PHYSICAL REVIEW B VOLUME 53, NUMBER 16 15 APRIL 1996-II

Self-consistent order-*N* density-functional calculations for very large systems

Pablo Ordejón
Department of Physics and Materials Research Laboratory, University of Illinois, Urbana, Illinois 61801
and Departamento de Física, Universidad de Oviedo, C/ Calvo Sotelo s/n, 33007 Oviedo, Spain

Emilio Artacho and José M. Soler
Instituto de Ciencia de Materiales Nicolás Cabrera and Departamento de Física de Materia Condensada,
C-III Universidad Autónoma de Madrid, 28049 Madrid, Spain

(Received 29 December 1995)

We present a method to perform fully self-consistent density-functional calculations that scales linearly with the system size and which is well suited for very large systems. It uses strictly localized pseudoatomic orbitals as basis functions. The sparse Hamiltonian and overlap matrices are calculated with an $O(N)$ effort. The long-range self-consistent potential and its matrix elements are computed in a real-space grid. The other matrix elements are directly calculated and tabulated as a function of the interatomic distances. The computation of the total energy and atomic forces is also done in $O(N)$ operations using truncated, Wannier-like localized functions to describe the occupied states, and a band-energy functional which is iteratively minimized with no orthogonality constraints. We illustrate the method with several examples, including carbon and silicon supercells with up to 1000 Si atoms and supercells of β-C_3N_4. We apply the method to solve the existing controversy about the faceting of large icosahedral fullerenes by performing dynamical simulations on C_{60}, C_{240}, and C_{540}.

A large effort has been devoted in the last few years to develop approximate methods to solve the electronic structure of large systems with a computational cost proportional to its size.[1] Several approaches are now sufficiently accurate and robust to obtain reliable results for systems with thousands of atoms. So far, however, most of these schemes have been useful only with simple Hamiltonians, like empirical tight-binding models, which provide an ideal setting for order-*N* calculations. First-principles order-*N* calculations have been performed mainly in the non-self-consistent Harris functional version[2] of the local-density approximation (LDA) for electronic exchange and correlation (XC) using minimal bases.[1,3] Linear scaling algorithms in fully self-consistent LDA have also been tried,[4] but the results are far from the linear scaling regime, due to the relatively small number of manageable atoms in those simulations. Hernandez and Gillan[5] have successfully produced LDA results in large silicon systems using a real-space grid method. The computational requirements that this kind of approach demands are, however, extremely large, and calculations must be performed in massive computational platforms.

We have developed a self-consistent density-functional formulation with linear scaling, capable of producing results for very large systems, whose computational demands are not overwhelmingly large, so that systems with many hundreds of atoms can be treated in modest computational platforms like work stations, and much larger systems can be treated in massive platforms. The method is based on the linear combination of atomic orbitals (LCAO) approximation as the basis of expansion of the electronic states. Nonorthogonal LCAO bases are very efficient, reducing the number of variables dramatically, compared to plane-wave (PW) or real-space grid approaches, so that larger systems can be studied. Also, LCAO can provide extremely accurate bases, always staying in the range of a few valence orbitals per atom.[6] As a first step, in this work we use minimal basis sets of one *s* and three *p* orbitals per atom, the extension to larger bases being perfectly possible within the present formulation. The choice of a basis obviously implies a possible error associated to its incompleteness. In the same way as for the error due to the linear scaling algorithm, the error in the basis can be reduced at the expense of an increase in computational effort. Its magnitude should be carefully checked, but also compared with other sources of error to ensure that an increase of the basis is really worthwhile.

