Adaptive finite-element method for electronic-structure calculations

Eiji Tsuchida and Masaru Tsukada
Department of Physics, Faculty of Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan
(Received 7 June 1996))

We present a method which combines the finite-element method with the adaptive curvilinear coordinates, and the method is applied to the electronic-structure calculation of a model potential system. Comparison with other real-space methods, such as the finite-difference method is also made, and the efficiency of our method is examined. In addition, it has some desirable properties, and will be useful as a part of the $O(N)$ method for self-consistent calculations. [S0163-1829(96)06236-4]

I. INTRODUCTION

Recently, a number of real-space methods have been proposed for the *ab initio* calculation of materials.[1–11] This is mainly because they have the following desirable properties compared with the conventional plane-wave method:[12] (i) Orbital formulation of linear scaling algorithms[13–16] requires that the wave functions are strictly localized in given regions of space. This is easily achieved in real-space methods. (ii) In many of the real-space methods, we can put more points in regions where the oscillation of the wave function is rapid. This *local* refinement reduces the computational effort considerably. (iii) They are suitable for parallel computers, because no Fourier transforms are necessary.

However, work in this direction is still in progress, and there seems to be much room for improvement. In this paper, we propose another real-space approach which combines the finite-element method (FEM) with the adaptive curvilinear coordinates (ACC's), and show that it is an efficient scheme with a number of desirable properties. In our previous works,[6] we developed the FEM for self-consistent electronic-structure calculations on arbitrary grids. It is advantageous, especially for all-electron calculations of materials, because the grid can be made to vary logarithmically near the nuclei. However, as a compensation for this high freedom of the grids, the smoothness of the wave function is not guaranteed.[17] This loss of smoothness is undesirable for describing the smooth pseudo-wave-functions of the valence electrons. To improve this deficiency, we restrict the freedom of the grids, and instead obtain the smoothness of the wave functions while retaining ease in describing localized wave functions.

II. FORMALISM

A. Basis functions

First we will discuss the basis functions. In Ref. 6, we used standard finite-element basis functions of the first and second orders. They allowed us to work on rather arbitrary grids, but we did not stick to this property, and limited the grids to those that are continuously transformed from uniform grids. Thus, as given below, we can use smooth basis functions whose derivatives are continuous.

For simplicity, we consider the one-dimensional case. A pair of functions,

$$S_0(x)=\begin{cases}1-3x^2+2x^3, & 0\leqslant x\leqslant 1\\ S_0(-x), & -1\leqslant x<0\\ 0, & \text{otherwise,}\end{cases} \tag{1}$$

$$S_1(x)=\begin{cases}x-2x^2+x^3, & 0\leqslant x\leqslant 1\\ -S_1(-x), & -1\leqslant x<0\\ 0, & \text{otherwise,}\end{cases} \tag{2}$$

whose shapes are shown in Fig. 1, have the following properties:

$$S_0(0)=1,\quad S_0'(0)=S_0(\pm 1)=S_0'(\pm 1)=0, \tag{3}$$

$$S_1'(0)=1,\quad S_1(0)=S_1(\pm 1)=S_1'(\pm 1)=0. \tag{4}$$

This means that S_0 corresponds to the value of the function at $x=0$, and S_1 corresponds to the derivative at $x=0$. Note also that they are strictly localized in region $[-1,1]$, and zero elsewhere. This is an important property for linear scaling algorithms.[13–16] On a three-dimensional uniform grid, the basis functions are given by the products of these functions:

$$\left\{S_i\left(\frac{x-n_x}{h}\right)S_j\left(\frac{y-n_y}{h}\right)S_k\left(\frac{z-n_z}{h}\right)\right\},\quad i,j,k=0,1 \tag{5}$$

where n_x, n_y, and n_z are the coordinates of the grid points, and h is the spacing of the grid. Therefore, eight basis functions are assigned to each grid point.[18] These functions and their similitudes are often called spline functions rather than the FEM, and was already introduced in Refs. 1 and 8. In particular, higher-order extensions and orthogonal basis functions are discussed in Ref. 1, but we do not use them in this paper.[19] It is possible, and not a bad idea, to work on uniform grids with these basis functions. But it will not be an efficient scheme, like the plane-wave method, when the electrons are localized. Thus we use these basis functions on ACC's, which are explained in Sec. II B.

