Adaptive localization regions for O(*N*) density functional theory calculations

2008 J. Phys.: Condens. Matter 20 294210

(http://iopscience.iop.org/0953-8984/20/29/294210)

IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **20** (2008) 294210 (6pp) doi:10.1088/0953-8984/20/29/294210

Adaptive localization regions for O(N) density functional theory calculations

J-L Fattebert

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

E-mail: fattebert1@llnl.gov

Received 15 January 2008, in final form 5 February 2008
Published 24 June 2008
Online at stacks.iop.org/JPhysCM/20/294210

Abstract

A linear scaling approach for general and accurate pseudopotential density functional theory calculations is presented. It is based on a finite difference discretization. Effective O(N) scaling is achieved by confining the orbitals in spherical localization regions. To improve accuracy and flexibility while computing the smallest possible number of orbitals, we propose an algorithm for adapting localization regions during computation. Numerical results for a polyacetylene chain and a magnesium oxide ring are presented.

1. Introduction

One way to think about linear scaling approaches based on localized orbitals—such as the one described in this paper—is to consider those as linear combination of atomic orbitals (LCAO) type methods with flexible orbitals, adapted to their environment. Localized orbitals are expanded in a generic numerical basis set which can be systematically improved to achieve a prescribed accuracy.

The plane waves (PW) method is an example of a systematically improvable basis set commonly used in pseudopotential density functional theory (DFT) simulations, in particular when one wants to avoid any bias due to the numerical basis set, when high numerical accuracy is needed or when dealing with systems such as metals with almost free electrons. The numerical solution of the Kohn–Sham equations—eigenstates—is expanded directly in terms of M elementary plane waves and the accuracy of the solution can be systematically improved by enlarging the basis set. In contrast, linear combination of atomic orbitals (LCAO) type methods use a numerical basis set made of atomic-like orbitals specific to the physical system that one tries to simulate and the solution is expressed as a linear combination of those. In the PW approach, because the numerical basis set is quite a bit larger than in the LCAO, we can limit the calculation to just the few eigenstates that we are interested in—N occupied states. In that case, the major computational work is in determining these eigenstates—O($M \cdot N^2$) scaling with $M \gg N$—and the work in the N-dimensional subspace defined by those states—such as inverting the overlap matrix or determining the occupation numbers by diagonalizing an $N \times N$ matrix—becomes quite light in comparison (O(N^3)). LCAO approaches in contrast can be viewed as methods dealing with a fixed and predetermined subspace, of dimension an order of magnitude larger than N, in which one needs to determine the occupation in the form of a single-particle density matrix.

In this paper we focus on an O(N) complexity algorithm as an alternative to PW, sharing some of the goals and features with this reference approach, in particular using an unbiased and systematically improvable numerical basis set, and calculating the minimal number of electronic states needed in this basis set. Note that the PW basis set is just one example of a systematically improvable and unbiased numerical basis set. Alternative approaches such as finite differences or finite elements achieve the same goal, each with their pros and cons, and have been used successfully applied in DFT simulations [1–5]. In this paper we will use the finite difference approach as a discretization scheme. The algorithm described below could be easily combined with other real-space discretizations however. To reduce computational complexity we will introduce spherical localization regions (LR) in which the electronic states will be confined. Each state is described by values at each point of the finite difference grid inside a given localization region, i.e. by O(1) degrees of freedom. To keep the number of electronic states in the calculation as small as possible, we will allow the localization regions to adapt their locations and sizes according to their environment. Looking at localized orbital linear scaling approaches as LCAO methods with a basis of flexible orbitals

adapted to their environment, one realizes that one may be able to reduce the number of those orbitals by making them more flexible and still describe accurately N electronic wavefunctions. In this paper we explore the limit of this idea: including in the calculation a number of localized orbitals equal to the number N of electronic states that we want to compute accurately.

