

# Efficient multiscale algorithms for solution of self-consistent eigenvalue problems in real space

Nimal Wijesekera<sup>1</sup>, Guogang Feng<sup>2</sup> and Thomas L. Beck<sup>1,2</sup>

<sup>1</sup>*Department of Physics*

<sup>2</sup>*Department of Chemistry*

*University of Cincinnati, Cincinnati, OH 45221-0172*

*nimalw@physics.uc.edu, fengg66@yahoo.com*

*becktl@email.uc.edu*

(Dated: February 3, 2005)

Real-space multiscale methods provide efficient algorithms for large-scale electronic structure calculations. In this paper we present novel multigrid strategies for solving self-consistent problems in density functional theory. The Full Approximation Scheme formulation of the multigrid method allows one in principle to move the expensive orthogonalization and Ritz projection operations to coarse levels. In addition, the effective potential may be updated on coarse levels during multiscale processing of the eigenfunctions. We investigate modifications of a previously proposed algorithm which are necessary to yield robust convergence rates. With these modifications, converged results are obtained within ten self-consistency iterations without orthogonalization or Ritz projection for the full occupied subspace on the fine level. Calculations are performed on three many-electron examples: the benzene, benzenedithiol, and glycine molecules. Recently developed relativistic separable dual-space Gaussian pseudopotentials are utilized to remove the core electrons.

PACS numbers:

## I. INTRODUCTION:

The solution of large-scale electronic structure problems involving hundreds of wavefunctions is a computationally demanding task. In recent years, a great deal of effort has been devoted to the development of efficient numerical methods for tackling these problems. The intense interest is stimulated by challenging applications arising in the realistic modeling of systems encountered in physics, chemistry, biophysics and the emerging nanoscale sciences. The main existing approaches for large-scale *ab initio* calculations can be roughly categorized as plane-wave basis set[1], localized numerical [?] or Gaussian basis set[2, 3] and real-space methods [4]. One appeal of the traditional basis set or real-space methods is the localized nature of the representation. This property has led to the development of methods whose complexity scales linearly with the number of electrons [? ]. In real-space calculations, the discretized partial differential equations are represented on grids using either finite-element or finite-difference formulations. The solutions can be obtained using iterative techniques due to the banded nature of the Hamiltonian. On single grids, however, stalling is a major problem of the iterative methods. The stalling results from the long-wavelength modes of the errors which are not efficiently removed on the fine scale[? ]. Multigrid methods overcome the stalling problem by decimating errors on multiple length scales. The short-wavelength errors are smoothed on fine levels, while the long-wavelength components are removed on coarse levels during the multigrid cycles. In the past a number of real-space multigrid solvers have been developed to solve the Kohn-Sham equations of density functional theory [5–14].

Previously we implemented a Full Approximation Scheme (FAS) nonlinear multigrid method [15] in solving the self-consistent Kohn-Sham equations [6]. Utilization of the FAS approach led to a significant improvement in convergence efficiency compared with linearized multigrid methods [? ? ]. In the original eigenvalue method of Ref. [15], the expensive orthogonalization and Ritz projection operations are performed on the fine level. Costiner and Ta’asan [?] advanced the FAS method by also moving the fine-scale separation techniques (orthogonalization and Ritz) to coarse levels. The FAS representation of the Ritz projection leads to a generalized eigenvalue problem on the coarse levels. They also introduced a backrotation step designed to prevent rotations in degenerate subspaces, sign changes, rescalings, and permutations of the eigenvectors. The overall algorithm is termed generalized Ritz/backrotation (GRBR). The backrotation step is necessary since the fine scale functions are corrected following the GRBR process on the coarse grid, and therefore the coarse-level eigenfunctions must correspond to their fine-level counterparts. In a second contribution, they adapted their eigenvalue method to solve large-scale self-consistent problems. They considered two approaches for handling the self consistency. First, the Poisson and eigenvalue problems were solved in a sequential fashion: once the effective potential was obtained on the finest level via solution of the Poisson equation (given a fixed updated charge density), the restricted potential was used during the multigrid solution for the eigenfunctions. Thus, the method cycles back and forth between the Poisson and eigenvalue problems much as in a traditional self-consistency code. Second, they updated the potential on coarse levels simultaneously with updates of the eigenfunction. In this way, the eigenfunctions and potential evolve toward the self-

consistent solution together on the coarse levels.

Here we present work aimed at applying these methods to non-periodic systems such as large molecules having tens of wavefunctions. Relativistic separable dual-space Gaussian pseudopotentials [16] are utilized to remove the core electrons, leaving 15-21 eigenfunctions for the test cases considered here (gly2? 30). The Gram-Schmidt orthogonalization and Ritz projection steps on the fine level (labeled by  $h$ , the fine grid spacing) scale as  $q^2 N_g^h$ , where  $q$  is the number of eigenfunctions and  $N_g^h$  is the total number of grid points. This scaling assumes the eigenfunctions span the whole physical domain. Moving these steps to coarser levels via the GRBR algorithm reduces the cost by a factor of eight for each level. Even though the coarse-level (labelled by  $H$ ) generalized Ritz process scales as  $q^2 N_g^H$ , Costiner and Ta'asan observed an effective  $q N_g^h$  algorithmic scaling which corresponds to the cost of updating the eigenfunctions on the fine level.

In our earlier paper [7], we applied the GRBR algorithm to fixed potential eigenvalue problems and self-consistent electronic structure problems. The fixed potential and small test self-consistent problems possessed clearly defined eigenvalue cluster structures; the eigenvalue clusters consist of degenerate or near-degenerate subspaces of the full occupied subspace (later in this paper we use the term cluster to denote any ordered and partitioned collection of eigenvalues). We showed that the GRBR method converged for the fixed-potential and small self-consistent problems such as the CO molecule and the Ne atom, all of which possess clearly defined eigenvalue clusters. The convergence rates slowed, however, after a few self-consistency iterations. The convergence rates could be restored by performing Gram-Schmidt orthogonalization occasionally on the fine level (once in every 3 – 5 self-consistency cycles). For larger systems such as benzene dithiol (which does not possess clearly defined eigenvalue clusters), the GRBR algorithm in its original form stalled prior to full convergence. The method did converge (slowly) with periodic fine-scale orthogonalization for the benzene molecule with its clear eigenvalue cluster structure.

These results led us to conclude that, for large molecules having tens of states and ambiguous eigenvalue cluster structures, GRBR alone could not bring sufficient separation of the wavefunctions on the fine level, and this results in stalling. The reasons for this are likely due to the relatively large number of wavefunctions of these large molecules, vague eigenvalue near-degeneracies (clusters), and the complexities of the potentials relative to those considered in [17, 18].

In this paper we describe a method to overcome these difficulties by performing Gram-Schmidt orthogonalization and Ritz projection on predetermined overlapping eigenvalue clusters on the fine level. These clusters need not necessarily correspond to degenerate subspaces. This is performed along with GRBR which is done on a coarser grid. Since the Gram-Schmidt

orthogonalization and Ritz projection are performed on eigenvalue/eigenfunction clusters, it is computationally considerably less expensive than the regular Gram-Schmidt orthogonalization and Ritz projection methods performed on the full occupied subspace on the finest level. Further we successfully tested this method with the simultaneous multigrid technique where the potential is updated on coarse levels. The calculations were all performed at the Kohn-Sham LDA level using a 12th-order finite difference representation of the Hamiltonian.

