

Real-space electronic-structure calculations: Combination of the finite-difference and conjugate-gradient methods

Ari P. Seitsonen

*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin-Dahlem, Germany
and Laboratory of Physics, Helsinki University of Technology, 02150 Espoo, Finland*

M. J. Puska, and R. M. Nieminen

Laboratory of Physics, Helsinki University of Technology, 02150 Espoo, Finland

(Received 10 January 1995)

We present a scheme for a rapid solution of a general three-dimensional Schrödinger equation. The Hamiltonian operator is discretized on a point grid using the finite difference method. The eigenstates, i.e., the values of the wave functions in the grid points, are searched for as a constrained (due to the orthogonality requirement) optimization problem for the eigenenergies. This search is performed by the conjugate-gradient method. We demonstrate the scheme by solving for the self-consistent electronic structure of the diatomic molecule P_2 starting from a given effective electron potential. Moreover, we show the efficiency of the scheme by calculating positron states in low-symmetry solids.

I. INTRODUCTION

The Kohn-Sham method of density-functional theory (DFT) (Ref. 1) casts the problem of many interacting electrons into the problem of solving one-particle Schrödinger equations. This is a tremendous simplification which makes large-scale first-principles electronic-structure calculations possible. However, the practical implementation of DFT still poses the problem of a fast and accurate solution of the Kohn-Sham equations for a general three-dimensional geometry. One way to reduce the calculational effort is the use of pseudopotentials and pseudo (valence) wave functions to decrease the number of electron states and to avoid difficulties due to the rapid oscillations of the valence wave functions in the ion core regions. The pseudowave functions can be expanded, e.g., in a plane-wave basis, that is, by Fourier analysis. To do this, one imposes periodic boundary conditions and the supercell approximation also for systems such as isolated defects in solids or finite atomic clusters. However, a mathematical manipulation can be tried in order to avoid the effects of the supercell approach.² By using plane waves, the numerical accuracy of the calculation can be systematically improved by increasing their number, i.e., the cutoff energy. Another feature of the plane-wave expansion is that the resolution is uniform in the supercell. This is sometimes wasteful as the requirement for high cutoff is often most stringent near the atomic cores. A severe problem is that the number of plane waves needed can become so large that the calculations become extremely demanding on computer time and memory space.

The finite-difference method has been used to solve the Schrödinger equation in a few contexts. Kimball and Shortley³ introduced a method where the wave function is obtained through an iterative relaxation scheme.

This method has been later applied^{4,5} to calculate the quantum-mechanical state of the hydrogen impurity in metals and to calculate positron states in solids.⁶ The energy eigenvalue decreases monotonically towards its converged value, but the drawback of the method is that the convergence becomes extremely slow for large point grids. Joly⁷ has applied the difference method to calculate the low-energy-electron-diffraction and the low-energy-positron-diffraction intensities. Baroni and Giannozzi⁸ introduced a Green's function method which employs the finite differences and a real-space grid, and they even obtained a linear scaling with the volume of the system.

Recently, Chelikowsky *et al.*⁹ showed that it is possible to use the finite-difference method within the pseudopotential scheme and solve for the electron wave functions in real space, at least for finite atomic clusters. The eigenstates, i.e., the eigenenergies and the corresponding values of the wave function in the mesh points, are determined by an iterative diagonalization of the ensuing Hamiltonian matrix. It is easy to improve systematically the numerical accuracy of the method by increasing the point density. Moreover, real-space methods are suited well for parallel computing. On the contrary, the plane-wave method utilizes fast Fourier transformation routines, which are not as straightforward to fully parallelize.

In this paper we report on an important improvement of the real-space technique: we combine the finite-difference method with an iterative conjugate-gradient (CG) solution for the lowest-lying eigenstates of the Kohn-Sham Hamiltonian. Moreover, we use the technique to solve the positron states in low-symmetry solids. For this application the method suits particularly well, because in the simplest case the potential is constructed from the electronic structure calculated without the

positron and does not depend on the positron density. Previously, the CG technique has been used in electronic-structure calculations^{10,11} based on the direct optimization of the Kohn-Sham total energy functional using the plane-wave basis set. Now the energy eigenvalues are optimized in the space spanned by the values of the wave functions in the grid points.