B. Adaptive curvilinear coordinates

The concept of ACC's for electronic-structure calculations was brought about by Gygi,[20] in combination with plane-wave basis sets. Since then, it has been used successfully several times,[21–23] and recently combined with the

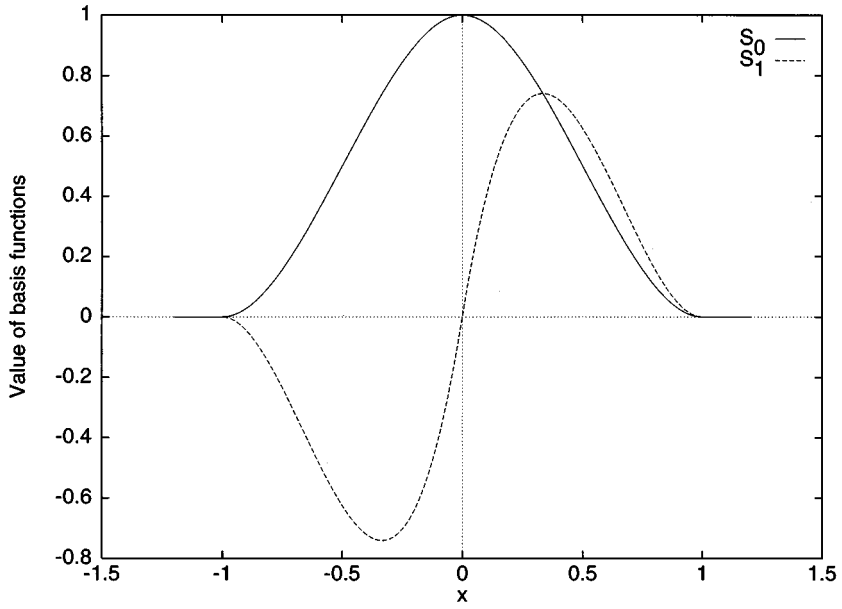

FIG. 1. The basis functions S_0 and S_1. Note that S_0 corresponds to the value of the function at $x=0$, and S_1 corresponds to the derivative at $x=0$. For numerical reasons, S_1 is multiplied by 5.0.

finite-difference (FD) approach.[7] The basic idea of ACC's is as follows. We first define a coordinate transformation $\boldsymbol{\xi}\leftrightarrow\mathbf{x}$, where $\mathbf{x}$ denotes the Euclidean coordinates which are usually used, and $\boldsymbol{\xi}$ denotes the curvilinear coordinates. Suppose we choose a transformation such that the projection of a uniform grid in ξ space onto x space becomes dense where the potential varies rapidly. Then, if we work in ξ space, the resolution can be enhanced *locally*, and the calculation will be performed efficiently. There is no unique way to define the mesh, so we follow the method of Ref. 7 and define the transformation as below:

$$\boldsymbol{\xi}=\mathbf{x}+\sum_{\alpha}(x-\mathbf{R}_{\alpha})f_{\alpha}(|\mathbf{x}-\mathbf{R}_{\alpha}|), \tag{6}$$

$$f_{\alpha}(r)=A_{\alpha}\frac{a_{\alpha}}{r}\tanh\left(\frac{r}{a_{\alpha}}\right)\exp\left[-\left(\frac{r}{b_{\alpha}}\right)^{2}\right], \tag{7}$$

where $\mathbf{R}_{\alpha}$ denotes the atomic positions, A_{α} is the degree of distortion, a_{α} is the range of magnification, and b_{α} is the range of distortion. In practice, we need to calculate the Jacobian matrix J, its derivatives, and the metric tensor g^{ij} [24]