The structure of the paper is as follows. First we describe the fine-level Ritz projection multigrid method [15]. Then the GRBR algorithm and the fine-level Ritz projection on preselected clusters are discussed. Next the simultaneous update of the potential on coarse levels is described. This is followed by discussions of the application of the relativistic separable dual-space Gaussian pseudopotentials and the iterative relaxation scheme. In the section concerning computational implementation, a detailed description of the technical details of the numerical calculations is given. The convergence results for the three large-molecule test cases are then presented. Finally, we discuss the algorithms and computational results and present our conclusions.

## II. MULTIGRID CYCLE WITH FINE-LEVEL RITZ PROJECTION

Let us denote the eigenvalue problem with  $q$  wavefunctions by

$$HU = U\Lambda. \quad (1)$$

Here  $H$  is the Hamiltonian which is an  $N_g^h \times N_g^h$  square matrix where  $N_g^h$  is the total number of grid points in the real-space domain on the fine level. We label the finest grid level  $l$  and the coarser levels with  $k$  (the coarsest level has the smallest  $k$ ). The matrix  $U$  is a  $q \times N_g^h$  matrix and its columns are the wavefunction values on the grid. The matrix  $\Lambda$  is a  $q \times q$  diagonal matrix with the eigenvalues along the diagonal. We denote the exact grid solution  $U$  and the current approximation  $u$ . We obtain an initial fine-scale approximation  $u$  via a full multigrid cycle as discussed in Section VIII.

Once an initial approximation is obtained on the finest level  $l$  and a few relaxation steps are performed there, the problem is passed to the next coarser level  $k$ . On the coarse level, the eigenvalue problem takes the form

$$H^k u^k = u^k \Lambda + \tau^k. \quad (2)$$

where  $\tau^k$  is the defect correction defined as

$$\tau^k = I_k^{k+1} \tau^{k+1} + H^k (I_{k+1}^k u^{k+1}) - I_{k+1}^k (H^{k+1} u^{k+1}). \quad (3)$$

The operator  $I_{k+1}^k$  is the restriction operator which involves a local trapezoid-rule average of the function. The defect correction is zero on the finest level. The first term

in eq. 3 includes a contribution from the previous level's defect correction on grids at least two levels removed from the finest scale. If the exact solution from the fine grid is inserted into the coarse-grid eigenvalue equation, it is easily seen that an identity is obtained where the restricted fine-grid eigenfunction solves the coarse-grid equation. This formulation thus satisfies the important condition of zero correction at convergence.

We also note that the same formulation is used to solve the Poisson equation during the self-consistent solution. For the Poisson problem the Hamiltonian in eqn. 2 is replaced by the Laplacian operator, and the first term on the rhs of eqn. 2 is replaced by  $-4\pi\rho$  on the coarse level, where  $\rho$  is the charge density. The coarse-grid density is obtained by restricting the fine-grid density.

On the coarse level, iterative relaxation is performed subject to constraints designed to maintain eigenfunction orthonormality on the finest grid [15]. We also explored use of Kaczmarz relaxation on the coarse levels as in refs. [17]; no clear advantage was observed, however. A correction is then made back on the next higher level  $k+1$  ( $k+1=l$  for the finest level):

$$u^{k+1} = u^{k+1} + I_k^{k+1}(u^k - I_{k+1}^k u^{k+1}). \quad (4)$$

where  $I_k^{k+1}$  is the interpolation operator. Either linear or cubic interpolation was used throughout our work (specified below).

Once back on the finest grid, the eigenvalues can be updated using the Rayleigh quotient:

$$\lambda = \langle u^l | H^l | u^l \rangle / \langle u^l | u^l \rangle. \quad (5)$$

Following relaxation steps on the fine level, Gram-Schmidt orthogonalization and Ritz projection are performed to improve the occupied subspace. The complete process of moving to a sequence of coarser levels followed by corrections and relaxations steps on succeeding finer levels until the finest level is reached is termed a V-cycle. The Ritz projection step is discussed in the next section (see eqs. 8 and 11 for the fine-scale version). The above discussion can be extended to any number of levels. For the Poisson problem, the multigrid cycle is continued all the way up to a coarsest level which has only a single interior point. For the eigenvalue problem, the coarsest level must possess enough structure to (at least approximately) sample the oscillations in the eigenfunctions. For the eigenvalue problem, we utilized three grids levels ( $17^3$ ,  $33^3$ , and  $65^3$ ) in all of the calculations reported here.

### III. GENERALIZED RITZ PROJECTION (GRP) AND BACK ROTATION (BR)

The eigenvalue problem can be rewritten using the relation

$$Hv = v\Lambda \quad (6)$$

where

$$v = uE \quad (7)$$

Thus

$$HuE = uE\Lambda \quad (8)$$

Here  $v$  and  $u$  are  $q \times N_g^h$  matrices and  $E$  is a  $q \times q$  matrix to be found. When eqn. 8 is transferred to a coarse level using an FAS transfer, it takes the form

$$HuE = uE\Lambda + \tau E \quad (9)$$

where  $\tau$  is the defect correction vector. Multiplying eqn. 9 on the left by  $u^T$  the generalized eigenvalue problem results:

$$u^T(Hu - \tau)E = (u^T u)E\Lambda \quad (10)$$

This can be solved for  $E$  using standard linear algebra packages. The solution gives  $E$  and a new set of eigenvalues  $\lambda$ . New wavefunctions and defect corrections  $\tau$  are obtained as linear combinations of the previous values:

$$v = u^{new} = u^{old} E \quad (11)$$

$$\tau^{new} = \tau^{old} E \quad (12)$$

On the fine grid where the defect correction  $\tau$  is zero, the generalized Ritz projection reduces to the standard Ritz projection. Notice that the eigenvalues are the same on all levels at convergence by construction.

In the generalized Ritz projection method, rotations of solutions in subspaces of close eigenvalues, rescaling, sign changes, and permutations of the solutions may occur. Therefore the backrotation (BR) algorithm was introduced by Costiner and Ta'asan [17] to deal with these problems. The backrotation algorithm goes as follows. Once  $E$  and  $\lambda$  are found by solving eqn. 10, the eigenvalues in  $\Lambda$  are sorted and the columns of the  $E$  matrix are permuted accordingly. The clusters of degenerate and non-degenerate clusters of  $\Lambda$  are identified. For each diagonal block of  $E$  the dominant elements of the block are brought to the diagonal by permuting the columns of  $E$  and the diagonals of  $\Lambda$ . Define  $F$  as a block diagonal matrix having only the diagonal blocks of  $E$ . Each non-degenerate diagonal block of  $F$  is replaced with the identity matrix. Set  $E = EF^{-1}$  and change the signs of the columns of  $E$  such that the diagonal elements are all positive. The columns of  $E$  are normalized and the wavefunctions and the defect corrections are updated as per eqns. 11 and 12. Extensive details of the backrotation step are given in ref. [17].