Our method uses standard LDA techniques[7] for the valence electrons, the core electrons being replaced by pseudopotentials.[8] The basis orbitals used in this work are the *s* and *p* pseudoatomic orbitals defined by Sankey and Niklewski,[9] in the context of non-self-consistent Harris functional methods. These are slightly excited pseudoatomic orbitals $\phi_\mu(\mathbf{r})$, obtained by solving the valence electron problem in the isolated atom, with the same pseudopotential and LDA approximations, and with the boundary condition that the atomic orbitals are strictly localized, vanishing outside a given radius r_c. This radius cutoff is in principle orbital dependent, but we do not make explicit this dependence in the equations only for simplicity in the notation. The great advantage of these orbitals is that they give rise to sparse overlap and Hamiltonian matrices (since matrix elements between distant orbitals exactly vanish) and they display the same structure as in conventional tight binding. The extent of the interactions and the sparseness of the matrices depend on the cutoff radius r_c of each atom. These are not critical as long as the maxima of the atomic wave functions are well within r_c. For an analysis of the quality of pseudoatomic orbitals as a basis for solid state calculations we refer the reader to Ref. 10.

0163-1829/96/53(16)/10441(4)/$10.00

The LDA Hamiltonian matrix elements for a given particle density are obtained using a combination of techniques, adopting the most convenient one for each term of the Hamiltonian. In a prior step, to avoid dealing with the long range of the pseudopotentials, we rewrite the Kohn-Sham Hamiltonian by adding and subtracting the Hartree potential created by the neutral-atom charge $n_0(\mathbf{r})$, defined as

$$n_0(\mathbf{r})=\sum_i n_i^{\mathrm{NA}}(\mathbf{r}-\mathbf{R}_i), \tag{1}$$

where i runs over the atoms in the system, and n_i^{NA} is the spherical atomic charge density of the atom i in its neutral, isolated state with ρ_μ^0 electrons on each orbital ϕ_μ. If we define $\delta n(\mathbf{r})=n(\mathbf{r})-n_0(\mathbf{r})$, where $n(\mathbf{r})$ is the actual charge density, the Hartree potential can be decomposed into two contributions V_H^δ and V_H^0, created by $\delta n(\mathbf{r})$ and $n_0(\mathbf{r})$, respectively. Using Eq. (1), V_H^0 can be expressed as a sum of atomic contributions. Also, the pseudopotential is decomposed into a short-range nonlocal term V_{NL} and a long-range local term V_L. Following Sankey and Niklewski,[9] we define the neutral atom potential of a given atom at $\mathbf{R}_i$ as

$$V_{\mathrm{NA}}(\mathbf{r}-\mathbf{R}_i)=V_L(\mathbf{r}-\mathbf{R}_i)+e^2\int\frac{n_i^{\mathrm{NA}}(\mathbf{r}-\mathbf{R}_i)}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'. \tag{2}$$

V_{NA} is short ranged, since the core attraction and the electron Coulomb repulsion of the neutral atom charge cancel each other beyond r_c. The Kohn-Sham Hamiltonian is finally obtained as

$$\begin{aligned}H^{\mathrm{KS}}=&\frac{p^2}{2m}+\sum_i\left[V_{\mathrm{NL}}(\mathbf{r}-\mathbf{R}_i)+V_{\mathrm{NA}}(\mathbf{r}-\mathbf{R}_i)\right]\\&+V_H^\delta(\mathbf{r})+V_{\mathrm{XC}}(\mathbf{r}).\end{aligned} \tag{3}$$

The overlap, kinetic energy term, neutral atom potential, and nonlocal part of the pseudopotential, are all independent of the charge density $n(\mathbf{r})$, and their matrix elements between atomic orbitals can be expressed as sums of two-center $[S_{\mu\nu}=\langle\phi_\mu|\phi_\nu\rangle$ and $\langle\phi_\mu|p^2/2m|\phi_\nu\rangle]$ or three-center $[\langle\phi_\mu|V_{\mathrm{NL}}(\mathbf{r}-\mathbf{R}_i)|\phi_\nu\rangle$ and $\langle\phi_\mu|V_{\mathrm{NA}}(\mathbf{r}-\mathbf{R}_i)|\phi_\nu\rangle]$ integrals, which only depend on the relative positions of pairs or triplets of atoms. We follow the method proposed by Sankey and Niklewski[9] to compute all these integrals: they are calculated beforehand and tabulated as a function of the relative position of the centers. These tables are used during the simulation, to calculate all the nonzero integrals by interpolation. The details of the procedure can be found in Ref. 9. Since all these integrals are zero for distant enough atoms, their number scales linearly with the size of the system, as well as the computation time. The contributions of these terms to the Hamiltonian are computed only once for a given atomic configuration, since they do not depend on the self-consistent charge.