II. FINITE-DIFFERENCE AND CONJUGATE-GRADIENT TECHNIQUES

In the finite-difference method for the Schrödinger equation the values of all functions (potentials, wave functions, charge densities) are given on the mesh points. The Laplacian operator of the kinetic energy is replaced by a difference approximation for the second derivative. In our applications we have used a (uniform) simple cubic mesh, and we found that the second derivative of the wave function in one direction can be approximated accurately by a numerical operator employing five adjacent mesh points.

The technique requires a starting potential for the Schrödinger equation. For the P_2 dimer we have first solved in the local-density approximation (LDA) for electron exchange and correlation the electronic structure by using the norm-conserving nonlocal pseudopotentials¹² and the plane-wave basis. From the self-consistent results the sum of the Hartree and exchange-correlation potentials of the pseudovalence density and the local part of the pseudopotential are determined at the real-space grid points. For the positrons we construct the potential by superimposing free atoms.⁶ The positron potential consists of a Coulomb part due to nuclei and electrons and of a correlation part, which is calculated in LDA (Ref. 13) from the superimposed electron density.

In the calculation of the wave functions we have applied the CG method, described by Payne *et al.*¹¹ We start from a trial wave function ψ_i^0 . Here the superscript stands for the iteration number and the subscript denotes the different electron states for the dimer. For the positron we calculate only the ground state wave function. The trial wave functions may be random numbers at the grid points, but especially in the case of positron states we have found more efficient a function which is constant in the (interstitial) regions with a low potential and vanishes (near the nuclei) where the potential is high. To start the iterations, the steepest descent vector

$$\zeta_i^m = -(H - \lambda_i^m)\psi_i^m \quad (1)$$

is calculated. Above λ_i^m is the present approximation for the energy eigenvalue,

$$\lambda_i^m = \langle \psi_i^m | H | \psi_i^m \rangle. \quad (2)$$

The Hamiltonian H contains the kinetic energy operator and the local potential which is kept constant during the wave function iterations. In our calculations for the electron states the Hamiltonian contains also an operator due to the nonlocality of the pseudopotential. The

fact that the local potential does not depend on the wave function means that the function to be optimized is the energy eigenvalue. In the case of several eigenstates of a dimer we first calculate the one with the lowest energy. We then obtain the next energy state and orthogonalize the steepest descent direction against the lowest energy state. Thereafter we continue to higher energy states and orthogonalize against the lower states. In the case of a large system, a very large number of eigenstates has to be solved. For this process a more efficient method should be found. For example, Wang and Zunger¹⁴ showed that when the plane-wave expansion is used with a non-self-consistent potential the Lanczos algorithm becomes faster than the preconditioned CG method when the system size increases, but it suffers from the well-known ‘‘Lanczos phenomena.’’

After the calculation of the steepest descent direction the preconditioning step follows. In the case of the positron states we found that without preconditioning the amplitude of the steepest descent vector during the first iterations is much larger near the nuclei than in the interstitial or open defect regions, where the positron wave function finally will reside. Therefore for the positron calculations we use a preconditioning which shifts the weight of the steepest descent vector from the ion core regions to the open regions of the lattice. We do this by first calculating for every grid point the relative distance between the energy eigenvalue λ_i^m and the potential $V(\mathbf{r})$ at that point

$$x = A \frac{|\lambda_i^m - V(\mathbf{r})|}{T_i^m}, \quad (3)$$

where $T_i^m = \langle \psi_i^m | (-\hbar^2 \nabla^2 / 2m) | \psi_i^m \rangle$ is the kinetic energy. The preconditioning is then performed by multiplying the steepest descent vector at the point \mathbf{r} by the following expression suggested by Teter *et al.*:¹⁰

$$K(\mathbf{r}) = \frac{27 + 18x + 12x^2 + 8x^3}{27 + 18x + 12x^2 + 8x^3 + 16x^4}. \quad (4)$$

Above A is an adjustable factor which controls the amplitude ratios of the steepest descent vector between the ion core and open lattice regions. We have found that around the maximum speedup (the number of iterations needed reduces by preconditioning typically by a factor of 5) the iteration process does not depend strongly on A and we have used the value of 0.1. When calculating the pseudovalence electron states the amplitude of the steepest descent vector is rather evenly distributed during the first iterations and the use of preconditioning function given above is not effective. The effects of the preconditioning depend also on the type of the trial wave function chosen.