$$(J)_{ij}=\frac{\partial\xi_i}{\partial x_j},\quad (J)_{ij,k}=\frac{\partial}{\partial\xi_k}\left(\frac{\partial\xi_i}{\partial x_j}\right), \tag{8}$$

$$g^{ij}=(J)_{ik}(J)_{jk}. \tag{9}$$

These quantities can be calculated analytically from (6) and (7). For convenience, the wave function $\psi(x)$ is written as the product[7]

$$\psi(x)=|J|^{1/2}\phi(\xi(x)), \tag{10}$$

and $\phi(\xi)$ is expanded by the basis functions explained in Sec. II A in ξ space. In this representation, the kinetic and potential energies are expressed as follows:

$$E_{\mathrm{kin}}=\int\frac{\partial\psi}{\partial x_i}\frac{\partial\psi}{\partial x_i}\,d\mathbf{x} \tag{11}$$

$$=\int\left[A^{ij}(\boldsymbol{\xi})\frac{\partial\phi}{\partial\xi_i}\frac{\partial\phi}{\partial\xi_j}+B^{j}(\boldsymbol{\xi})\frac{\partial\phi}{\partial\xi_j}\phi+C(\boldsymbol{\xi})\phi^{2}\right]d\boldsymbol{\xi}, \tag{12}$$

where

$$A^{ij}\equiv g^{ij},\quad B^{j}\equiv(J)_{jm}(J)_{im,i},\quad C\equiv\tfrac{1}{4}(J)_{pr,p}(J)_{qr,q} \tag{13}$$

and

$$E_{\mathrm{pot}}=\int V(\mathbf{x})\psi^{2}(\mathbf{x})d\mathbf{x} \tag{14}$$

$$=\int V(\boldsymbol{\xi})\phi^{2}(\boldsymbol{\xi})\mathbf{d}\xi. \tag{15}$$

The derivatives are calculated analytically, and the integrals are calculated numerically if necessary.[25] In the present paper, we do not consider self-consistent cases, so the total energy is given by $E_{\mathrm{tot}}=E_{\mathrm{kin}}+E_{\mathrm{pot}}$. To obtain the ground-state energy and electronic density, we minimize E_{tot} with respect to the expansion coefficients of the wave functions. Unconstrained minimization[14,15] was employed as in Ref. 6.

III. APPLICATION

To compare the performance of our method and other real-space methods, we calculated the ground state of the system which was once used by Gygi.[20] The potential has the form

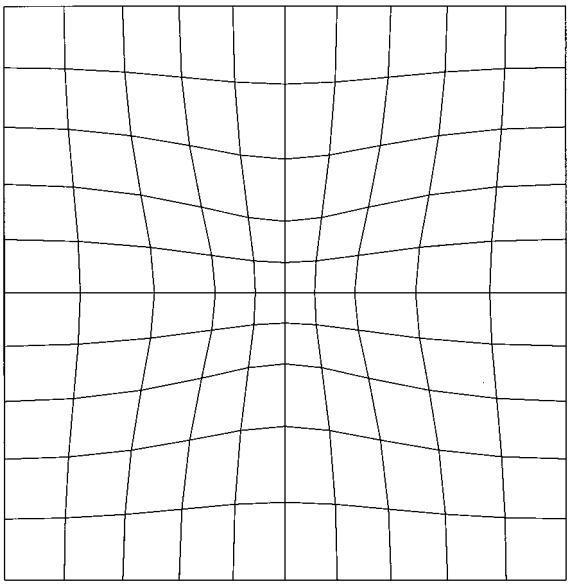

FIG. 2. The projection of a uniform grid in ξ space onto x space. It is clearly seen that the grid spacing is small near the center, where the potential is rapidly varying.