The matrix elements of the Hartree potential $V_H^\delta(\mathbf{r})$ created by the charge $\delta n(\mathbf{r})$ and the exchange-correlation potential $V_{\mathrm{XC}}[n(\mathbf{r})]$ both depend on the self-consistent charge. To calculate these integrals we compute $n_0(\mathbf{r})$, $n(\mathbf{r})$, and $\delta n(\mathbf{r})$, for a given LCAO density matrix, at the points of a regular grid in real space. This is straightforward since the basis orbitals are defined in real space. Poisson's equation for the Hartree potential can then be solved by the standard fast Fourier transform (FFT) method, assuming a supercell geometry, or by the multigrid method.[11] In spite of its $N\ln N$ scaling, we presently use FFT's for simplicity, since this part represents a minor contribution to the total computational load. Note that only two FFT's are necessary per cycle of self-consistency (SCF cycle), in contrast with PW-based calculations, where an FFT is required for each electronic state. The LDA XC potential is trivially computed on each point of the grid. Once the value of the Hartree and the XC potentials are known at every point, the integrals $\langle\phi_\mu|V_H^\delta|\phi_\nu\rangle$ and $\langle\phi_\mu|V_{\mathrm{XC}}|\phi_\nu\rangle$ are computed by direct summation on the grid. These sums are carefully done to minimize the amount of numerical workload involved. Only the nonzero integrals (between orbitals on atoms closer than $2r_c$) are computed, and only the points of the grid for which both orbitals are nonzero contribute to each integral. We use sparse-matrix multiplication techniques optimized for this class of operations. As a result, the computation of the integrals scales linearly with the size of the system.

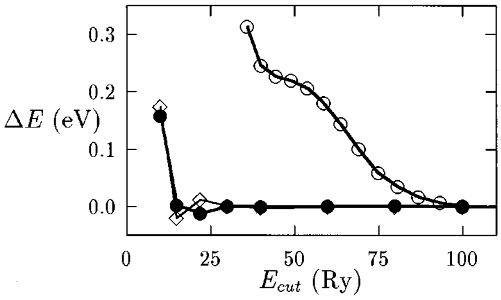

FIG. 1. Convergence of the total energy per carbon atom vs grid fineness (given by the cutoff E_{cut} of the plane waves that it can represent). The results of the present method are shown for a diamond supercell with 64 atoms (full circles) and for a C_3 cluster (diamonds). Open circles show results of conventional plane-wave calculations for diamond (Ref. 12).

It is important to stress that the convergence with grid spacing of our method is different from that in standard PW calculations, which are known to require large PW cutoffs for systems containing atoms with hard pseudopotentials. In Fig. 1 we show the convergence of the total energy per atom (referred to the converged value) for carbon, as a function of E_{cut}, the kinetic energy cutoff of the plane waves that the grid can represent. Full circles are for a diamond supercell of 64 atoms, whereas diamonds are for a cluster of three carbon atoms in a supercell of $15\times15\times15$ Å^3. In both cases, the results are converged to below 2 meV/atom for a cutoff of 30 Ry. This is in sharp contrast with results of PW calculations[12] (open circles), in which the cutoff necessary to achieve convergence (with the same pseudopotential) is much higher. Note, moreover, that the energy cutoff in our case refers to the grid representation of the charge density, whereas in the PW case it refers to the wave functions, which implies an even higher (four times) cutoff in the charge density. The reason for the fast convergence of our approach is that most of the Hamiltonian terms (most importantly the kinetic energy and the neutral atom potential) are not computed in the grid.