The preconditioned steepest descent vector is orthogonalized against the present state to be calculated and against the lower energy states already determined. Thereafter the conjugate preconditioned steepest descent vector is found by the help of that vector of the previous iteration and finally the conjugate vector is orthogonalized against the present wave function and normalized.

The new wave function is constructed as a linear combination from the present wave function and the conjugate vector. The mixing of the functions in the combination is determined following the description of Payne *et al.*¹¹ by requiring an approximative minimization of the energy eigenvalue in this subspace.

III. TEST CALCULATIONS

We demonstrate the real-space conjugate-gradient (RCG) method by calculating the electronic structure of the P_2 dimer. It is a good test case because accurate plane-wave results, with which the comparison will be done, can be obtained with a reasonable energy cutoff. Moreover, the comparisons can be done for four molecular orbitals of different character occupied by the valence electrons. The phosphorous nuclei are separated in our calculations by the experimental bond length of 2.89 Å. The calculations are performed by using a cubic cell with the sides of 8.20 Å.

We have first performed self-consistent electronic-structure calculations with the plane-wave method using several energy cutoffs up to 30 Ry. The local potentials obtained are then fed in the RCG code, and the energy eigenvalues and wave functions are determined. Both in the plane-wave and in the RCG calculations periodic boundary conditions were used for the wave functions. Because for a given energy cutoff the potential is the same in both methods the different energy eigenvalues and wave functions arising from the two methods can be directly compared. Finally, we have iterated the real-space technique in order to find the fully self-consistent potential and electronic structure. This we have done by starting from the potential obtained by the plane-wave method using a 15 Ry energy cutoff; less than 10 self-consistency iterations are then needed for convergence. For solving the Hartree potential we use the direct integration which in practice is a sum of the potentials due to uniformly charged Wigner-Seitz spheres at the grid points. In the large-scale calculations one should use iterative methods or rely on the fast Fourier transformation as in Ref. 2.

The energy eigenvalues of the molecular orbitals calculated using the plane-wave expansion are shown in Fig. 1 as a function of the cutoff energy. The finite-difference results, which are obtained after a given plane-wave calculation using the same local potential, are also shown. The number of points in the finite-difference grid increases as the cutoff energy increases: 32^3 points are used with 6, 8, and 10 Ry, 40^3 with 12, 14, and 15 Ry, 42^3 with 18 Ry, and finally 56^3 points with the cutoff energy of 30 Ry. Because the eigenvalues change smoothly as the point density changes stepwise we conclude that the grid point densities used are large enough in order not to influence remarkably the RCG results. The errors due to the use of a limited number of points in the finite difference form for the kinetic energy operators are small. The convergence of the kinetic energy as a function of the order of the difference form is fast: In the worst case of the lowest energy state the eigenvalue has in practice

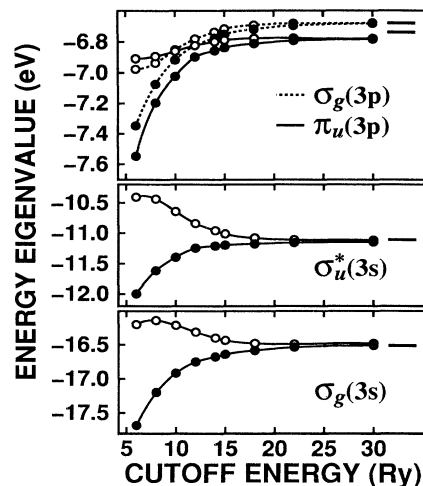


FIG. 1. Energy eigenvalues of the P_2 molecular orbitals as a function of the energy cutoff used in the plane-wave calculation. The open circles are the results of the plane-wave calculation. The black circles refer to the finite-difference calculations in which the potential of the corresponding plane-wave calculation is used. The thick lines on the right-hand side are the results of self-consistent finite-difference calculation.

