Electronic Structure Methods for Predicting the Properties of Materials: Grids in Space

James R. Chelikowsky*, Yousef Saad**, Serdar: Öğüt* and Igor Vasiliev*

*Department of Chemical Engineering and Material Science,

**Department of Computer Science,

Minnesota Supercomputing Institute,

University of Minnesota,

Minneapolis, Minnesota 55455, USA

Andreas Stathopoulos

Department of Computer Science,

College of William and Mary,

Williamsburg, VA, 23187, USA

(October 29, 1999)

Abstract

If the electronic structure of a given material is known, then many physical and chemical properties can be accurately determined without resorting to experiment. However, determining the electronic structure of a realistic material is a difficult numerical problem. The chief obstacle faced by computational materials and computer scientists is obtaining a highly accurate solution to a complex eigenvalue problem. We illustrate a new numerical method for calculating the electronic structure of materials. The method is based on discretizing the *pseudopotential density functional method* (PDFM) in real space. The eigenvalue problem within this method can involve large,

sparse matrices with up to thousands of eigenvalues required. An efficient and accurate solution depends increasingly on complex data structures that reduce memory and time requirements, and on parallel computing. This approach has many advantages over traditional plane wave solutions, *e.g.*, no fast Fast Fourier transforms (FFT's) are needed and, consequently, the method is easy to implement on parallel platforms. We demonstrate this approach for localized systems such as atomic clusters.

I. INTRODUCTION: THE ELECTRONIC STRUCTURE PROBLEM.

A fundamental problem in condensed matter physics is the prediction of the electronic structure of complex systems such as amorphous solids and glasses or small atomic clusters. Many materials properties can be predicted if an accurate solution of the electronic structure for the system of interest exists. For example, the structural properties of a material can be determined if the total electronic energy of the system is known as a function of atomic positions. Likewise, response functions such as optical and dielectric constants can be determined if the electronic wave functions are known. Beyond the scientific merit of verifying experimental results and establishing the validity of new scientific concepts, these electronic structure calculations also facilitate testing of hypothetical materials without laboratory experiments.

There are numerous approaches to the electronic structure problem. [1] These approaches range from simple empirical methods where experiment is used to fix adjustable parameters to first principles methods where no experimental data is needed. Here we focus on a first principles approach. It is important to recognize the advantages of first principles methods. Such methods avoid *ad hoc* constructs and the prejudice of "preconceived" ideas relative to the nature of the chemical bonds in condensed matter. The construction of efficient first principles methods is among the most challenging tasks in computational materials science today. The heart of the computation problem is to obtain highly accurate values for the

total electronic energy of matter from the solution of a large, Hermitian eigenvalue problem. Part of the challenge stems from the fact that the number of eigenvalues and eigenvectors (*i.e.*, eigenpairs) required can be very large, say in the order of thousands. This number is proportional to the number of atoms in the system which can be in the thousands (or more) for realistic models.

The electronic structure of matter is described by a many body wave function Ψ which obeys the Schrödinger equation:

$$\mathcal{H}\Psi = \mathcal{E}\Psi,$$

where $\mathcal{H}$ is the Hamiltonian operator for the system and $\mathcal{E}$ is the total energy. This expression can be simplified through several approximations. These approximations are all based on the removal of degrees of freedom. For example, the Born-Oppenheimer approximation separates the nuclear degrees freedom and the electronic degrees of freedom. Within this approximation, the nuclear coordinates are treated as classical objects. Another simplification is the utilization of density functional theory [2–4] to map the many body problem on to a one-electron problem. These two approximations yield the following:

$$\left[\frac{-\hbar^2\nabla^2}{2m} + V_{tot}[\rho(\vec{r}), \vec{r}]\right]\psi_i(\vec{r}) = E_i\psi_i(\vec{r}), \tag{1}$$

where $\hbar$ is Planck's constant, m is the electron mass, V_{tot} is the total potential at some point $\vec{r}$ in the system, and $\rho(\vec{r})$ is the charge density at that point. The potential depends explicitly on the charge density, which in turn depends on the wave functions ψ_i as follows:

$$\rho(\vec{r}) = -e \sum_i |\psi_i(\vec{r})|^2 \tag{2}$$

where the sum is over occupied states. The electronic structure problem can be viewed as a nonlinear eigenvalue problem because of the nonlinear dependence of the operator on the left-hand side on the eigenfunctions.