Generalized Ritz projection coupled with backrotation (GRBR) performed on a coarse level is supposed to bring separation of eigenstates on the fine level. But this worked with limited success in our case studies involving self-consistent electronic structure calculations [7].

#### IV. FINE-SCALE RITZ PROJECTION PERFORMED ON CLUSTERS ALONG WITH GRBR

As discussed above, the original GRBR algorithm stalled for large systems with ambiguous eigenvalue cluster structures. We found that the incorporation of the following algorithm in addition to GRBR could alleviate those difficulties. My E method?? Refer to previous discussion. New results?? Mine with periodic GS?

The general idea is to perform fine-scale Ritz projection on previously selected eigenvalue clusters in addition to the GRBR processing on coarse levels. Cluster structures are chosen at the beginning of the multigrid V-cycles on the fine level. As an example suppose we have a system with six eigenfunctions. The eigenvalues along with the wavefunctions are sorted in the increasing order as  $\lambda_1 \dots \lambda_6$ . Suppose that two clusters are chosen. Fine-level orthogonalization and Ritz projection are performed on two overlapping clusters such as  $(\lambda_1 \dots \lambda_4)$  and  $(\lambda_4 \dots \lambda_6)$ . The inclusion of the overlap element ( $\lambda_4$ ) during the fine-level separation was found to be crucial. The coarse-level (middle level in all our calculations) GRBR process is performed on clusters with no overlap such as  $(\lambda_1, \lambda_2, \lambda_3)$  and  $(\lambda_4, \lambda_5, \lambda_6)$ . This brought good enough separation of the wavefunctions on the fine level to achieve full convergence for the cases where GRBR alone stalled.

This fine-level separation technique is relatively inexpensive compared to performing orthogonalization and Ritz Projection on the full occupied subspace as in the regular Ritz projection method. If  $m$  clusters consisting of  $p$  wavefunctions are chosen, the cost of orthogonalization and Ritz projection on clusters scales as  $mp^2N_g^h$ . The ratio of the scalings of these two methods goes as  $1/m$ . Therefore there will be an  $m$ -fold decrease in computational time for large systems.

If small clusters are chosen  $m$  will be larger and the savings of computational time will be higher. But in our experience it is found that, with too small clusters, the algorithm tends to stall. The size of the clusters can be gradually increased until a suitable size is found. Using this approach, we observed comparable convergence rates to the full fine-level Ritz projection. LARGER SYSTEM??

#### V. SIMULTANEOUS MULTIGRID METHOD

Due to the nonlinear FAS formulation of the self-consistent problem, it is also possible to update the coarse-grid effective potential simultaneously with the eigenfunctions [18]. This contrasts with the sequential algorithm in which the fixed coarse-grid potential is the restricted value from the fine level.

The simultaneous update of the potential on the coarse grids requires some care. The density must be updated on the coarse grids from which the coarse-level

effective potential is obtained. A key advantage of the FAS approach is that the coarse-level density and potential terms can be constructed with defect correction terms such that, at convergence, the coarse-level values are precisely those restricted from the fine grid. The fine-grid eigenfunctions are normalized so as to preserve charge conservation throughout the various grids at convergence. This yields the optimal representation of the coarse-grid potential. In order to effect these points, new defect corrections are required for the charge density and the exchange-correlation potential.

We define  $\bar{\rho}^k(u^k)$  for coarse levels  $k$  as

$$\bar{\rho}^k(u^k) = 2 \sum_{i=1}^q |u_i^k|^2 \quad (13)$$

It is important to note that  $\bar{\rho}^k(u^k)$  is not a true charge density on coarse levels since, even at convergence, the restricted eigenfunctions are neither normalized nor orthogonal on coarse grids. Once wavefunctions are updated on level  $k$ , the quantity  $\bar{\rho}^k$  is generated using (eqn. 13). We then construct the coarse level density via

$$\rho^k = \bar{\rho}^k(u^k) + I_{k+1}^k \rho^{k+1} - \bar{\rho}^k(I_{k+1}^k u^{k+1}) \quad (14)$$

It is clear that, at convergence, the first and third terms cancel and we are left with the charge density restricted from the finer level. Therefore, the charge conservation is maintained on all levels since the restriction operation preserves the total charge. The last two terms can be defined as a defect correction for  $\rho$ . Once  $\rho^k$  is updated, the Coulomb potential can be found by solving the coarse-grid Poisson problem using again the FAS technique [4]. Using this FAS approach with the above-defined charge density, the coarse-level Coulomb potential is precisely the restricted fine-level potential.

On the coarse level  $k$  the exchange-correlation potential  $v_{xc}^k$  is updated as soon as  $\rho^k$  is obtained. We define a form for the coarse-grid  $v_{xc}^k$  which also yields the restricted fine-level  $v_{xc}^l$ :

$$v_{xc}^k = v_{xc}^k(\rho^k) + I_{k+1}^k v_{xc}^{k+1} - v_{xc}^k(I_{k+1}^k \rho^{k+1}) \quad (15)$$

The last two terms of eqn. 15 can be considered a defect correction for  $v_{xc}^k$ . For  $v_{xc}$  we have used either the VWN form or the formula given in [19]. The pseudopotentials (presented below) do not depend on the valence charge density, so they do not require modification on coarse levels.

#### VI. PSEUDOPOTENTIALS

Relativistic separable dual-space Gaussian pseudopotentials [16] are used in all of our studies. They consist of local and nonlocal components. The local component

is

$$v_{loc}(r) = \frac{-Z_{ion}}{r} \operatorname{erf}\left(\frac{r}{\sqrt{2}r_{loc}}\right) + \exp\left[-\frac{1}{2}\left(\frac{r}{r_{loc}}\right)^2\right] \\ \times \left[ C_1 + C_2\left(\frac{r}{r_{loc}}\right)^2 + C_3\left(\frac{r}{r_{loc}}\right)^4 \right. \\ \left. + C_4\left(\frac{r}{r_{loc}}\right)^6 \right] \quad (16)$$

Here erf is the error function and  $Z_{ion}$  is the net charge of the nucleus plus the core electrons.  $C_1, C_2, C_3, C_4$  and  $r_{loc}$  are constants. The variable  $r$  is the radial distance from the nucleus.

The nonlocal component is

$$v_l(\mathbf{r}, \mathbf{r}') = \sum_{i=1}^3 \sum_{j=1}^3 \sum_{m=-l}^l Y_{l,m}(\hat{\mathbf{r}}) p_i^l(r) h_{i,j}^l p_j^l(r') Y_{l,m}^*(\hat{\mathbf{r}}') \quad (17)$$

The  $Y$ 's are spherical harmonics and  $l$  is the angular momentum quantum number. The  $h_{i,j}^l$ 's are given parameters. The values of the constants  $C$ s and  $h_{i,j}^l$  are given in the original paper [16]. The full nonlocal portion of the pseudopotential is constructed by summing over the  $l$  components.