converged with the 7-point form and the change is only 25 meV when the 5-point form is substituted by the 11-point form. Therefore the fact that the finite-difference eigenenergies are consistently below the plane-wave ones means that the description of the wave functions is better in the finite-difference method than in the plane-wave method, in which the spurious oscillations of the wave functions far from the nuclei raise the kinetic energy (see Fig. 2 below). Indeed, we have calculated the energy eigenvalues corresponding to the plane-wave wave function in our RCG code and they differ from those of the plane-wave calculation by the inaccuracy of the form for the kinetic energy.

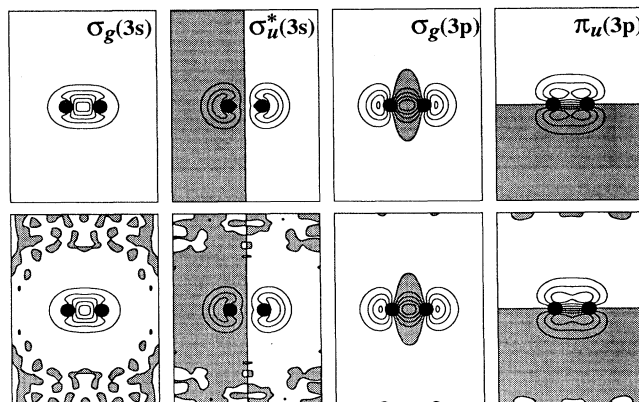


FIG. 2. Molecular orbitals of the P_2 dimer. The upper and lower panels show results from the finite-difference and plane-wave calculations, respectively. The contour spacing is one-sixth of the maximum value. The regions of negative values are shadowed and the positions of nuclei are given as black dots.

The finite-difference eigenvalues rise with increasing cutoff energy reflecting changes in the potential. (The potential depends on the cutoff energy through the accuracy of the electron density.) When calculating the two lowest eigenenergies by the plane-wave method this increase in the potential as a function of the cutoff energy is compensated by the improvement of the wave function leading to a decreasing trend. A deficiency of the the plane-wave expansion with small cutoff energies is that the order of the two highest occupied states $\pi_u(3p)$ and $\sigma_g(3p)$ is reversed when the cutoff energy decreases. The plane-wave and the finite-difference eigenenergies converge towards each other when the cutoff energy increases. The energy eigenvalues obtained by the self-consistent RCG calculations are shown in the right-hand side of Fig. 1. The starting potential on a grid of 40^3 points has been obtained from the plane-wave calculation with the cutoff energy of 15 Ry. The self-consistent levels are shifted so that the lowest level coincides with that obtained from the potential corresponding to the 30 Ry cutoff energy in the plane-wave calculation. During the self-consistency iterations the potential improves because of the better description of the wave function in the RCG method compared to the plane-wave calculations. Then the upper levels coincide quite well with the best previous calculations with the 30 Ry cutoff energy. This proves the accuracy of the self-consistent scheme in the real space.

The eigenfunctions of the occupied P_2 molecular orbitals are shown in Fig. 2. The plane-wave and the RCG wave functions shown correspond to the same potential obtained by the plane-wave calculation with 15 Ry cutoff energy. The wave functions are found to be in remarkably good agreement near the nuclei. The difficulty of plane-wave expansions to describe the nearly vanishing wave functions far from the nuclei is also clearly demonstrated.

Our test runs for positron states include systems from simple metals through semiconductors and insulators to high- T_c superconducting materials. We have found that this method means huge savings in the computer time in comparison with the Kimball-Shortley iteration scheme used earlier. This is true especially for positron states localized at defects. Below we demonstrate the efficiency of the new method in the case of a positron trapped by a Cd vacancy in the II-VI compound semiconductor CdTe.