Note that the idea of representing the electronic structure using floating orbitals was already proposed 40 years ago with the floating spherical Gaussian orbital model [6]. In this simple model, only four parameters per orbital—center coordinates and radius—are optimized to minimize the total energy, while the Gaussian shape is fixed. In that perspective the model described in this paper additionally optimizes the shape of the orbitals, making use of today's much larger computer power and leading to a much greater accuracy.

Numerous ideas have been proposed in the last 15 years to achieve linear scaling in electronic structure calculations (see for instance [7]). A few of those approaches are closely related to our methodology and essentially try to achieve the same goal, that is linear scaling with PW accuracy using localized orbitals. Following the idea of Hernandez and Gillan [8], the CONQUEST code was developed using B-spline finite elements as the underlying numerical basis [9]. Note that this code also offers a fixed orbital LCAO-like option. A similar approach, but using a plane wave-like basis set for localized orbitals was implemented in the ONETEP code [10]. Tsuchida and Tsukada [11] on the other hand used a finite elements discretization in conjunction with the Kim–Mauri–Galli functional. In contrast to the methodology presented below, those techniques, as well as the technique described in [12], use fixed localization regions centered on atoms. This lack of flexibility is compensated for by a larger number of localized orbitals.

In this paper we focus on the problem of defining appropriate localization regions in which to confine localized orbitals. We recently proposed an algorithm for automatically adapting the centers of localization regions. It essentially boils down to a regular update of the localization centers using the center of charge of the confined orbital as the new target. Such an approach presents several advantages over using fixed localization regions. In particular, accuracy is improved by optimizing the positions of localization regions. It also enables molecular dynamics simulations without spurious Pulay forces [13, 14]. In this paper we extend this idea and explore a new algorithm to adapt not only the position of the localization regions, but also their sizes. This is motivated by the fact that calculations going beyond the ground state of insulators often present some less localized maximally localized generalized Wannier functions (MLGWF), or a less peaked distribution of their spreads. Also in complex and multi-species systems, one can expect anomalous shapes for some MLGWF as observed for instance for amorphous silicon [15]. Another area where size adaptation could be quite useful is in studying systems for which electronic structure properties vary significantly with system size and for which determining *a priori* the size of LR based on error measurement on smaller reference systems may not be adequate.

2. Computational method

2.1. Linear scaling strategy

Our approach is based on the idea proposed in [12] and later extended in [13, 14]. The first step to achieving linear scaling is to express the DFT energy functional in terms of a set of general non-orthogonal orbitals $\{\phi_i\}_{i=1}^N$ [16]:

$$\begin{aligned} E_{\mathrm{KS}}[\{\phi_i\}_{i=1}^N] = &\sum_{i,j=1}^{N} K_{ij} \int_\Omega \phi_i(\mathbf{r}) \left(-\nabla^2\right) \phi_j(\mathbf{r})\, \mathrm{d}\mathbf{r} \\ &+ \frac{1}{2} \int_\Omega \int_\Omega \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{r}_2 + E_{\mathrm{XC}}[\rho] \\ &+ \sum_{i,j=1}^{N} 2K_{ij} \int_\Omega \phi_i(\mathbf{r}) (V_{\mathrm{ext}}\phi_j)(\mathbf{r})\, \mathrm{d}\mathbf{r} \end{aligned} \tag{1}$$

where ρ is the electronic density

$$\rho(\mathbf{r}) = 2 \sum_{i,j=1}^{N} K_{ij} \phi_i(\mathbf{r}) \phi_j(\mathbf{r}). \tag{2}$$

The $N \times N$ matrix K is a single-particle density matrix which expresses the occupation of each electronic state for a set of general non-orthogonal orbitals. If only the occupied valence states are being included in the computation, it is simply S^{-1}, the inverse of the overlap matrix. The energy E_{XC} models the exchange and correlation between electrons.

Given an external potential V_{ext}—defined by the various atomic species, their respective positions and pseudopotentials—the ground state of the physical system is obtained by minimizing the energy functional (1). We discretize this energy functional on a uniform real-space mesh. We represent the orbitals, the electronic density and the potential by their values at each node of the mesh. To approximate the Laplacian operator, we use the Mehrstellen finite difference scheme [2].