The variable  $p_i^l(r)$  is a projector and is defined as

$$p_i^l(r) = \frac{2r^{l+2(i-1)} \exp\left(-\frac{r^2}{2r_i^2}\right)}{r_i^{l+(4i-1)/2} \sqrt{\Gamma\left(l + \frac{4i-1}{2}\right)}} \quad (18)$$

where  $\Gamma$  is the gamma function. The projectors satisfy the normalization condition

$$\int_0^\infty p_i^l(r) p_i^l(r) r^2 dr = 1 \quad (19)$$

The separable nature of the pseudopotential is utilized to obtain an efficient real-space integration for the pseudopotential acting on the eigenfunctions [20]. The decay properties of the projectors allow restriction of the domain to a small region around each nucleus (specified below). Spin-orbit effects were neglected for the calculations presented here (check with Nimal??). On coarse levels, the pseudopotential is applied in the same form as on the fine level over the same physical domain (but sampling fewer grid points).

## VII. RELAXATION WITH SHIFT PARAMETER

Consider an eigenvalue equation for a single state (where  $\lambda$  and  $u$  are the current approximations):

$$Hu = \lambda u \\ H = L + D + U \quad (20)$$

The Hamiltonian matrix  $H$  is written using the lower-triangular part  $L$ , the diagonal part  $D$ , and the upper-triangular part  $U$ .

Originally we used Gauss-Seidel relaxation with a shift parameter  $\mu$  for the eigenvalue problem [15]:

$$u^{n+1} = u^n + (D - \mu I)^{-1} [\tau - (Du^n + Lu^{n+1} + Uu^n - \lambda Iu^n)]. \quad (21)$$

In the iteration equation  $u^{n+1}$  is the updated vector and  $u^n$  is the old vector.

Later we extended the iterative scheme to successive overrelaxation (SOR) with a shift parameter  $\mu$  and found improved convergence rates in our case studies:

$$u^{n+1} = u^n + (D - \mu I)^{-1} [\tau - (Du^n + Lu^{n+1} + Uu^n - \lambda Iu^n)]\omega. \quad (22)$$

where  $\omega$  is the overrelaxation parameter and  $\omega = 1$  gives the usual Gauss-Seidel relaxation (eqn. 21).

Choosing the appropriate parameters  $\omega$  and  $\mu$  for the relaxation scheme is important; otherwise consistent convergence may not be obtained. We determined optimal convergence parameters for the relaxation by numerical experimentation during the test calculations. Values used for  $\omega$  and  $\mu$  are given for all the cases we studied in (Tables I,II, and III). When nonlocal pseudopotentials are present the diagonal term  $D$  of the  $H$  matrix has to include the effect of the nonlocal pseudopotentials. For that purpose  $\int v(\mathbf{r}, \mathbf{r}') d\mathbf{r}'^3$  is added to  $D$ . It is worth noting that  $\int v(\mathbf{r}, \mathbf{r}') d\mathbf{r}'^3$  can be evaluated and stored at the beginning of the program since it does not involve the wavefunctions. We use this SOR scheme only on the fine grid; Gauss-Seidel relaxation is used on coarse levels (why??). Why integrate PP as above?? Ref GF paper.

## VIII. COMPUTATIONAL IMPLEMENTATION

Computational results for self-consistent eigenvalue problem for non periodic systems with pseudopotentials are presented. Benzene, glycine and benzenedithiol molecules are used in this study. Benzene and glycine each have 15 wavefunctions and benzenedithiol has 21 wavefunctions. The number of wavefunctions comes as a result of using pseudopotentials [16, 19]. If all electron calculations were done benzene, glycine and benzenedithiol will have 21, 20 and 37 wavefunctions respectively. The reason for choosing these molecules are due to different structures and degeneracies they present. Benzene molecule has clear symmetry and degeneracies of eigenvalues while glycine has no symmetry and shows almost no degeneracy. Benzenedithiol shows moderate degeneracy and it has a higher number of wavefunctions (21).

All the calculations are performed on a three-dimensional real-space grid using the finite-difference method with 12th-order accuracy for the Laplacian. The trapezoid rule is used for integration formula. Three

grid levels are utilized, comprising  $17^3$ ,  $33^3$  and,  $65^3$  grid points. Gauss-Seidel (equ.21) and SOR (equ.22) relaxation schemes with a shift parameter  $\mu$  are used. Interpolation and restriction operations are performed using linear interpolation and full weighting average respectively. The SOR relaxation scheme is used only on the fine grid ( $65^3$ ), and on the coarse grids ( $17^3$  and  $33^3$ ) the Gauss-Seidel method is always used. In describing the algorithms, the application of the nonlocal part of the pseudopotential operator on a wavefunction is denoted by  $NPP$  which is a function of  $\mathbf{r}$ :

$$NPP = \int v_{nl}(\mathbf{r}, \mathbf{r}') u(\mathbf{r}') d\mathbf{r}'^3, \quad (23)$$

where  $v_{nl}(\mathbf{r}, \mathbf{r}')$  is the nonlocal part of the pseudopotential, and  $u(\mathbf{r}')$  is the wavefunction.

## IX. RITZ PROJECTION METHOD

Initially on the coarse grid ( $A$  in Fig. 1), random numbers are assigned to all wavefunctions except boundary points which are set to zero and fixed throughout the calculations. Next, the wavefunctions are orthogonalized and the effective potential is found. This is followed by the application of  $NPP$ . All these operations prepare the Hamiltonian matrix, and in the next step the Ritz projection uses this matrix. The  $NPP$  is evaluated once again and the eigenvalues are updated using the Rayleigh quotient. The wavefunctions are relaxed once using the Gauss-Seidel method and normalized. The last four operations are repeated twice, and all the operations are repeated about 50 times. At this point, the residues of eigenvalues are around  $10^{-5}$ , and further iterations have almost no effect in reducing the residues. Next, the wavefunctions are interpolated to the next higher level ( $B$  in Fig. 1) and the same operations which are performed at  $A$  are repeated once at  $B$ . The wavefunctions are restricted to the coarse level ( $C$  in Fig. 1), and the defect correction is found. On the coarse grid  $C$ , the application of  $NPP$ , relaxation, and constraints are repeated about 5 times. The solution is then passed to the next higher level ( $D$  in Fig. 1) by an FAS correction step. The same operations which are performed at  $B$  are repeated once at  $D$ . This completes the two-level multigrid  $v$ -cycle. The two-level  $v$ -cycle is repeated about 20 times, and the solution is interpolated to the next higher level ( $E$  in Fig. 1) which is the fine level in these calculations. The interpolated wavefunctions provide a better initial guess on the fine level, and this is the main advantage of using the full multigrid scheme (FMG). Orthonormalization of wavefunctions, update of the effective potential, application of  $NPP$ , and Ritz projection are performed once. This is followed by the application of  $NPP$ , updating the eigenvalues, relaxation (SOR) and normalization twice, the application of  $NPP$ , and relaxation (SOR) and normalization twice. The convergence rate can be improved by doing more relaxations on the fine grid though it has a