Once the Kohn-Sham Hamiltonian has been obtained, we use a recently proposed order-N method[13,14] to compute the band structure energy $E_{\rm BS}$ (sum of occupied eigenvalues). In this approach, a modified band energy functional[13,14] is minimized[15] with respect to the electronic orbitals, by means of a conjugate gradients (CG) algorithm, to yield $E_{\rm BS}$. The orthonormality of the occupied states does not need to be imposed, but it is obtained as a result of the minimization of the energy functional. The elimination of the orthogonalization is the first step to achieve an order-N scaling. The second is the use of localized, Wannier-like wave functions (LWF) to describe the electronic states entering the minimization of the energy functional. Truncation of these localized functions beyond a given cutoff R_c from the center of the LWF provides a linear scaling algorithm. The errors involved in this truncation, which can be reduced arbitrarily by increasing the value of R_c, are analyzed in detail in Ref. 1.

After the band energy has been minimized and the LWF's obtained, the new charge density is computed, completing a so-called SCF cycle. From the density, a new Hamiltonian is produced, the procedure being repeated until self-consistency in the charge density or the Hamiltonian is achieved. At this point, the total energy can be computed as

$$E_{\rm tot}=E_{\rm BS}-\frac{e^2}{2}\int V_H(\mathbf{r})n(\mathbf{r})d\mathbf{r}+\frac{e^2}{2}\int V_H^0(\mathbf{r})n_0(\mathbf{r})d\mathbf{r}+\int[\epsilon_{\rm XC}(n)-V_{\rm XC}(n)]n(\mathbf{r})d\mathbf{r}+U_{ii\text{-}ee}, \tag{4}$$

where $V_H(\mathbf{r})$ is the Hartree potential of the self-consistent charge $n(\mathbf{r})$, and, following Sankey and Niklewski,[9] we have defined

$$U_{ii\text{-}ee}=\frac{e^2}{2}\sum_{ll'}{}'\frac{Z_lZ_{l'}}{|\mathbf{R}_l-\mathbf{R}_{l'}|}-\frac{e^2}{2}\int V_H^0(\mathbf{r})n_0(\mathbf{r}). \tag{5}$$

As in the case of the Hamiltonian, we have added and subtracted the electrostatic energy of the neutral atom charge $n_0(\mathbf{r})$ to obtain Eq. (4). The advantage, again, is that $U_{ii\text{-}ee}$ can be expressed as a sum of short-range contributions, which is easy to evaluate in $O(N)$ operations,[9] avoiding the problems related with the long-range character of the ionic core interactions. The integrals appearing in Eq. (4) are calculated in the real-space grid.

In molecular dynamics (MD) simulations and geometrical optimizations the atomic forces are needed. We compute them using a variation of the Hellman-Feynman theorem, which includes Pulay-like corrections to account for the fact that the basis set is not complete and moves with the atoms. The force on atom i is

$$\mathbf{F}_i=-\sum_{\mu\nu}\rho_{\mu\nu}\frac{\partial H^0_{\mu\nu}}{\partial\mathbf{R}_i}+\sum_{\mu\nu}E_{\mu\nu}\frac{\partial S_{\mu\nu}}{\partial\mathbf{R}_i}-\frac{\partial U_{ii\text{-}ee}}{\partial\mathbf{R}_i}+2\sum_{\mu}\rho^0_{\mu}\left\langle\frac{\partial\phi_\mu}{\partial\mathbf{R}_i}\middle|V_H^\delta\middle|\phi_\mu\right\rangle-2\sum_{\mu\nu}\rho_{\mu\nu}\left\langle\frac{\partial\phi_\mu}{\partial\mathbf{R}_i}\middle|V_H^\delta+V_{\rm XC}\middle|\phi_\nu\right\rangle, \tag{6}$$