$$V(r)=V_0\exp\left[-\left(\frac{r}{r_c}\right)^2\right], \tag{16}$$

with $V_0=-32$ a.u., and $r_c=0.5$ a.u., and is periodically arrayed at intervals of 5 a.u. in three dimensions. This system can be regarded as a model for strongly localized systems. The parameters of ACC's are $A=1.0$, $a=0.5$ a.u., and $b=2.0$ a.u., respectively. The projection of a uniform grid in ξ space onto x space is given in Fig. 2. The calculated results are shown in Fig. 3. Compared with the other real-space methods, the remarkable efficiency of our method can be seen.[26] Note also that the energy curves of other methods on uniform grids oscillate as the number of variables changes. This means that the value of the energy largely depends on whether the peak of the potential is on the grid or between the grid.[27] Thus we must be careful about this point when we perform molecular-dynamics simulations or optimization of the ionic positions, and compare the energy of different configurations. In comparison, the energy curve of our method converges rapidly and monotonously as the number of variables increases.

IV. DISCUSSION

The major properties of our method are as follows: (i) We can make the grid so that it is dense near the center of localization, so strongly localized systems, such as first-row elements, can be treated without any serious difficulties. This leads to a significant reduction in computational costs. (ii) No Fourier transform is necessary, since all calculations are performed in real space. This is important, especially for massively parallel computers. (iii) Orbitals strictly localized in given regions of space are naturally expressed in this approach, so our method is suited to linear scaling algorithms[13–16] for *ab initio* calculations. (iv) The variational principle is valid in our approach, since the derivatives and inner products of the basis functions are calculated exactly. This is in contrast with the FD approach, in which the energy often converges from below. (v) The matrices are generally less sparse than in the FD approach, though the sparsity does not depend on whether a uniform grid or an adaptive grid is used.[28] One solution to this problem is to use the orthogonal basis functions proposed by White, Wilkins, and Teter,[1] but we have not checked their accuracies yet. (vi) Setting up ACC's may look like complicated and time-consuming work, but the Jacobian, its derivatives, and the metric tensor have high symmetry, and calculating them is in principle a linear scaling procedure, so it will be done easily, especially on parallel computers.

V. CONCLUSION

We have presented a method for *ab initio* calculation of electronic structures based on the combination of the FEM and ACC's. Our method has been shown to give good results for a model system, compared with other real-space ap-

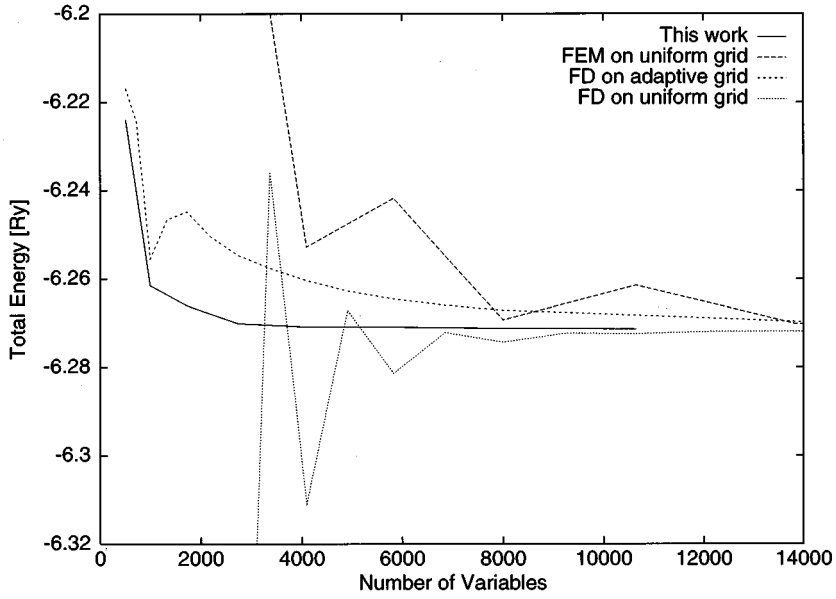

FIG. 3. The calculated results of various real space approaches are shown. The FD approach on adaptive grid (Ref. 7) was performed with order 3, and the same parameters of ACC's were used as in the present work. The FD approach on uniform grid (Ref. 5) was performed with order 6.