To achieve an effective linear scaling, localization constraints on the orbitals are added to this minimization procedure. Namely we impose that each orbital strictly vanishes outside a specific spherical localization region. We also adapt the minimization algorithm to satisfy the constraints along the way. Our minimization procedure is based on the steepest descent directions preconditioned by multigrid. The process is accelerated by a nonlinear extrapolation scheme [13]. Note that we compute the matrix K in $\mathrm{O}(N^3)$ operations using a standard diagonalization procedure for an $N \times N$ matrix. Because N is kept small by adapting the localization regions (see section 2.3), operations on $N \times N$ matrices remain relatively cheap for the range of values that we are currently targeting (N up to 5000)

Restricting orbitals to confinement regions introduces some approximation since there is in general no solution to our original problem which satisfies exactly the localization constraints. This truncation error however decays rapidly with the size of the localization regions. We observed numerically an exponential convergence for both the energy and forces in various systems [13]. By choosing localization regions large enough one can even reduce truncation error to a

quantity smaller than the discretization. The minimum size of localization regions necessary for achieving a prescribed accuracy directly affects the crossover point, that is the minimal system size for which any gain can be achieved by an O(N) approach. Also, because the energy functional is no longer invariant under rotations inside the occupied space in the presence of localization constraints, local minima are possible and have been observed [14].

2.2. Maximally localized generalized Wannier functions

The concept of maximally localized generalized Wannier functions (MLGWF) is very useful in justifying and understanding O(N) approaches based on orbital localization. Given a position operator $\hat{X}$, one can define the center of charge associated with $\hat{X}$ for an orbital ϕ as

$$\bar{X}(\phi) = \langle\phi|\hat{X}|\phi\rangle. \tag{3}$$

The spread associated with the operator $\hat{X}$ for the orbital ϕ is defined by

$$\sigma_{\hat{X}}(\phi) = \left[\langle\phi|\left(\hat{X} - \langle\phi|\hat{X}|\phi\rangle\right)^2|\phi\rangle\right]^{1/2} \tag{4}$$

and quantifies the degree of localization of ϕ around its center of charge $\bar{X}(\phi)$.

The spread functional

$$\sigma^2_{\hat{X}}(\{\phi_i\}_{i=1}^N) = \sum_{i=1}^N \sigma^2_{\hat{X}}(\phi_i) \tag{5}$$

expresses the total spread for the N orbitals $\{\phi_i\}_{i=1}^N$.

Now given a set of orthonormal functions $\{\psi_i\}_{i=1}^N$, we can define an orthogonal transformation

$$\phi_i = \sum_{j=1}^N a_{ij}\psi_j$$

where a_{ij} are the matrix elements of an orthogonal matrix A. Finding a matrix A which minimizes the sum of the spread functional (5) associated with m position operators, positions operators for direction x, y, z typically,

$$\sigma^2_{\{\hat{X}^{(k)}\}_{k=1}^m}(\{\phi_i\}_{i=1}^N) = \sum_{k=1}^m \sigma^2_{\hat{X}^{(k)}}\left(\{\phi_i\}_{i=1}^N\right) \tag{6}$$

can be formulated as a simultaneous diagonalization problem and efficiently solved using the algorithm proposed in [17]. The functions which minimize this spread functional—not unique in general—are by definition the maximally localized generalized Wannier functions (MLGWF) introduced by Marzari and Vanderbilt [18]. For periodic boundary conditions, the position operator for extended systems proposed by Resta [19] should be used. To avoid introducing complex numbers, we use the equivalent six trigonometric operators proposed in [17] ($m = 6$).

Being able to find a representation of the electronic subspace in terms of localized orbitals such as MLGWF justifies O(N) algorithms based on imposing localization constraints on orbitals. This is illustrated in figure 1 where MLGWF and localized orbitals obtained by minimization under localization constraints are plotted side by side for the C_2H_4 molecule. Their similarity is striking, even if MLGWF are orthogonal while localized orbitals are not.