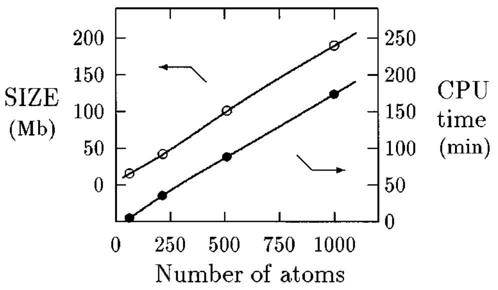

FIG. 2. CPU time per SCF cycle and job memory for a simulation of Si supercells with different sizes. Times measured in an IBM PowerPC with 17 Mflops (Linpack 100×100).

where $H^0=p^2/2m+V_{\rm NL}+V_{\rm NA}$, and $\rho_{\mu\nu}$ and $E_{\mu\nu}$ are the density and energy-density matrices, respectively.[9] The first three terms are calculated interpolating the table data,[9] whereas the last two terms are computed by numerical integration in the grid, as was done for the matrix elements of the Hartree and XC potentials in the Kohn-Sham Hamiltonian.

In order to show the linear scaling of the method, we have performed calculations on supercells of silicon in the diamond structure, with different numbers of atoms from 64 to 1000. Only the Γ point was used to sample the Brillouin zone, the cutoff for the charge density grid was 12 Ry, and the LWF's were truncated at 4.5 Å. Figure 2 shows the linear behavior of the CPU time and memory requirements with the number of atoms. The CPU time shown represents the average cost to perform a SCF step in a MD simulation, including the calculation of the charge density and Hamiltonian matrix elements, the minimization of the band structure energy, and the calculation of the atomic forces. The band structure energy minimization within each SCF cycle required an average of 20 CG iterations, while the number of SCF cycles depends largely on the simulation temperature, length of the time step, and mixing algorithm for self-consistency. So far, in comparable simulation conditions, no significant dependence of the number of CG iterations and SCF cycles on the size of the system has been observed. As we can see, in the present method both the CPU time and memory requirements are small enough to permit the calculation of a system of 1000 silicon atoms in a very modest work station.

As an example of a system with partially ionic character, and with atoms with compact orbitals, we have performed calculations on the β phase of C_3N_4, which was proposed as a potentially very hard material by Liu and Cohen.[17] The calculations were done in supercells of 42 and 224 atoms, with nearly identical results. A cutoff of 200 Ry for the charge density grid was used. We obtain an accuracy better than 1% in both the lattice constants and the several inequivalent bond lengths, and 10% in the bulk modulus, compared to other LDA calculations.[17] These results contrast with those of the non-self-consistent Harris functional, which yield errors of 5% and 16% for the distances and bulk modulus, respectively, showing that self-consistency is essential to obtain reliable results in this partially polar system.

We have applied our method to study the structure of large, single-shell, icosahedral fullerene clusters. These are important to understand the observed sphericity of multishell fullerenes. For the single-shell clusters, elasticity theory, as well as empirical potential calculations, predict markedly polyhedral shapes. Calculations performed by Itoh *et al.*,[16] using the Harris functional order-N method,[1] agree qualitatively with these results. However, similar non-self-consistent calculations[3] predict that even the large clusters are spherical. Here we have repeated the calculations with self-consistent LDA using the present method, thus improving on the non-self-consistent nature of the former calculations. Using a dynamical quenching algorithm, we have computed the equilibrium structure of three icosahedral fullerene clusters: C_{60}, C_{240}, and C_{540}. A supercell geometry was assumed, with a cubic cell with sides of 12 Å for C_{60}, 22 Å for C_{240}, and 34 Å for C_{540}. The calculations were done using a cutoff of 100 Ry for the representation of the charge density in reciprocal space, and a different localization radius for σ- and π-type Wannier functions (2.5 and 4.0 Å, respectively).[16] Increasing the localization radius to 4.1 Å (both for σ and π), and/or increasing the grid cutoff to 150 Ry in the simulations changes the relaxed interatomic distances less than 0.4%. The results are summarized in Table I. We see that our results are very similar to those obtained by Itoh *et al.*, and confirm that, except for C_{60}, the single-shell clusters tend to be polyhedral, instead of spherical, and that this polyhedral character is more pronounced as the cluster size increases.