proaches. A straightforward extension of the present scheme should include many electron self-consistent calculations with nonlocal pseudopotentials.[29,30] However, self-consistent calculations have already been performed with the FEM (Ref. 6) or ACC's,[7] and there seems to be no serious obstacle. The goal of the real-space approaches will be to perform *ab initio* molecular-dynamics calculations with time proportional to the number of atoms. Our scheme can be a promising candidate for this purpose.

ACKNOWLEDGMENTS

One of the authors (E.T.) is associated with the Japan Society for the Promotion of Science, and would like to thank Katsuyoshi Kobayashi for fruitful discussions. This work was partially supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan, and the facilities of the Supercomputer Center, Institute for Solid State Physics, University of Tokyo.

[1] S. R. White, J. W. Wilkins, and M. P. Teter, Phys. Rev. B **39**, 5819 (1989).

[2] J. Bernholc, J.-Y. Yi, and D. J. Sullivan, Faraday Soc. Discuss. **92**, 217 (1991); E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **52**, R5471 (1995).

[3] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1992).

[4] K. Cho, T. Arias, J. D. Joannopoulos, and P. K. Lam, Phys. Rev. Lett. **71**, 1808 (1993).

[5] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994); J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad, Phys. Rev. B **50**, 11 355 (1994).

[6] E. Tsuchida and M. Tsukada, Solid State Commun. **94**, 5 (1995); Phys. Rev. B **52**, 5573 (1995).

[7] F. Gygi and G. Galli, Phys. Rev. B **52**, R2229 (1995).

[8] J. F. Annett (unpublished).

[9] G. Zumbach, N. A. Modine, and E. Kaxiras, Solid State Commun. (to be published).

[10] S. Wei and M. Y. Chou (unpublished).

[11] V. Rao and R. Martin (unpublished).

[12] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992), and references therein.

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

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

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

[16] W. Kohn, Chem. Phys. Lett. **208**, 167 (1993).

[17] More exactly, the wave functions are continuous, but the derivatives of them are not continuous (C^0). Higher-order interpolation functions can improve the results, but even in this case, the continuity of the derivatives are not guaranteed.

[18] It may seem that higher-order terms, like $S_1(x)S_1(y)S_1(z)$, are unnecessary. However, our experience shows that omitting these terms sometimes leads to poor convergence, so we include all the terms.

[19] We tried the fifth-order basis functions, to find that they were somewhat better than the third-order basis functions (S_0, S_1). However, this amounts to assigning 27 basis functions per site, and the sparsity of the matrices will be considerably reduced.

[20] F. Gygi, Europhys. Lett. **19**, 617 (1992).

[21] F. Gygi, Phys. Rev. B **48**, 11 692 (1993); **51**, 11 190 (1995).

[22] A. Devenyi, K. Cho, T. A. Arias, and J. D. Joannopoulos, Phys. Rev. B **49**, 13 373 (1994).

[23] D. R. Hamann, Phys. Rev. B **51**, 7337 (1995); **51**, 9508 (1995).

[24] Summation over repeated indices is implied throughout the paper.

[25] For numerical integration, we used the Gauss-Legendre integration scheme of order 4 or 5.

[26] The original calculation of Gygi is not included in the figure, since his calculation was performed under the condition the grid was also allowed to vary during the minimization. However, it is possible to fulfill the same condition in our approach.

[27] This fact was more directly confirmed by moving the peak of the potential from one grid point to another grid point, and calculating the energy along the path, with the grid fixed.

[28] The number of nonzero elements in one line of a matrix is given by (number of basis functions per grid point) times 3^3. Therefore, for the basis functions we considered, the degree of sparsity is 216.

[29] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

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