Within the *local density approximation* theory [2], the potential V_{tot} maybe written as a sum of three distinct terms, specifically,

$$V_{tot}(\vec{r}) = V_{ion}(\vec{r}) + V_H(\vec{r}) + V_{xc}[\vec{r}, \rho(\vec{r})], \tag{3}$$

where V_{ion} is the unscreened potential. In the case of an atom, it would correspond to the bare nuclear potential. V_H is the Hartree potential, and V_{xc} is the exchange-correlation potential. Once the charge density $\rho(r)$ is known, the Hartree potential is obtained by solving the Poisson equation:

$$\nabla^2 V_H = -4\pi e\ \rho(r). \tag{4}$$

The exchange-correlation potential depends on the charge density at the point of interest. Both potentials V_H and V_{xc} have a local character. The density functional approximation reduces the number of degrees to those of a "one-electron" problem.

Within the local density approximation, the total potential and the wave functions are interdependent through the charge density. Equations 1, 3, and 4 constitute a set of non-linear equations. These are typically solved by the construction of a self consistent field (SCF). The procedure is usually initiated by superposing atomic charge densities to obtain an approximate charge density for the system of interest. From this density, the "input" Hartree and exchange-correlation potentials are formed. One solves a *Kohn-Sham* eigenvalue problem [2]:

$$\left[\frac{-\hbar^2\nabla^2}{2m} + V^p_{ion}(\vec{r},\rho(\vec{r})) + V_H(\vec{r}) + V_{xc}(\vec{r})\right]\psi_i(r) = E_i\psi_i(r), \tag{5}$$

for the eigenvalues and eigenvectors using the input potentials. With the eigenvalues and eigenvectors determined, we can obtain an "output" charge density. Using the "input" and "output" charge densities new V_H and V_{xc} potentials can be obtained. If the input and output charge densities are identical, then a self-consistent field is obtained. Since the superposition of atomic charge densities are not identical to the charge density in condensed matter phases, the input and output densities are significantly different. The input and output densities are mixed and a new density is formed and input into a new SCF cycle. The resulting V_H and V_{xc} are inserted into Eq. 5 and new eigenvalues and eigenvectors are obtained. This process is repeated until the difference between input and output potentials is below some specified tolerance.

The total electronic structure of the material, E_{tot}^{el}, can be written as

$$E_{tot}^{el} = \sum_i E_i - \frac{1}{2}\int d^3r\ V_H(\vec{r})\ \rho(\vec{r}) + \frac{1}{2}\int d^3r\ [E_{xc}[\rho(\vec{r})] - V_{xc}[\rho(\vec{r})]\ \rho(\vec{r}) \tag{6}$$

The total energy of the system is given by

$$E\{\vec{R}\} = E_{tot}^{el} + E_{ion-ion}\{\vec{R}\} \tag{7}$$

The second term represents the ion-ion interaction, *i.e.*, the Coulombic interaction between the ion-cores whose positions are given by $\{\vec{R}\}$. If we are given, $E\{\vec{R}\}$, any property related to the structure of matter can be calculated, at least in principle.

II. SOLVING THE EIGENVALUE PROBLEM

A major difficulty in solving the eigenvalue problem in Eq. 5 are the length and energy scales involved. The inner (core) electrons are highly localized and tightly bound compared to the outer (valence electrons). A simple basis function approach is frequently ineffectual. For example, a plane wave basis might require 10^{5-6} waves to represent converged wave functions for a core electron whereas only 10^2 waves are required for a valence electron. The pseudopotential overcomes this problem by removing the core states from the problem and replacing the all electron potential by one that replicates only the chemically active, valence electron states [5]. By construction, the pseudopotential reproduces the valence state properties such as the eigenvalue spectrum and the charge density outside the ion core. The unscreened pseudopotential, $V_{ion}^{p}(\vec{r})$ replaces $V_{ion}(\vec{r})$ in Eq. 5.