2.3. Adaptive localization regions

To get a linear scaling algorithm through orbital localization, one has to choose appropriately the localization regions (LR) in which the orbitals are going to be confined. If the dimension of the subspace spanned by the set of localized orbitals is much larger than the number of occupied states N, that choice is less important since the larger number of degrees of freedom can compensate for the lack of accuracy of the underlying orbitals. The most straightforward solution is then to have orbitals localized in regions centered on atoms in numbers large enough for describing at least the valence states around each atom. Using such an approach, one may end up with trial subspaces that are much larger than N. Techniques usually used in LCAO algorithms can then be used to reduce complexity in determining the single-particle density matrix X and fill up the occupied states [8]. Resulting unoccupied states usually have the role of a numerical buffer and the density of states for the highest energies is not to be considered physical [20].

Our current approach is different. As in the standard O(N^3) PW approach, we solve for and represent only the states that we are interested in. To do that, we let the localization regions adapt to their environment, optimizing both their location and their size during a calculation. On the basis of the analogy between MLGWF and localized orbitals, we use the centers of charge defined by (3) and spreads of the localized orbitals defined by (4) to iteratively update the localization regions. LR centers are moved towards centers of charge, while localization radii are rescaled to be proportional to the spread of the confined orbitals. To avoid a diverging total spread during the optimization process, the total volume of all the LR is kept fixed to a constant value V_0.

In summary, localization regions are adapted following this simple scheme:

Until convergence, do:

- Iteratively minimize functional (1) for fixed LR for a fixed number of iterations or until convergence to a certain tolerance is reached.
- For each orbital $\phi_i, i = 1, \ldots, N$, compute the m components of the centroids of charge,

$$Q_i^{(k)} = \langle\phi_i|\hat{X}^{(k)}|\phi_i\rangle, \qquad k = 1, \ldots, m,$$

and the spreads

$$\sigma_i = \left[\sum_{k=1}^m \langle\phi_i|\left(\hat{X}^{(k)} - Q_i^{(k)}\right)^2|\phi_i\rangle\right]^{1/2}.$$

- Move LR centers to $\mathbf{Q}_i$.
- Set LR radii to $R_c^{(i)} = \alpha\sigma_i$ with α such that $(4\pi/3)\sum_{i=1}^N (R_c^i)^3 = V_0$.