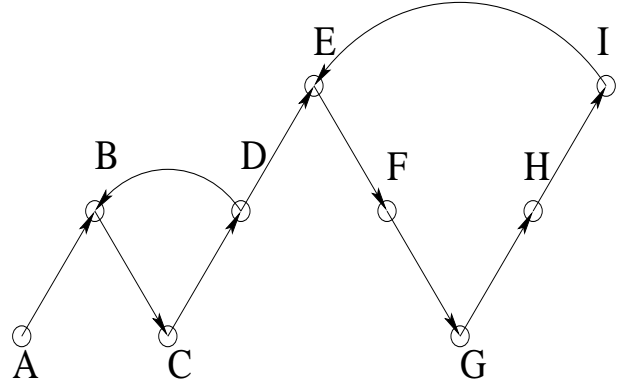


FIG. 1: This figure shows the three-level full multigrid process (FMG).

limit. We increased the number of relaxations to 8 per  $v$ -cycle, 4 at each end of the  $v$ -cycle by following the above order of operations on the fine grid. Further increase in the number relaxations did not improve the convergence rate. Next, the solution is restricted to the middle level ( $F$  in Fig. 1), where the defect correction is found first. The application of  $NPP$ , relaxation, and constraint are repeated about 5 times at  $F$ , and an additional application of  $NPP$  is performed. Then the solution is restricted to the next coarser level ( $G$  in Fig. 1), and the defect correction is evaluated. The operations which are performed at  $C$  are performed about 20 times at  $G$ . The solution is transferred to the next higher level ( $H$  in Fig. 1) by an FAS correction step. The same operations which are done at  $F$  are performed at  $H$ , and the solution is passed to the fine level ( $I$  in Fig. 1) using an FAS correction. The same operations which are done at  $E$  are performed at  $I$ . This completes the three-level  $v$ -cycle. The three-level  $v$ -cycle can be repeated until a desired accuracy is reached.

On the fine level ( $65^3$ ), the Gram-Schmidt orthonormalization and the Ritz projection are performed only twice (once at each end of the  $v$ -cycle) and relaxation is performed altogether 8 times. The Poisson problem is solved twice at each end of the  $v$ -cycle. The Gauss-Seidel and the SOR relaxation schemes with a shift parameter  $\mu$  are used, and the numerically optimized values of the parameters ( $\mu$  and  $\omega$ ) are shown in Tables. I, II, and III.

## X. RITZ PROJECTION PERFORMED ON CLUSTERS ALONG WITH GRBR

The implementation of the Ritz projection performed on clusters along with GRBR is done on the three-level  $v$ -cycle (E-F-G-H-I in Fig. 1). The Ritz projection method is used prior to that. Therefore, only the three-level  $v$ -cycle is described.

Upon entering the three-level  $v$ -cycle ( $E$  in Fig. 1), the normalization, update of the effective potential, orthonormalization on clusters, application of  $NPP$ , and Ritz projection on clusters are performed only once. The reason

| System         | $\mu$ | $\omega$ |
|----------------|-------|----------|
| Benzene        | -30.0 | 1.0      |
| Glycine        | -20.0 | 1.0      |
| Benzenedithiol | -30.0 | 1.0      |

TABLE I: The values of  $\mu$  and  $\omega$  used on the coarsest level  $17^3$ .

for doing these operations once at entering the  $v$ -cycle is that it has brought smooth convergence initially.

Next at  $E$ , the wavefunctions are normalized and the effective potential is updated. This is followed by the application of  $NPP$ , relaxation and normalization (twice using SOR), Ritz projection on clusters, application of  $NPP$ , relaxation and normalization (twice using SOR), and application of  $NPP$ . Then the solution is restricted to the middle level ( $F$  in Fig. 1), and the defect correction is evaluated. This is followed by the application of  $NPP$ , relaxation, application of  $NPP$ , and GRBR. These operations are repeated about 5 times, and an additional application of  $NPP$  is made. The solution is restricted to the coarser grid ( $G$  in Fig. 1) and the defect correction is found. Application of  $NPP$ , relaxation, and constraints are performed about 20 times at  $C$ . Then the solution is transferred to the next higher level ( $H$  in Fig. 1) using an FAS correction step. The same operations which are done at  $F$  are performed at  $H$ . Next, the solution is transferred to the fine level ( $I$  in Fig. 1) by an FAS correction step. The operations which are done at  $E$  are performed at  $I$ . This completes the three-level  $v$ -cycle. It can be repeated until a desired accuracy is reached.

The following clusters are used for the Ritz projection performed on clusters along with GRBR method. For benzene two clusters of 8 and 7 elements are chosen and for glycine two clusters of 7 and 8 elements are used. For benzenedithiol which has 21 wavefunctions, two clusters of 10 and 11 elements are chosen. The values of the parameters ( $\mu$  and  $\omega$ ) which are used for the relaxation schemes are shown in Tables I, II, and III. These values are obtained through numerical optimization.

## XI. SIMULTANEOUS TECHNIQUE

The simultaneous method with Ritz projection and the simultaneous method with Ritz projection performed on clusters along with GRBR are performed more or less similar to the Ritz projection method and the Ritz projection performed on clusters along with GRBR method, respectively. The major difference is that the effective potential is updated on the coarse grids which is expected to bring improved convergence rates. However, in our calculations we did not find any major improvements in the convergence rates compared to the other methods. Slightly different values for  $\mu$  and  $\omega$  are used in the relaxation as shown in (Tables. I,II and III).

| System         | $\mu(c)$ | $\mu(m)$ | $\omega(c)$ | $\omega(m)$ |
|----------------|----------|----------|-------------|-------------|
| Benzene        | -30.0    | -20.0    | 1.0         | 1.0         |
| Glycine        | -20.0    | -20.0    | 1.0         | 1.0         |
| Benzenedithiol | -30.0    | -20.0    | 1.0         | 1.0         |

TABLE II: The values of  $\mu$  and  $\omega$  for the two-level multigrid process. The levels (coarse (c) and middle (m)) are specified within the bracket.

| System         | $\mu(c)$ | $\mu(m)$ | $\mu(f)$ | $\omega(c)$ | $\omega(m)$ | $\omega(f)$ | $\omega_1(f)$ |
|----------------|----------|----------|----------|-------------|-------------|-------------|---------------|
| Benzene        | -20.0    | -20.0    | 0.0      | 1.0         | 1.0         | 1.7         | 1.5           |
| Glycine        | -20.0    | -20.0    | 0.0      | 1.0         | 1.0         | 1.7         | 1.5           |
| Benzenedithiol | -20.0    | -20.0    | 0.0      | 1.0         | 1.0         | 1.7         | 1.5           |

TABLE III: The values of  $\mu$  and  $\omega$  for the three-level multigrid process are shown. The levels (coarse (c), middle(m), and fine(f)) are specified within the bracket. The values of  $\omega$  in the last column are used in the simultaneous algorithm.