Figure 3 shows the positron wave function at the Cd vacancy. The trapping is a result of the reduction of the positron-nucleus repulsion at the vacancy relative to that in the perfect crystal lattice. The trapping is reflected in the increase of the positron lifetime against the annihilation with an electron, because the electron density at the defect is also reduced. For example, according to our calculations based on the LDA for the annihilation rate,^{13,15} the positron lifetime increases from 276 ps to 288 ps.

According to Fig. 3 the positron wave function at the Cd vacancy is relatively delocalized. This results in a quite small lifetime increase due to the trapping and it also makes the calculations demanding because the supercell needed for reliable calculations becomes large. We have used a supercell of 215 atoms and a finite-difference

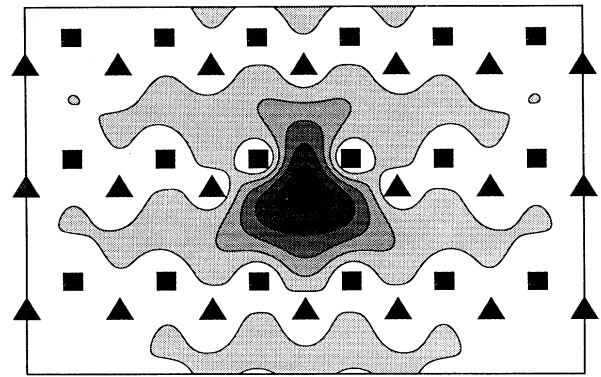


FIG. 3. Positron wave function at the Cd vacancy in CdTe. The figure shows a region of the (110) plane limited by the borders of the calculation cell. The contour spacing is one-sixth of the maximum value. The darker shading indicates larger amplitude of the wave function. The triangles and squares denote the positions of the Cd and Te nuclei, respectively.

grid consisting of 72^3 points. This means that the distance between neighboring points is 0.27 Å. For comparison, the use of a supercell of 63 atoms decreases the positron lifetime estimate by 3 ps, whereas the effect of doubling the grid spacing (reducing the number of grid points by the factor of 8) is less than 1 ps.

Figure 4 shows the convergence of energy eigenvalue for the positron at the Cd vacancy as a function of the cpu time used.¹⁶ The convergence is the difference of the present eigenvalue and the eigenvalue iterated close to the numerical accuracy of the computer. It can be seen that in order to attain a convergence of 10^{-5} Ha the Kimball-Shortley iteration requires about 30 times the time taken by our optimized preconditioned conjugate-gradient scheme. The effect of the preconditioning is a

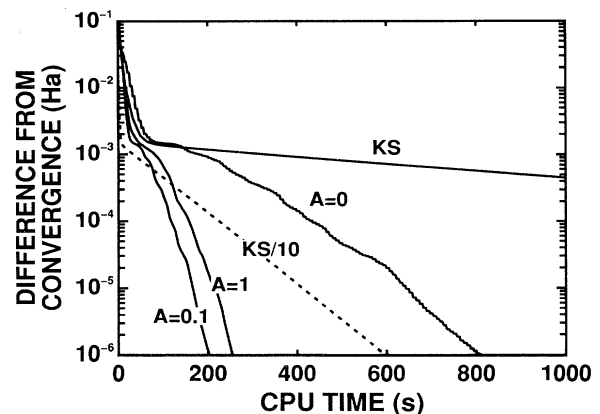


FIG. 4. Convergence of the energy eigenvalue for a positron at the Cd vacancy in CdTe. The results of the Kimball-Shortley (KS, the dashed line shows the values with the time scaled by the factor of 10) and the conjugate-gradient method without preconditioning ($A = 0$), as well as with preconditioning with $A = 0.1$ or $A = 1$ [Eq. (3)], are shown.

factor of about 4. It can be also seen that the speedup does not depend very strongly on the parameter A used in Eq. (3).