TABLE I. Average radius ($\overline{r}$), standard (σ_s), and maximum deviation [$\sigma_m=(r_{\max}-r_{\min})/2$] of radii, and nonplanarity angle[3] ϕ around pentagons (going from 0° for a planar pentagonal site to 12° for a truncated icosahedron) for the fullerene clusters. We compare the results of the present work with those of Itoh *et al.* (Ref. 16) obtained with the Harris functional.

	This work				Itoh *et al.*			
	$\overline{r}$ (Å)	$\sigma_s/\overline{r}$	$\sigma_m/\overline{r}$	ϕ	$\overline{r}$ (Å)	$\sigma_s/\overline{r}$	$\sigma_m/\overline{r}$	ϕ
C_{60}	3.59	0.000	0.000	12.0°	3.55	0.000	0.000	12.0°
C_{240}	7.18	0.023	0.027	8.5°	7.06	0.021	0.028	7.9°
C_{540}	10.69	0.038	0.054	9.6°	10.53	0.033	0.053	9.2°

In conclusion, we have presented an efficient method for self-consistent LDA calculations with linear scaling. We have analyzed the performance versus system size and grid cutoff, and shown that simulations of systems with hundreds of atoms are possible with small work stations. This should open the possibility of very large scale *ab initio* simulations in the near future.

We acknowledge R. M. Martin and Paul von Allmen for many useful discussions, and D. A. Drabold and O. F. Sankey for allowing us the use of many of their codes. P.O. is indebted to R. M. Martin and J. B. Adams for continuous support and encouragement. This work was partially supported by DOE Grant No. DEFG 02-91ER45439 and DGICYT (Spain) Grant No. PB92-0169.

---

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

[2] J. Harris, Phys. Rev. B **13**, 1770 (1985).

[3] D. York, J. P. Lu, and W. Yang, Phys. Rev. B **49**, 8526 (1994); J. P. Lu and W. Yang, *ibid.* **49**, 11 421 (1994).

[4] W. Hierse and E. B. Stechel, Phys. Rev. B **50**, 17 811 (1994).

[5] E. Hernandez and M. Gillan, Phys. Rev. B **51**, 10 157 (1995); (unpublished).

[6] R. Poirier, R. Kari, and I. G. Csizmadia, *Handbook for Gaussian Basis Sets* (Elsevier Science, New York, 1985), and references therein.

[7] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965). The exchange-correlation potential is taken from D. M. Ceperley and G. J. Alder, Phys. Rev. Lett. **45**, 566 (1980), as parametrized by J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[8] G. B. Bachelet, D. R. Hamman, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[9] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).

[10] D. Sanchez-Portal, E. Artacho, and J. M. Soler (unpublished); Solid State Commun. **95**, 685 (1995).

[11] W. H. Press *et al.*, *Numerical Recipes* (Cambridge University, New York, 1989).

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

[13] P. Ordejón, D. A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **48**, 14 646 (1993); F. Mauri, G. Galli, and R. Car, *ibid.* **47**, 9973 (1993).

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

[15] The variables of the minimization are the coefficients $c_{\mu i}$ of the expansion of the localized wave functions in terms of the atomic orbitals: $\psi_i=\Sigma_\mu c_{\mu i}\phi_\mu$.

[16] S. Itoh, P. Ordejón, D. A. Drabold, and R. M. Martin (unpublished).

[17] A. Y. Liu and M. L. Cohen, Phys. Rev. B **41**, 10 727 (1989); A. Reyes-Serrato, D. H. Galván, and I. L. Garzón, *ibid.* **52**, 6293 (1995), and references therein.