Since the pseudopotential is weak, simple basis sets such as a plane wave basis are extremely effective. For example, in the case of crystalline silicon only 50-100 plane waves need to be used. The resulting matrix representation of the Schrödinger operator is dense on the Fourier (plane wave) space, but it is not formed explicitly, Instead, matrix-vector product operations are performed with the help of fast Fourier transforms (FFT). This approach is akin to spectral techniques used in solving certain types of partial differential equations. The plane wave method uses a basis of the form:

$$\psi_{\vec{k}}(\vec{r}) = \sum_{\vec{G}} \alpha(\vec{k}, \vec{G}) \exp(i(\vec{k} + \vec{G}) \cdot \vec{r}) \tag{8}$$

where $\vec{k}$ is the wave vector, $\vec{G}$ is a reciprocal lattice vector and $\alpha(\vec{k}, \vec{G})$ represent the coefficients of the basis. In a plane wave basis, the Laplacian term of the Hamiltonian is represented by a diagonal matrix. The potential term V_{tot}^{p} gives rise to a dense matrix. In practice, these matrices are never formed explicitly, since with appropriate use of FFT we can easily operate with this matrix by going back and forth between real-space and Fourier space. Indeed, in real space it is trivial to operate with the potential term which is represented by a diagonal matrix, and in Fourier space it is trivial to operate with the Laplacian term which is also represented by a diagonal matrix. The use of plane wave bases also leads to natural preconditioning techniques which are obtained by simply employing a matrix obtained from a smaller plane wave basis, neglecting the effect of high frequency terms on the potential. For periodic systems, where $\vec{k}$ is a good quantum number, the plane wave basis coupled to pseudopotentials is quite effective. However, for non-periodic systems such as clusters, liquids or glasses, the plane wave basis must be combined with a *supercell method* [5]. The supercell repeats the localized configuration to impose periodicity to the system. There is also again a parallel to be made with spectral methods which are quite effective for simple periodic geometries, but lose their superiority when more generality is required. In addition to these difficulties the two FFTs performed at each iteration can be costly, requiring $n \log n$ operations, where n is the number of plane waves, versus $O(N)$ for real space methods where N is the number of grid points. Usually, the matrix size $N \times N$ is larger than $n \times n$ but only within a constant factor. This is exacerbated in high performance environments where FFTs require an excessive amount of communication and are particularly difficult to implement efficiently.

Another popular basis employed with pseudopotentials include Gaussian orbitals [6]. Gaussian bases have the advantage of yielding analytical matrix elements provided the potentials are also expanded in Gaussians. However, the implementation of a Gaussian basis is not as straightforward as with plane waves. For example, numerous indices must be

employed to label the state, the atomic site, and the Gaussian orbitals employed. On the positive side, a Gaussian basis yields much smaller matrices and requires less memory than plane wave methods. For this resason Gaussians are especially useful for describing transition metal systems.

An alternative approach is to avoid the use of a basis. For example, one can use a real space method that avoids the use of plane waves and FFT's altogether. This approach has become popular and different versions of this general approach been implemented by several groups. Here we illustrate a particular version of this approach called the *Finite-Difference Pseudopotential Method* (FDPM) [7].

A real space approach overcomes some of the complications involved with non-periodic systems, and although the resulting matrices can be larger than with plane waves, they are sparse and the methods are easier to parallelize. Even on sequential machines, we find that real space methods can be an order of magnitude faster than the traditional approach.

Our *real space* algorithms avoid the use of FFT's by performing all calculations in real physical space instead of Fourier space. A benefit of avoiding FFT's is that the new approaches have very few global communications. In fact, the only global operation remaining in *real space* approaches is that of the inner products. These inner products are required when forming the orthogonal basis used in the generalized Davidson procedure as discussed below.