### A. Poison Problem and Exchange Correlation Potential

In all these calculations except simultaneous method the Poison equation (equ.24) is always solved on the fine grid for the whole system (electrons plus nuclei) and then the nuclear contribution is replaced by the smooth local component of the pseudopotential (equ.16).

$$\nabla\phi = -4\pi\rho \quad (24)$$

Boundary values for the poison problem are given using multipole expansion up to quadrupole term. On coarse levels restricted values of the coulomb potential are used. Since the nuclear charge distribution for a large molecule can not always be placed on grid points the charges are distributed to the neighboring grid points. It is done by splitting the charge according to the fraction of volume. Fig. 2 shows how it is done in 2-dimension and it can be done similarly using volume in 3-dimension. Exchange-correlation potential is found using the formula given in

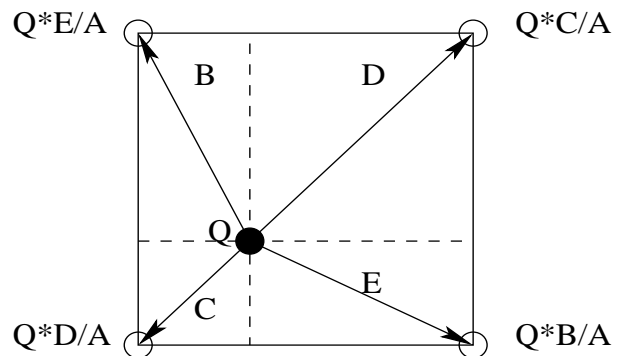


FIG. 2: This shows how the charge  $Q$  which is not on a grid point is distributed to neighboring grid points in 2-dimension.  $B, C, D$  and  $E$  are the divided areas and  $A$  is the total area of the square. This can be done in 3-dimension using volume

[19, p. 1709]. VWN formula for the exchange correlation potential can be used if desired, and it also works in our algorithm. On coarse levels restricted exchange correlation potential of the fine grid is used except in simultaneous techniques.

## B. Nonlocal Pseudopotential

The act of nonlocal component on a wavefunction (equ.??) can be written as

$$\begin{aligned} & \int v(\mathbf{r}, \mathbf{r}') u(\mathbf{r}') d\mathbf{r}'^3 = \\ & \times \sum_l \sum_{i,j=1}^3 \sum_{m=-l}^l Y_{l,m}(\hat{\mathbf{r}}) p_i^l(r) \\ & \times h_{i,j}^l \int p_j^l(r') Y_{l,m}(\hat{\mathbf{r}}) u(\mathbf{r}') d\mathbf{r}'^3 \end{aligned} \quad (25)$$

This form of  $\int v(\mathbf{r}, \mathbf{r}') u(\mathbf{r}') d\mathbf{r}'^3$  is possible because of the separable nature of the nonlocal pseudopotential. The advantage of this form is that the integration on the right side of (equ.25) can be done only once for an update of all the points  $\mathbf{r}$  of the nonlocal domain. In other words the integration on the right side of (equ.25) need not be done for each grid point  $\mathbf{r}$ .

The nonlocal component of the pseudopotential (equ.??) is performed only over a small cube around each nucleus. This is possible because of the radial projectors (equ.18) tend to zero outside the covalent radius of an atom. The domain sizes used in our calculations are 11 grid points (per side) on the fine grid and 9 grid points on the mid level and 5 grid points for the coarse level. (equ.??) is implemented on all grid levels without any modifications.

## XII. COMPUTATIONAL RESULTS

Convergence rates are analyzed by plotting  $\text{Log}_{10}|E - E_{exact}|$  vs number of  $v$ -cycle (self-consistent iterations) and  $E$  is the total energy given by (equ.??) [21, p. 1709].

The total energy  $E$  is evaluated and stored at the end of each three-level  $v$ -cycle. The total energy, calculated at the end of the algorithm, is considered as the converged value  $E_{con}$  of the total energy. The benzene, glycine, and benzenedithiol molecules are used to test the convergence efficiencies of the algorithms.

The Ritz projection method, the Ritz projection performed on clusters along with GRBR, and the simultaneous method show almost the same convergence rates for all test cases as shown in Figs. 4, 5, 6. In all these cases,  $\text{Log}_{10}|E - E_{con}|$  drops to  $-6$  in about 10  $v$ -cycles. The simultaneous method with the Ritz projection performed on clusters along with GRBR shows the slowest convergence rate in all three cases.

The most important result in this research is the good convergence rates shown by the Ritz projection performed on clusters along with the GRBR algorithm; this method scales as  $N_e^2$  on the fine grid, whereas the Ritz projection method scales as  $N_e^3$ .

Comparison of convergence efficiencies of four different algorithms including the multigrid method, is shown in Ref.[5] for the test cases of a 64-atom Si cell and a 64-atom diamond cell. Even though the systems ( 64-atom Si cell and a 64-atom diamond cell) studied are not exactly similar to our case studies, we used those convergence rates to compare the convergence rates of this research. The fastest algorithm in Ref.[5] is the multigrid algorithm with subspace diagonalization, and  $\text{Log}_{10}|E - E_{con}|$  reaches  $-6$  in about 40 self-consistent iterations Fig. 3. In our case studies,  $\text{Log}_{10}|E - E_{con}|$  drops to  $-6$  in about 10 self-consistent iterations ( $v$ -cycles) in all four algorithms except for the simultaneous method with Ritz projection performed on clusters along with GRBR; in that case,  $\text{Log}_{10}|E - E_{con}|$  reaches  $-6$  in about 20  $v$ -cycles. One reason for the faster convergence rates observed in our research is the use of the FAS nonlinear scheme instead of the linearized coarse-grid-correction scheme (CGC) [?] as used in the work of Ref.[5]. In the CGC method, the full function is relaxed on the fine grid, and the residual equation is relaxed for the error on the coarse grids. In the FAS scheme, the full function is relaxed on all grids, and it is possible to update the eigenvalues on the coarse grids. Highly efficient plane-wave based calculation of the Kohn-Sham ground state of metallic systems with pseudopotentials can be found in the study of Kresse and Furthmüller in Ref.[? ]. In their study, the logarithm of the total energy difference drops to below  $-6$  within 10 self-consistent iterations Fig. 3, and those results are comparable to our work.

It is important to note the degree of orthogonality of the wavefunctions of the Ritz projection performed on clusters along with the GRBR method. The orthogonality of the wavefunctions between different clusters increases gradually with the increasing number of  $v$ -cycles. Initially the degree of the orthogonality starts around  $10^{-3}$  ( $\langle u_i | u_j \rangle \simeq 10^{-3}$ ), and it improves to  $10^{-14}$  ( $\langle u_i | u_j \rangle = 10^{-14}$ ) at full convergence. The contour plots of density of wavefunctions could be obtained from these calculations as an additional outcome, and one of such plots is shown in Figs. 7.