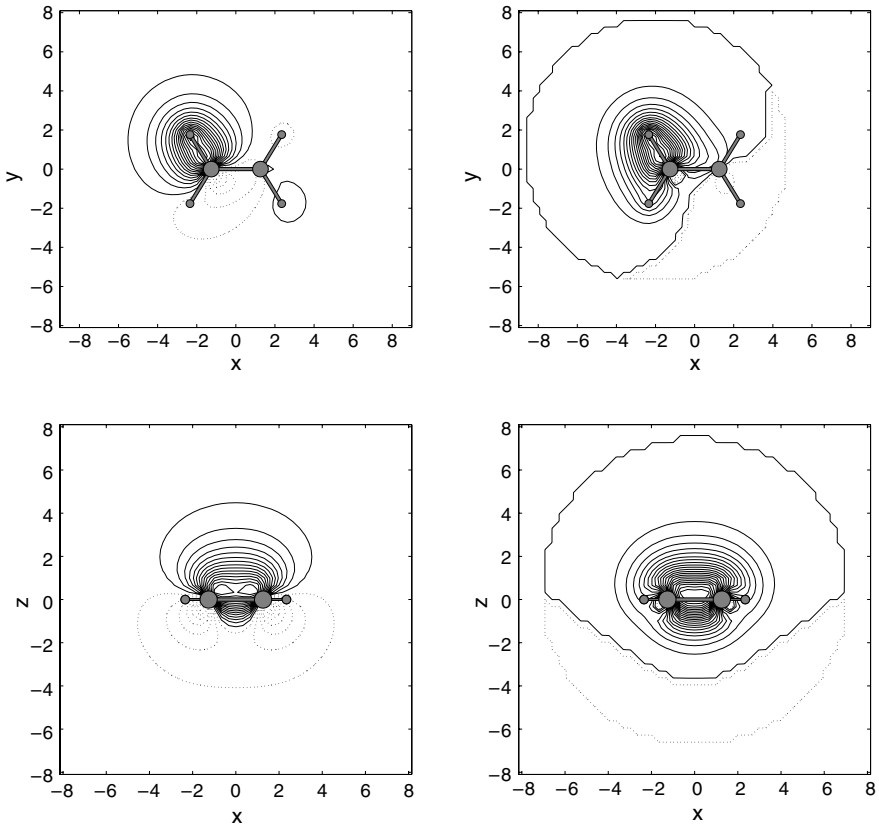

Figure 1. Two types of MLGWF in C_2H_4 (left) and the corresponding strictly localized orbitals obtained by O(N) scheme (right). Isovalues are orbital dependent. Dotted lines correspond to negative isovalues. Isovalues for localized orbitals are chosen in such a way that the limit of the localization regions appears clearly.

In a molecular dynamics simulation or geometry optimization, just one cycle of this algorithm is usually sufficient at every ionic step. But multiple iterations are required for a new configuration with no good initial guess.

3. Numerical results

To illustrate the algorithms described in this paper, we use a polyacetylene chain $(C_2H_2)_8$ as a test system. This is a convenient system since it contains a small number of electrons but is large enough (periodic cell of length 37.12 bohr along the chain axis) to fully contain all the localization regions in our test and include totally disconnected ones. It allows affordable O(N^3) reference calculations for comparing with results from the approximative O(N) scheme. We use the same geometry as is used in other theoretical studies [21]. This system is made up of 40 fully occupied valence states. We also consider the lowest three empty conduction states which are fully localized on the molecule. We use the local density approximation (LDA) to model the exchange and correlation term.

We start by computing MLGWF for the subspace spanned by the 40 fully occupied states as well as for the subspace including three additional (unoccupied) states. Results are shown in figure 2. The result obtained in the first case is quite natural: two MLGWF on each C=C bond, one on each C–C and C–H bond. When including three conduction states, the situation becomes more complicated.

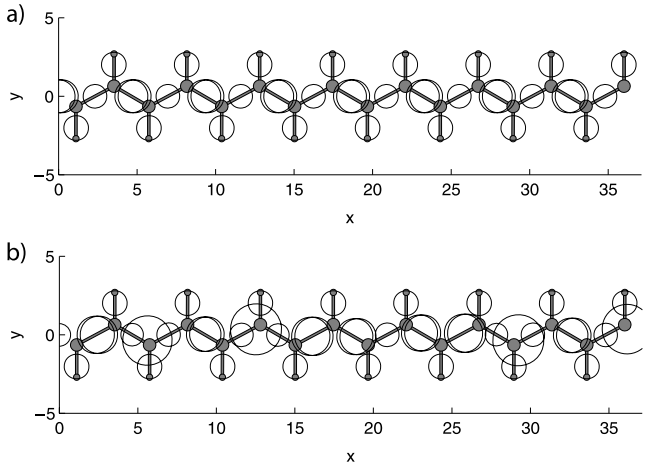

Figure 2. Projection in molecular plane of MLGWF for polyacetylene chain $(C_2H_2)_8$. Open circles have radii equal to half their MLGWF spread and are centered at Wannier centers. Circles with coincidental projections have been shifted for a better display. (a) Number of MLGWF equal to number of fully occupied states (40). (b) Same calculation with three additional empty states.

For the calculations with adaptive localization regions, we first compute the ground state of the system without any unoccupied states. LRs are centered at Wannier centers and all have the radius of 7 bohr which is large enough for obtaining

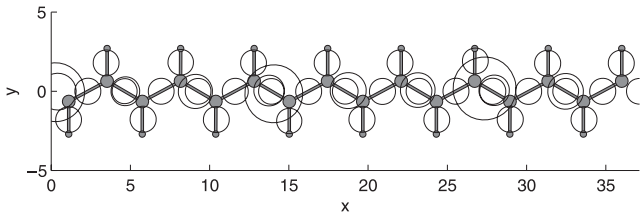

Figure 3. Positions and spreads of adaptive localized orbitals in polyacetylene chain $(C_2H_2)_8$. Open circles have radii equal to half the spreads.