Our approach utilizes finite difference discretization on a real space grid. A key aspect to the success of the finite difference method is the availability of *higher order finite difference expansions* for the kinetic energy operator, *i.e.*, expansions of the Laplacian [10]. Higher order finite difference methods significantly improve convergence of the eigenvalue problem when compared with standard finite difference methods. If one imposes a simple, uniform grid on our system where the points are described in a finite domain by (x_i, y_j, z_k), we approximate $\frac{\partial^2 \psi}{\partial x^2}$ at (x_i, y_j, z_k) by

$$\frac{\partial^2 \psi}{\partial x^2} = \sum_{n=-M}^{M} C_n \psi(x_i + nh, y_j, z_k) + O(h^{2M+2}), \tag{9}$$

where h is the grid spacing and M is a positive integer. This approximation is accurate to $O(h^{2M+2})$ upon the assumption that ψ can be approximated accurately by a power series in h. Algorithms are available to compute the coefficients C_n for arbitrary order in h [10].

With the kinetic energy operator expanded as in Eq. 9, one can set up a one-electron Schrödinger equation over a grid. One may assume a uniform grid, but this is not a necessary requirement. $\psi(x_i, y_j, z_k)$ is computed on the grid by solving the eigenvalue problem:

$$-\frac{\hbar^2}{2m}\left[\sum_{n_1=-M}^{M} C_{n_1}\psi_n(x_i+n_1h, y_j, z_k) + \sum_{n_2=-M}^{M} C_{n_2}\psi_n(x_i, y_j+n_2h, z_k) + \sum_{n_3=-M}^{M} C_{n_3}\psi_n(x_i, y_j, z_k+n_3h)\right] + [\ V_{ion}(x_i, y_j, z_k) + V_H(x_i, y_j, z_k) + V_{xc}(x_i, y_j, z_k)\ ]\,\psi_n(x_i, y_j, z_k) = E_n\ \psi_n(x_i, y_j, z_k). \tag{10}$$

If we have L grid points, the size of the full matrix resulting from the above problem is $L \times L$.

A complicating issue in setting up an algorithm is the ionic pseudopotential term. This term is easy to cast in Fourier space, but it may also be expressed in real space. The interactions between valence electrons and pseudo-ionic cores may be separated into a local potential and a Kleinman and Bylander [8] form of a nonlocal pseudopotential in *real space* [9],

$$V_{ion}(\vec{r})\psi_n(\vec{r}) = \sum_a V_{loc}(|\vec{r}_a|)\psi_n(\vec{r}) + \sum_{a,\ n,lm} G^a_{n,lm} u_{lm}(\vec{r}_a)\Delta V_l(r_a) \tag{11}$$

$$K^a_{n,lm} = \frac{1}{< \Delta V^a_{lm} >} \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)\psi_n(\vec{r}) d^3r \tag{12}$$

and $< \Delta V^a_{lm} >$ is the normalization factor,

$$< \Delta V^a_{lm} > = \int u_{lm}(\vec{r}_a)\Delta V_l(r_a) u_{lm}(\vec{r}_a) d^3r, \tag{13}$$

where $\vec{r}_a = \vec{r} - \vec{R}_a$, and the u_{lm} are the atomic pseudopotential wave functions of angular momentum quantum numbers (l, m) from which the l-dependent ionic pseudopotential, $V_l(r)$, is generated. $\Delta V_l(r) = V_l(r) - V_{loc}(r)$ is the difference between the l component of the ionic pseudopotential and the local ionic potential.