## XIII. DISCUSSION

We have successfully tested four nonlinear FAS multigrid techniques for solving self-consistent eigenvalue problem using relativistic separable dual-space Gaussian pseudopotentials for non periodic systems such as molecules having tens of wavefunctions. These are the usual Ritz projection, Ritz projection on clusters along with GRBR, the simultaneous technique and the simultane-

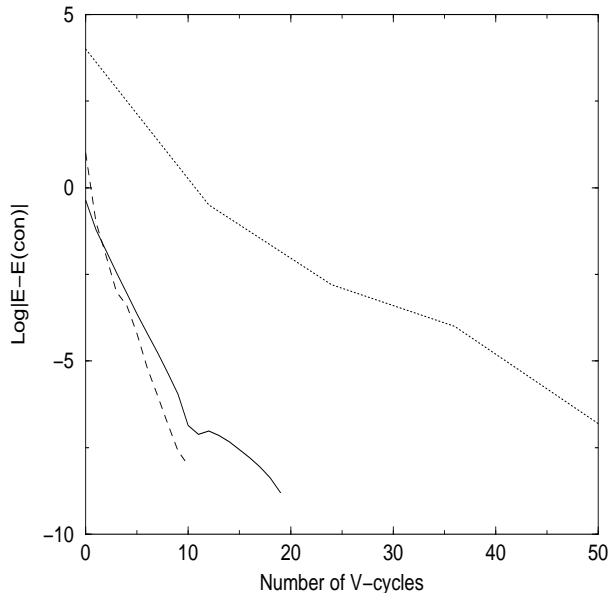


FIG. 3: This figure shows comparison of convergence rates to other works. The  $\log_{10}$  of the difference between the current total energy  $E$  a.u and the fully converged value  $E_{con}$  is plotted against the number of  $v$ -cycles (self-consistent iterations). The top curve is obtained using CGC method for 64-atom Diamond cell [5]. The solid curve in the bottom is the result of plane-wave calculations (metallic systems) [? ]. The dashed curve in the bottom is obtained by us for benzene using FAS multigrid method. The data for the top curve and the solid curve are obtained from Ref.[5] and Ref.[? ] respectively.

ous technique with Ritz projection on clusters along with GRBR.

Earlier we have shown [7] that the GRBR method worked with slower convergence rates for small molecules and atoms such as CO and Ne having fewer number of wavefunctions. Convergence rates could be restored performing occasional orthogonalization on the fine grid for these small molecules. But for large molecules such as benzene, glycine and benzenedithiol GRBR always stalled ( benzenedithiol ) or converged with much slower convergence rate (benzene). The reasons could be relatively large number of wavefunctions these molecules posses and the ambiguous degeneracies (clusters) presented and the potentials which are much less smooth compared to those of the smaller molecules and the those used by Costiner and Ta'asan [17, 18] in their calculations. Determining the clusters for large systems is another difficulty we encountered [7] since it has to be done automatically for large systems.

We have shown in this paper that the successful implementation of the generalized Ritz projection for large systems involving tens of orbitals could be achieved using the method of Ritz projection performed on clusters along with GRBR. In this method the cluster structures are chosen arbitrarily at the beginning of the three level

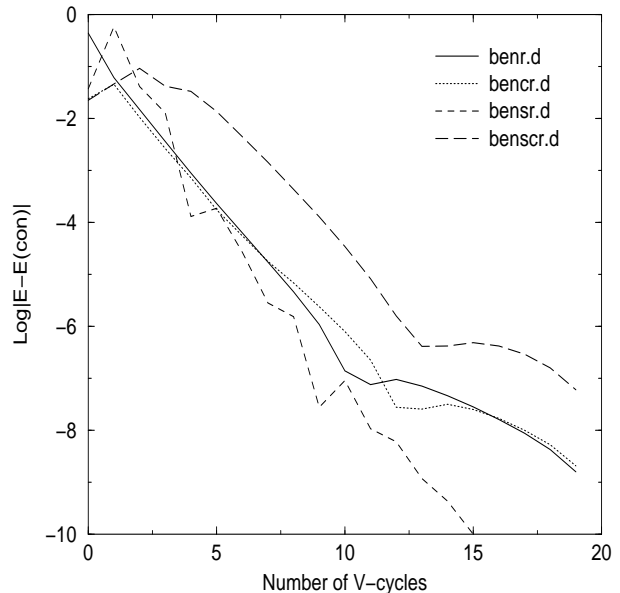


FIG. 4: Convergence rates for benzene. The  $\log_{10}$  of the difference between the current total energy  $E$  a.u and the fully converged value  $E_{con}$  is plotted against the number of  $v$ -cycles (self-consistent iterations). The plots are for regular Ritz projection (benr.d), Ritz projection performed on clusters along with GRBR (bencr.d), simultaneous method with Ritz projection (bensr.d) and for simultaneous method with Ritz projection performed on clusters along with GRBR (benscr.d). The fine grid spacing is  $h=0.3$  a.u.

$v$  - cycle. The Ritz projection is performed on those clusters on the fine grid in addition to GRBR on a coarse level (mid level in our cases). This removes the stalling observed in the original form of GRBR algorithm and good convergence rates are obtained in all cases studied.

In the method of Ritz projection on clusters along with GRBR the orthogonalization and the Ritz projection are performed on the clusters. Orthogonalization and Ritz projection performed on all the wavefunctions  $q$  scale as  $q^2 * N$  where  $N$  is the number of grid points. Performing these two operations on  $m$  overlapping clusters having  $p$  wavefunctions in each cluster scale as  $m * (p + 1)^2 * N$ . Therefore nearly an  $m$ -fold saving of computational time can be obtained over regular Ritz projection method. In our case studies it is about 50 percent saving. In addition to saving of computational time faster convergence rates (compared to Ritz projection methods) are obtained in all cases studied.

In these calculations the use of pseudopotentials play an important role since they provide smoother potentials around nuclei as opposed to the rough potentials of all electron calculations. Because of the projectors and the separable nature of the pseudopotentials, the implementation of the nonlocal part is not computationally expensive. Further it could be implemented on the coarse levels in the same way as on the fine grid without any modifications. On coarse levels the integration  $\int v(\mathbf{r}, \mathbf{r}')u(\mathbf{r}')d\mathbf{r}'^3$

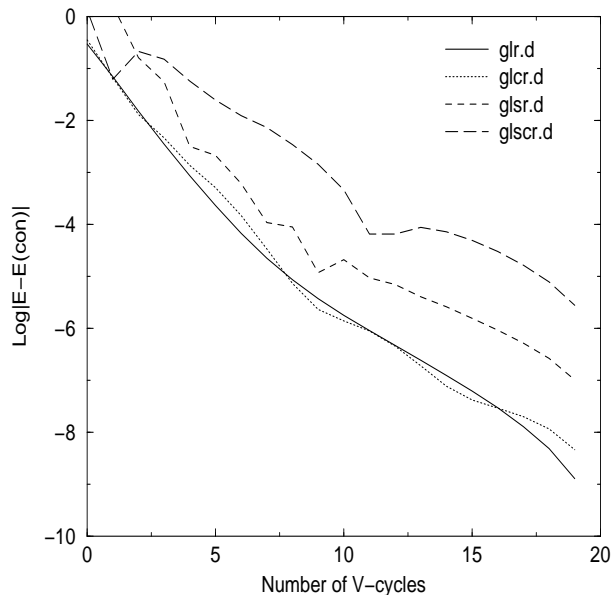


FIG. 5: Convergence rates for glycine. The  $\log_{10}$  of the difference between the current total energy  $E$  a.u and the fully converged value  $E_{con}$  is plotted against the number of  $v$ -cycles (self-consistent iterations). The plots are for regular Ritz projection (glr.d), Ritz projection performed on clusters along with GRBR (glcr.d), simultaneous method with Ritz projection (glsr.d) and for simultaneous method with Ritz projection performed on clusters along with GRBR (glscr.d). The fine grid spacing is  $h=0.3$  a.u.