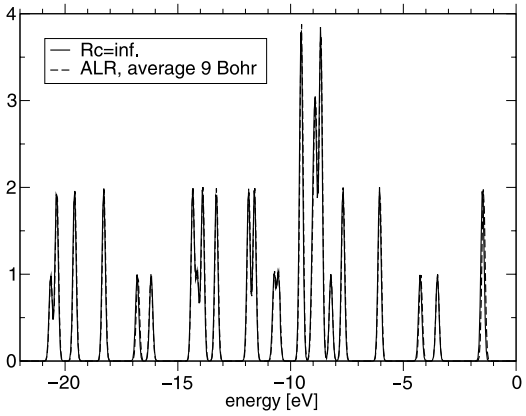

Figure 4. Density of states for polyacetylene chain $(C_2H_2)_8$ with three unoccupied states. Comparison between $O(N^3)$ calculation and adaptive localization regions approach with an average localization radius of 9 bohr.

an accurate density of states. Using the electronic density from that first calculation, we compute the Kohn–Sham potential and freeze it before adding three additional states. Three additional LRs associated with these states are also added and centered on three carbon atoms as uniformly as possible. We then run our adaptive localization algorithm with a total volume V_0 corresponding to an average localization radius of 9 bohr. This leads to localization radii that range from 7 to 17 bohr, thus allowing wider spreads for some states while preserving the accuracy of the occupied states. Positions and spreads of optimized orbitals are shown in figure 3. The resulting density of states is plotted in figure 4. It is almost indistinguishable from the reference $O(N^3)$ calculation. The error on the band gap is 0.022 eV. The localization radii are shown in figure 5. The localized orbital with the largest spread is shown in figure 6. It is still quite localized with respect to the global system, but spreads over five atoms unlike in systems with no unoccupied stages for which MLGWF typically spread over one bond/two atoms.

As a second numerical example, we consider a magnesium oxide ring $(MgO)_8$. For this application, we use the PBE exchange and correlation functional. In our pseudopotential approximation, we treat the semicore 2p states of the Mg atoms explicitly which leads to a total of 56 doubly occupied orbitals for the 16-atom ring. One major difference between this system and the previous example is the ionic nature of the chemical bonds between oxygen and magnesium atoms. The ratio between maximum and minimum spreads of MLGWF

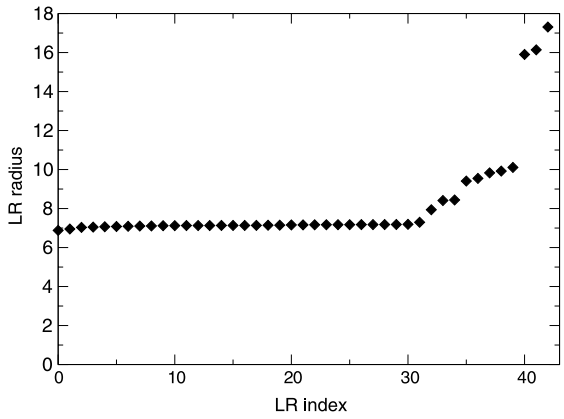

Figure 5. Polyacetylene chain $(C_2H_2)_8$: radii distribution for LRs with three unoccupied states after adaptation. The average localization radius was 9 bohr.