The grid we use is based on a points uniformly spaced in a three dimensional cube as shown in Figure 1, with each grid point corresponding to a row in the matrix. However, many points in the cube are far from any atoms in the system and their negligible charge may then be replaced by zero. Special data structures may be used to discard these points and keep only those having a nonzero charge. The size of the Hamiltonian matrix is usually reduced by a factor of two to three with this strategy, which is quite important considering the large number of eigenvectors which must be saved. Further, since the Laplacian can be represented by a simple stencil, and since all local potentials sum up to a simple diagonal matrix, the Hamiltonian need not be stored. Handling the ionic pseudopotential is complex as it consists of a local and a non-local term (Eqs. 11 and 12). In the discrete form, the nonlocal term becomes a sum over all atoms, a, and quantum numbers, (l, m) of rank-one updates:

$$V_{ion} = \sum_{a} V_{loc,a} + \sum_{a,l,m} c_{a,l,m} U_{a,l,m} U_{a,l,m}^{T} \tag{14}$$

where $U_{a,l,m}$ are sparse vectors which are only non-zero in a localized region around each atom, $c_{a,l,m}$ are normalization coefficients.

There are several difficulties with the eigen problems generated in this application in addition to the size of the matrices. First, the number of required eigenvectors is proportional to the atoms in the system, and can grow up to thousands. Besides storage, maintaining the orthogonality of these vectors can be a formidable task. Second, the relative separation of the eigenvalues becomes increasingly poor as the matrix size increases and this has an adverse effect on the rate of convergence of the eigenvalue solvers. Preconditioning techniques attempt to alleviate this problem.

On the positive side, the matrix need not be stored as was mentioned earlier and this reduces storage requirement. In addition, good initial eigenvector estimates are available at each iteration from the previous SCF loop. An iterative method should be able to use this information.

In this work, we developed a code based on the generalized Davidson [11] method, in

which the preconditioner is not restricted to be a diagonal matrix as in the Davidson method. The code addresses the problems mentioned above by using implicit deflation (locking), a windowing approach to gradually compute all the required eigenpairs, and special targeting and reorthogonalization schemes. A more detailed description can be found in [15].

A preconditioning technique we used in our approach is based on a filtering idea and the fact that the Laplacian is an elliptic operator [13]. The eigenvectors corresponding to the few lowest eigenvalues of ∇^2 are smooth functions and so are the corresponding wavefunctions. When an approximate eigenvector is known at the points of the grid, a smoother eigenvector can be obtained by averaging the value at every point with the values of its neighboring points. Assuming a cartesian (x, y, z) coordinate system, the low frequency filter acting on the value at the point (i, j, k), which represents one element of the eigenvector, is described by:

$$\psi_{i,j,k} := \frac{\psi_{i-1,j,k} + \psi_{i,j-1,k} + \psi_{i,j,k-1} + \psi_{i+1,j,k} + \psi_{i,j+1,k} + \psi_{i,j,k+1}}{12} + \frac{\psi_{i,j,k}}{2} \tag{15}$$

It is worth mentioning that other preconditioners that have been tried have resulted in mixed success. The use of shift-and-invert [14] involves solving linear systems with $A - \sigma I$, where A is the original matrix and the shift σ is close to the desired eigenvalue. These methods would be prohibitively expensive in our situation, given the size of the matrix and the number of times that $A - \sigma I$ must be factored. Alternatives based on an approximate factorization such as ILUT [15] are ineffective beyond the first few eigenvalues. Methods based on approximate inverse techniques have been somewhat more successful, performing better than filtering at additional preprocessing and storage cost. Preconditioning 'interior' eigenvalues, i.e., eigenvalues located well inside the interval containing the spectrum, is still a very hard problem. Current solutions only attempt to dampen the effect of eigenvalues which are far away from the ones being computed. This is in effect what is achieved by filtering and sparse approximate inverse preconditioning. These techniques do not reduce the number of steps required for convergence in the same way that shift-and-invert techniques do. However, filtering techniques are inexpensive to apply and result in fairly substantial

savings in iterations.

III. PARALLEL IMPLEMENTATION

For distributed memory parallel computers, the SPMD (Single Program Multiple Data) model has emerged as the most popular programming paradigm. In our implementation of the SCF procedure we have followed a hybrid of the SPMD and the master-worker paradigm. The master performs most of the preprocessing, computing of scalar values, and processing of the new potential at each SCF iteration. The master is also responsible for applying the mixing scheme on the potentials. The workers solve the eigenvalues and eigenvectors, update the charge density, and solve the Poisson equation for the Hartree potential in an SPMD fashion.