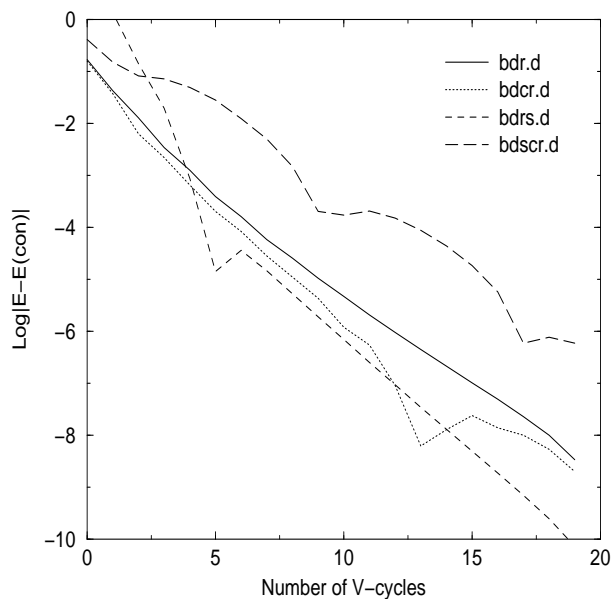


FIG. 6: Convergence rates for benzenedithiol. The  $\log_{10}$  of the difference between the current total energy  $E$  a.u and the fully converged value  $E_{con}$  is plotted against the number of  $v$ -cycles (self-consistent iterations). The plots are for regular Ritz projection (bdr.d), Ritz projection performed on clusters along with GRBR (bdcr.d), simultaneous method (bdrs.d) and for simultaneous method with Ritz projection performed on clusters along with GRBR (bdscr.d). The fine grid spacing is  $h=0.3$  a.u.

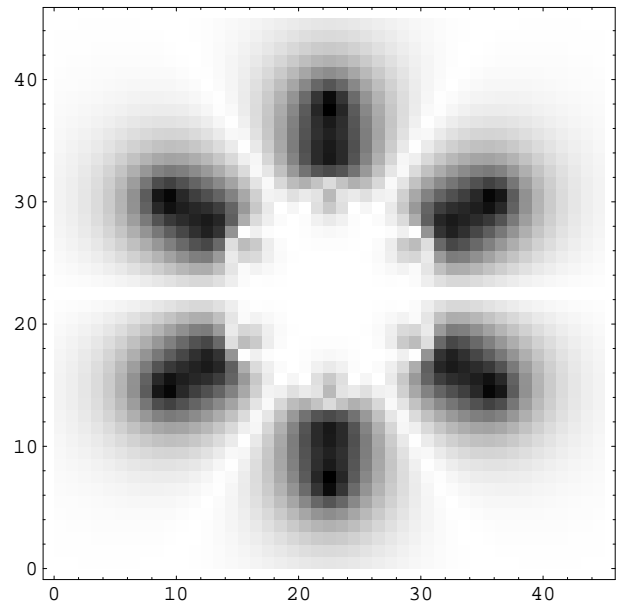


FIG. 7: This figure shows the square root value of the electron density across x-y plane for the 7th highest occupied state of benzene. The blackness is proportional to the square value of the electron density. The benzene molecule lies on the x-y plane at the center of the domain.

is performed on a less number of grid points than on the fine grid and perhaps if any defects (errors) generated by this, they would have been absorbed into the defect correction  $\tau$  and consistent convergence was possible.

We believe that these methods specially the Ritz projection on clusters along with GRBR could be extended to much larger systems having hundreds of wavefunctions. For such systems use of orthogonalization and Ritz projection would be extremely costly since both operations scale as  $q^2 * N$ . But choosing relatively small clusters and performing orthogonalization and Ritz projection on those smaller clusters would save much needed computational time.

## Acknowledgments

We gratefully acknowledge the support of the National Science Foundation (CHE-0112322) for this research. We also thank Achi Brandt and Shlomo Ta'asan for many helpful discussions.

- 
- [1] M. Payne, M. Teter, D. Allan, T. Arias, and J. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- [2] M. Challacombe, *Comput. Phys. Commun.* **128**, 93 (2000).
- [3] D. P. Sánchez-Portal, P. Ordejón, E. Artacho, , and J. M. Soler, *Intl. J. Quantum Chem.* **65**, 453 (1997).
- [4] T. L. Beck, *Rev. Mod. Phys.* **72**, 1041 (2000).
- [5] E. L. Briggs, D. J. Sullivan, and J. Bernholc, *Phys. Rev. B.* **54**, 14362 (1996).
- [6] J. Wang and T. L. Beck, *J. Chem. Phys.* **112**, 9223 (2000).
- [7] N. Wijesekera, G. Feng, and T. L. Beck, *J. of Theor. Comput. Chem.* **2**, 553 (2003).
- [8] M. Heiskanen, T. Torsti, M. J. Puska, and R. M. Nieminen, *Phys. Rev. B.* **63**, 245106 (2001).
- [9] F. Ancilotto, P. Blandin, and F. Toigo, *Phys. Rev. B.* **59**, 7868 (1999).
- [10] T. Hoshi and T. Fujiwara, *J. Phys. Soc. Jpn.* **66**, 3710 (1997).
- [11] I.-H. Lee, Y.-H. Kim, , and R. M. Martin, *Phys. Rev. B.* **61**, 4397 (2000).
- [12] N. A. Modine, G. Zumbach, and E. Kaxiras, *Phys. Rev. B.* **55**, 10289 (1997).
- [13] J.-L. Fattebert, *J. Comput. Phy.* **149**, 75 (1999).
- [14] J.-L. Fattebert and J. Bernholc, *Phys. Rev. B.* **62**, 1713 (2000).
- [15] A. Brandt, S. F. McCormick, , and J. Ruge, *SIAM J. Sci. Stat. Comput.* **4**, 244 (1983).
- [16] C. Hartwigsen, S. Goedecker, and J. Hutter, *Phys. Rev. B.* **58**, 3641 (1998).
- [17] S. Costiner and S. Ta'asan, *Phys. Rev. E.* **51**, 3704 (1995).
- [18] S. Costiner and S. Ta'asan, *Phys. Rev. E.* **52**, 1181 (1995).
- [19] S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B.* **54**, 1703 (1996).
- [20] R. D. King-Smith, M. C. Payne, and J. S. Lin, *Phys. Rev. B.* **44**, 13063 (1991).
- [21] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press - New York, 1989) .