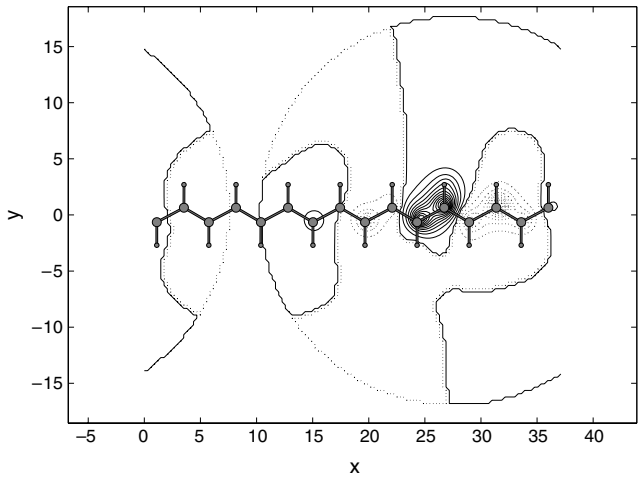

Figure 6. Contour plot of adapted localized orbitals with widest spread for polyacetylene chain $(C_2H_2)_8$ calculation with three unoccupied states and an average localization radius of 9 bohr.

is about two when considering just the occupied states. The ground state computation of this system can benefit from adaptive localization regions. Localization regions obtained for this system are plotted in figure 7. Table 1 shows the error measured on the energy and forces for calculations with three different localization sizes. It shows in particular an exponential decay of the error with the average radius of the localization regions.

4. Concluding remarks

The finite difference approach provides an appropriate framework in which to achieve plane wave accuracy and linear scaling using localized orbitals. Flexible adaptive localization regions enable calculation of the states of interest only, without the need to use a larger subspace including many unoccupied states with no physical significance. Adapting positions seems more important than adapting sizes for simple problems where most MLGWF have similar spreads. Localization region adaptivity in size becomes however important in complex

Table 1. Numerical results for $(MgO)_8$ ring: accuracy and range of localization radius in atomic units (au). The error is defined as the difference with the results obtained without localization constraints on the electronic wavefunctions.

Average adaptive radius	6 (bohr)	7 (bohr)	8 (bohr)
Error energy (Ha/atom)	3.4×10^{-3}	9.6×10^{-4}	3.1×10^{-4}
Average error forces (au)	2.7×10^{-3}	1.1×10^{-3}	3.5×10^{-4}
Range localization radius (bohr)	4.0–9.5	5.0–10.1	5.5–11.7

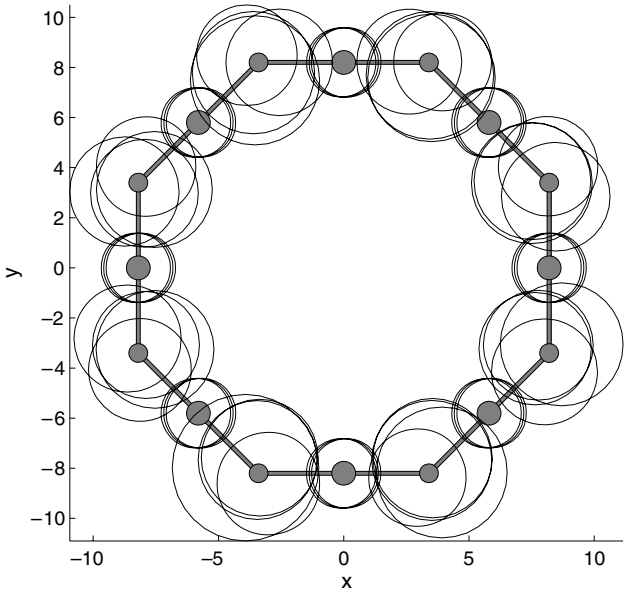

Figure 7. Projection in the molecular plane of positions and radii of adaptive localized orbitals (open circles) in $(MgO)_8$ ring for an average localization radius of 8 bohr. Radii have been rescaled by a factor 0.5 for a better display.

multi-species systems or when some conduction states need to be computed accurately.

While we have a good general understanding of the electronic structure representation in terms of MLGWF for the valence states in insulating systems, the situation is more complicated when we go beyond ground state calculations of perfect crystals or organic molecules. Having a better understanding of the electronic structure representation by a set of localized orbitals such as the MLGWF in general may be key for designing efficient linear scaling algorithms applicable to general and complex systems. Finally, error estimators functions of truncation radii would be extremely useful in an automatic LR adaptation scheme, in particular for systems for which electronic structure properties vary significantly with system size and for which determining *a priori* the size of LR based on error measurement on smaller reference systems may not be adequate.