There are several reasons dictating the master-worker choice. First, there are some inherently sequential parts in the code which require large memory but short execution time. It is also common that one of the nodes in a parallel environment is equipped with larger memory than the others. Second, the code calls several library routines which have been written by various research groups over a long period of time. Despite their importance, these routines take only a few seconds to execute. Parallelizing them all would require an inordinate amount of effort with doubtful results as to the achievable gains. Third, this paradigm allows incremental parallelization of the code, implementing first the most time consuming procedures, such as the eigensolver, then gradually adding parallelism to other parts. Correctness of the code is also easier to maintain by this strategy. Finally, the resulting code is portable to other parallel platforms without requiring large amounts of memory for all the worker processors.

The primary sources of parallelism intrinsically available in the application are: (1) the multitude of required eigenvectors, and (2) parallelism from spatial decomposition. Assigning each processor the task of calculating all the eigenpairs in a segment of the spectrum would provide excellent coarse grain parallelism and parallel efficiency. For each eigenpair,

one could use inverse iteration with some iterative method. However, the linear systems to be solved are highly indefinite and iterative methods for the inverse iteration converge extremely slowly. As was mentioned earlier shift-and-invert is impractical for the large matrices at hand. An alternative is to use a polynomial preconditioning approach. A polynomial p can be found such that the dominant eigenvalues of $p(A)$ are the transforms by p of the eigenvalues in the desired subinterval. Then these dominant eigenvalues and associated eigenvectors can be computed and the corresponding eigenvalues of the original matrix can then be evaluated. A major advantage with this approach is that global orthogonality does not need to be maintained since the eigenvectors of a Hermitian matrix are orthogonal if they are computed accurately enough; only eigenvectors associated with eigenvalues in a given subinterval must be orthogonalized during the computation. This is a workable approach but the book-keeping required in order to ensure that no eigenvalues are missed and that they are all represented only once may be quite cumbersome. In addition high degree polynomials may be needed that reduce the gains from parallelism.

Instead of this 'spectral decoupling' idea, we have adopted a domain decomposition approach based on partitioning the physical space. The problem is mapped onto the processors in a data parallel way because of the fine granularity parallelism present in the matrix-vector multiplication and orthogonalization operations. The rows of the Hamiltonian (and therefore the rows of the eigenvectors and potential vectors) are assigned to processors according to a partitioning of the physical domain. The subdomains can be chosen naturally as sub-cubes or slabs of the cube, but since the zero-charge areas can be arbitrarily distributed in the domain, a general partitioning is more appropriate. This is illustrated in Figure 2 We have designed the mapping routines to be independent of the partitioner, requiring only a function $P(i, j, k)$ which returns the number of the processor where point (i, j, k) resides. This facilitates the use of many publicly available partitioning tools. We have tested two ways of partitioning. The first is a greedy approach that optimizes load balancing by ordering the points and assigning the same number of points to each processor, but it ignores the amount of communication which is induced. The second approach uses the popular partitioning

package METIS [16] which seeks to optimize both load balancing and the communication volume between processors.

Since the matrix is not actually stored, an explicit reordering can be considered so that the rows on a processor are numbered consecutively. Under this conceptually easier scheme, only a list of pointers is needed that denote where the rows of each processor start. The nonlocal part of the matrix, which is a sum of rank-one updates, is mapped in a similar way. For each atom and for each pair of quantum numbers, a sparse vector $U_{a,l,m}$ in Eq. 14 is partitioned according to the rows it contributes to. Even though the number of non-zero elements of the U-vectors is small, their partitioning is fairly well balanced if the matrix partitioning is well balanced. With this mapping, the large storage requirements of the program are distributed.