The real-space techniques are suitable for parallel computing. One obvious approach is to decompose the space grid to domains treated in different processors, with matching (and communication) at boundaries. We have tested this approach and found that in typical applications the cpu time per processor is for two processors about 55% of the cpu time needed in a conventional serial calculation with one processor. The corresponding value for four processors is about 30%. Another enhancement of the technique is to increase the grid resolution using, e.g., multigrid techniques. For moving nuclei, an adaptive grid scheme can be implemented.¹⁷

IV. CONCLUSIONS

We have shown that the finite-difference method combined with the conjugate-gradient optimization techniques can be efficiently used to solve the wave functions and energy eigenvalues. To become really competitive in electronic-structure calculations real-space methods still call for optimization and parallelization, and algorithmic development, e.g., to obtain accurate forces on the atoms. However, the progress made so far is certainly promising enough to warrant efforts in this area. The technique presented in this paper is already very powerful in low-symmetry applications where few (or none) self-consistency loops are required. Positron states in solids are a typical example.

-
- ¹ R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.*, **61**, 689 (1989).
- ² R. N. Barnett and U. Landman, *Phys. Rev. B* **48**, 2081 (1993).
- ³ G. E. Kimball and G. H. Shortley, *Phys. Rev.* **45**, 815 (1934).
- ⁴ H. Sugimoto and Y. Fukai, *Phys. Rev. B* **22**, 670 (1980).
- ⁵ M. J. Puska and R. M. Nieminen, *Phys. Rev. B* **29**, 5382 (1984).
- ⁶ M. J. Puska and R. M. Nieminen, *J. Phys. F* **13**, 333 (1983).
- ⁷ Y. Joly, *Phys. Rev. Lett.* **68**, 950 (1992); **72**, 392 (1994).
- ⁸ S. Baroni and P. Giannozzi, *Europhys. Lett.* **17**, 547 (1992).
- ⁹ J. R. Chelikowsky, N. Troullier, and Y. Saad, *Phys. Rev. Lett.* **72**, 1240 (1994).
- ¹⁰ M. P. Teter, M. C. Payne, and D. C. Allan, *Phys. Rev. B* **40**, 12255 (1989).
- ¹¹ M. C. Payne, M. P. Teter, D.C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- ¹² D. R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* **43**, 1494 (1979). We use the pseudopotentials parametrized by R. Stumpf, X. Gonze, and M. Scheffler (unpublished). The nonlocality is treated using the technique proposed by Kleinmann and Bylander [*Phys. Rev. Lett.* **48**, 1425 (1982)].
- ¹³ J. Arponen and E. Pajanne, *Ann. Phys. (N.Y.)* **121**, 343 (1979); *J. Phys. F* **9**, 2359 (1979).
- ¹⁴ L.-V. Wang and A. Zunger, *Comput. Mat. Sci.* **2**, 326 (1994).
- ¹⁵ E. Boroński and R. M. Nieminen, *Phys. Rev. B* **34**, 3820 (1986).
- ¹⁶ We use one node of an IBM SP1 cluster.
- ¹⁷ F. Gygi, *Phys. Rev. B* **48**, 11 692 (1993).

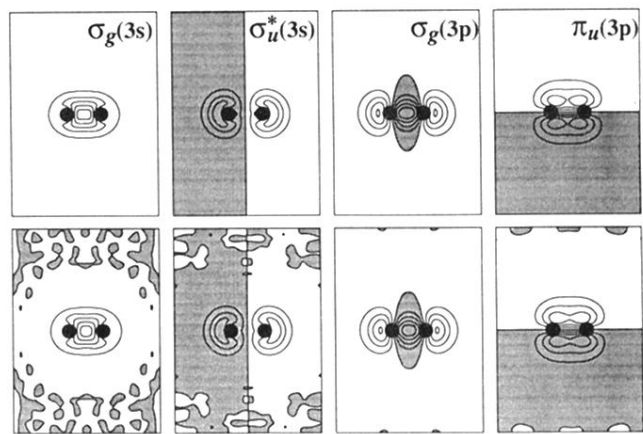


FIG. 2. Molecular orbitals of the P_2 dimer. The upper and lower panels show results from the finite-difference and plane-wave calculations, respectively. The contour spacing is one-sixth of the maximum value. The regions of negative values are shadowed and the positions of nuclei are given as black dots.