Acknowledgments

This work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Financial support from US Department of Energy/SciDAC project on Quantum Simulation of Materials and Nanostructures is gratefully acknowledged. The author also thanks E W Draeger for his critical reading of the manuscript.

References

[1] Chelikowsky J R, Troullier N and Saad Y 1994 Finite-difference-pseudopotential method: electronic structure calculations without a basis *Phys. Rev. Lett.* **72** 1240–3
[2] Briggs E L, Sullivan D J and Bernholc J 1996 Real-space multigrid-based approach to large-scale electronic structure calculations *Phys. Rev.* B **54** 14362–75
[3] Tsuchida E and Tsukada M 1995 Electronic-structure calculations based on the finite-element method *Phys. Rev.* B **52** 5573–8
[4] Pask J E and Sterne P A 2005 Finite elements methods in *ab initio* electronic structure calculations *Modelling Simul. Mater. Sci. Eng.* **13** 71–96
[5] Fattebert J L, Hornung R D and Wissink A M 2007 Finite element approach for density functional theory calculations on locally-refined meshes *J. Comput. Phys.* **223** 759–73
[6] Frost A A 1967 Floating spherical gaussian orbital model of molecular structure. I. Computational procedure. LiH as an example *J. Chem. Phys.* **47** 3707
[7] Goedecker S 1999 Linear scaling electronic structure methods *Rev. Mod. Phys.* **71** 1085–123
[8] Hernandez E and Gillan M J 1995 Self-consistent first-principles technique with linear scaling *Phys. Rev.* B **51** 10157–60
[9] Bowler D R, Choudhury R, Gillan M J and Miyazaki T 2006 Recent progress with large-scale *ab initio* calculations: the CONQUEST code *Phys. Status Solidi* b **243** 989
[10] Skylaris C-K, Haynes P D, Mostofi A A and Payne M C 2005 Introducing ONETEP: linear-scaling density functional simulations on parallel computers *J. Chem. Phys.* **122** 084119
[11] Tsuchida E and Tsukada M 1998 Large-scale electronic-structure calculations based on the adaptive finite-element method *J. Phys. Soc. Japan* **67** 3844–58
[12] Fattebert J-L and Bernholc J 2000 Towards grid-based O(N) density-functional theory methods: optimized non-orthogonal orbitals and multigrid acceleration *Phys. Rev.* B **62** 1713–22
[13] Fattebert J-L and Gygi F 2004 Linear scaling first-principles molecular dynamics with controlled accuracy *Comput. Phys. Commun.* **162** 24–36
[14] Fattebert J-L and Gygi F 2006 Linear-scaling first-principles molecular dynamics with plane-waves accuracy *Phys. Rev.* B **73** 115124
[15] Fornari M, Marzari N, Peressi M and Baldereschi A 2001 Wannier function characterization of floating bonds in a-Si *Comput. Mater. Sci.* **20** 337
[16] Galli G and Parrinello M 1992 Large scale electronic structure calculations *Phys. Rev. Lett.* **69** 3547–50
[17] Gygi F, Fattebert J-L and Schwegler E 2003 Computation of maximally localized Wannier functions using a simultaneous diagonalization algorithm *Comput. Phys. Commun.* **155** 1–6
[18] Marzari N and Vanderbilt D 1997 Maximally localized generalized Wannier functions for composite energy bands *Phys. Rev.* B **56** 12847
[19] Resta R 1998 Quantum-mechanical position operator in extended systems *Phys. Rev. Lett.* **80** 1800–3
[20] Skylaris C-K and Haynes P D 2007 Achieving plane wave accuracy in linear-scaling density functional theory applied to periodic systems: a case study on crystalline silicon *J. Chem. Phys.* **127** 164712
[21] Rohra S, Engel E and Görling A 2006 Exact-exchange Kohn–Sham formalism applied to one-dimensional periodic electronic systems *Phys. Rev.* B **74** 045119