The tools that we developed for mapping, setting up the data structures and performing the communication, are independent of the nature of our problem and can be embedded in other applications for unstructured stencil computations, which use any of the general data structures described in the following sections.

In the Davidson algorithm, the basis vectors and long work arrays, follow the same distribution as the eigenvectors. Thus, all vector updates (saxpy operations) can be performed in parallel, and all reduction operations (e.g., sdot operations) require a global reduction (e.g., global sum) of the partial results on each processor.

The matrix-vector multiplication is performed in three steps. First, the contributions of the diagonals (potentials and the Laplacian diagonal) is computed in parallel on all processors. Second, the contribution of the Laplacian is considered on the rows of each processor. As in the sequential code, this is performed by using the stencil information. In the parallel implementation communication is necessary, since some of the neighboring points of the local subdomain may reside on different processors. For this reason, each processor maintains the following data structure, which maps the local grid points to the local rows, and appends the needed interface points from other processors at the end of the local row list:

$$index(i,j,k) = \begin{cases} \text{row number in local ordering,} & \text{if } (i,j,k) \text{ is on local processor} \\ \text{index below local ordering,} & \text{if } (i,j,k) \text{ is a needed interface row} \\ \text{special index,} & \text{if } (i,j,k) \text{ is not considered (zero charge)} \end{cases} \tag{16}$$

The workers build this and other supporting data structures during the setup phase, by locating which of their rows are needed in the stencils of other processors. In the second step of the matrix-vector multiplication, this interface information is exchanged among nearest neighbors and the stencil multiplication can proceed in parallel. In the third step, each of the rank-one updates of the nonlocal components is computed as a sparse, distributed dot product. All local dot products are first computed before a global sum of their values takes place. The solution of Eq. 4 for the Hartree potential with the Conjugate Gradient method and the preconditioning operation also require the stencil, and therefore, they have the same communication pattern as the second step of the matrix-vector multiplication.

Orthogonalization is an expensive phase, and as the number of required eigenvectors increases, it is bound to dominate the cost. Reorthogonalization is performed every time a vector norm reduces significantly after orthogonalization. Although reorthogonalization recovers the numerical accuracy lost in the Gram-Schmidt procedure, its nature is sequential and induces several synchronization points. In the current application, global sums of the dot-products are delayed so that only one synchronization is needed. In addition, by performing the reorthogonalization test through easily obtained estimates of the vector norms, we introduce only two synchronization points in the procedure.

To demonstrate scalability of the code, we examined a large quantum dot involving 191 silicon atoms and 148 hydrogen atoms. The matrix size involved 83,200 grid points, *i.e.*, in principle the Hamiltonian matrix contain 83,200 $\times$ 83,200 entries. For the electronic structure calculation, 560 eigenvalues were obtained. The overall scalability is illustrated in Figure 3. Clusters or quantum dots present a difficult problem as the environment for each atom can be very different, *e.g.*, a surface atom has far fewer neighbors than does an interior atom. Owing to this issue, our scale up efficiency of $\sim$80% is quite good.

We note that there are several groups utilizing this real space approach [17]. There are some notable differences between the current approaches and these approaches. A non-uniform grid is often incorporated. Non-uniform grids can be used to accommodate systems with highly heterogeneous environments. As an example, consider a system with two atomic constituents: one with highly localized wave functions and another with delocalized wave functions. For a uniform grid, the spacing will be fixed by the highly localized species and, consequently, will be "over converged" for the delocalized species. Non-uniform grids can be adapted so that regions with rapidly fluctuating wave functions are represented by a fine grid and regions with slowly fluctuating wave functions are represented by a coarse grid. This advantage can be considerable for some systems, but the real space approach loses its ease of implementation in this case. For example, the Hamiltonian matrix loses its highly structured form and expressions for the interatomic forces become quite complex. Also, implementing and optimizing a uniform grid, especially in systems where the atoms are allowed to move, can significantly increase the computational load. To date, no molecular dynamics simulations have been performed with non-uniform grids for this reason.

