter*, 2002, **14**, 2717.
- 6 J. E. Pask and P. A. Sterne, *Modell. Simul. Mater. Sci. Eng.*, 2005, **13**, R71.
- 7 L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman and R. Schneider, *J. Chem. Phys.*, 2008, **129**, 014109.
- 8 M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus and W. de Jong, *Comput. Phys. Commun.*, 2010, **181**, 1477.
- 9 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175.

- 10 W. Yang and T.-S. Lee, *J. Chem. Phys.*, 1995, **103**, 5674.
- 11 X.-P. Li, R. W. Nunes and D. Vanderbilt, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 10891.
- 12 S. Goedecker and L. Colombo, *Phys. Rev. Lett.*, 1994, **73**, 122.
- 13 S. Goedecker and M. Teter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1995, **51**, 9455.
- 14 S. Goedecker, *J. Comput. Phys.*, 1995, **118**, 261.
- 15 S. Goedecker, *Rev. Mod. Phys.*, 1999, **71**, 1085.
- 16 J. D. Cloizeaux, *Phys. Rev.*, 1964, **135**, A685.
- 17 J. D. Cloizeaux, *Phys. Rev.*, 1964, **135**, A698.
- 18 W. Kohn, *Phys. Rev.*, 1959, **115**, 809.
- 19 R. Baer and M. Head-Gordon, *Phys. Rev. Lett.*, 1977, **79**, 3962.
- 20 S. Ismail-Beigi and T. A. Arias, *Phys. Rev. Lett.*, 1999, **82**, 2127.
- 21 S. Goedecker, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 3501.
- 22 L. He and D. Vanderbilt, *Phys. Rev. Lett.*, 2001, **86**, 5341.
- 23 W. Hierse and E. B. Stechel, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17811.
- 24 E. Hernández and M. J. Gillan, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1995, **51**, 10157.
- 25 C.-K. Skylaris, P. D. Haynes, A. A. Mostofi and M. C. Payne, *J. Chem. Phys.*, 2005, **122**, 084119.
- 26 D. R. Bowler and T. Miyazaki, *J. Phys.: Condens. Matter*, 2010, **22**, 074207.
- 27 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103.
- 28 J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón and D. Sánchez-Portal, *J. Phys.: Condens. Matter*, 2002, **14**, 2745.
- 29 N. Marzari and D. Vanderbilt, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1997, **56**, 12847.
- 30 I. Daubechies, *Ten lectures on wavelets*, SIAM, 1992, vol. 61.
- 31 S. Mohr, L. E. Ratcliff, P. Boulanger, L. Genovese, D. Caliste, T. Deutsch and S. Goedecker, *J. Chem. Phys.*, 2014, **140**, 204110.
- 32 C. Hartwigsen, S. Goedecker and J. Hutter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 3641.
- 33 M. Krack, *Theor. Chem. Acc.*, 2005, **114**, 145.
- 34 A. Willand, Y. O. Kvashnin, L. Genovese, A. Vázquez-Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch and S. Goedecker, *J. Chem. Phys.*, 2013, **138**, 104109.
- 35 L. Lin, M. Chen, C. Yang and L. He, *J. Phys.: Condens. Matter*, 2013, **25**, 295501.
- 36 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- 37 P. Pulay, *Mol. Phys.*, 1969, **17**, 197, <http://www.tandfonline.com/doi/pdf/10.1080/00268976900100941>.
- 38 Y. M. Niquet, L. Genovese, C. Delerue and T. Deutsch, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 161301.
- 39 S. Mohr, P. Pochet, M. Amsler, B. Schaefer, A. Sadeghi, L. Genovese and S. Goedecker, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 041404.
- 40 I. Garg, H. Sharma, K. Dharamvir and V. Jindal, *J. Comput. Theor. Nanosci.*, 2011, **8**, 642.
- 41 M. Riad Manaa, H. a. Ichord and D. W. Sprehn, *Chem. Phys. Lett.*, 2003, **378**, 449.
- 42 D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, 1980, **45**, 566.
- 43 S. Krishnan, G. Brenet, E. Machado-Charry, D. Caliste, L. Genovese, T. Deutsch and P. Pochet, *Appl. Phys. Lett.*, 2013, **103**, 251904.
- 44 K. A. Wilkinson, N. D. M. Hine and C.-K. Skylaris, *J. Chem. Theory Comput.*, 2014, **10**, 4782.