An alternative approach to the finite difference code is to use a *finite elements* method. The finite element approach shares some of the advantages associated with methods based on non-uniform grids. Finite elements can be adapted to enhance convergence over specific regimes in real space. In addition, finite element approaches are variational, since they correspond a basis oriented approach. (In contrast, finite difference methods are not variational with the grid spacing and the total energy can converge from above or below.) However, finite element methods also share the disadvantages of non-uniform grids, *i.e.*, they are difficult to implement, and more computationally intensive.

IV. PROPERTIES OF CONFINED SYSTEMS: CLUSTERS

The electronic and structural properties of atomic clusters stands as one of the outstanding problems in materials science. Clusters possess properties that are characteristic of

neither the atomic nor solid state. For example, the energy levels in atoms may be discrete and well separated in energy relative to kT. In contrast, solids have continuum of states (energy bands). Clusters may reside between these limits, *i.e.*, the energy levels may be discrete, but with a separation much less than kT.

Real space methods are ideally suited for investigating these systems. In contrast to plane wave methods, real space methods can examine non-periodic without introducing artifacts such as supercells. Also, one can easily examine charged clusters. In supercell configurations, unless a compensating background charge is added, the Coulomb energy diverges for charged clusters. A closely related issue concerns electronic excitations. In periodic systems, it is nontrivial to consider localized excitation, *e.g.*, exciting an atom in one cell, excites all atoms. Density functional formalisms avoid these issues by considering localized or non-periodic systems.

A. Structure

Perhaps the most fundamental issue in dealing with clusters is the structure. Before any accurate theoretical calculations can be performed for a cluster, the atomic geometry must be known. However, determining the atomic structure of clusters can be a formidable exercise. Serious problems arise from the existence of multiple local minima in the potential-energy-surface of these systems. This is especially true for some clusters such as those involving semiconducting species. In these clusters, strong many body forces can exist.

A convenient method to determine the structure of small clusters is *simulated annealing*. Within this technique, atoms are randomly placed within a large cell and allowed to interact at a high (usually fictive) temperature. The atoms will sample a large number of configurations. As the system is cooled, the number of high energy configurations sampled is restricted. If the anneal is done slowly enough, the procedure should quench out structural candidates for the ground state structures.

Langevin molecular dynamics appears well suited for such simulated anneals. In Langevin dynamics, the ionic positions, $\mathbf{R}_j$, evolve according to

$$M_j \ \ddot{\mathbf{R}}_j = \ \mathbf{F}(\{\mathbf{R}_j\}) - \gamma M_j \ \dot{\mathbf{R}}_j + \mathbf{G}_j \tag{17}$$

where $\mathbf{F}(\{\mathbf{R}_j\})$ is the interatomic force on the j-th particle, and $\{\mathrm{M}_j\}$ are the ionic masses. The last two terms on the right hand side of Eq. (17) are the dissipation and fluctuation forces, respectively. The dissipative forces are defined by the friction coefficient, γ. The fluctuation forces are defined by random Gaussian variables, $\{\mathbf{G}_i\}$, with a white noise spectrum:

$$\langle G_i^\alpha(t)\rangle = 0 \quad \text{and} \quad \langle G_i^\alpha(t) G_j^\alpha(t')\rangle = 2\gamma \ M_i \ k_B \ T \ \delta_{ij} \ \delta(t-t') \tag{18}$$

The angular brackets denote ensemble or time averages, and α stands for the Cartesian component. The coefficient of T on the right hand side of Eq. (18) insures that the fluctuation-dissipation theorem is obeyed, *i.e.*, the work done on the system is dissipated by the viscous medium ([18,19]). The interatomic forces can be obtained from the Hellmann-Feynman theorem using the pseudopotential wave functions.

Our simulations can be contrasted with other techniques such as the Car-Parrinello method. We do not employ fictitious electron dynamics; at each time step the system is quenched to the Born-Oppenheimer surface. Our approach requires a full-self consistent treatment of the electronic structure problem; however, because the interatomic forces are true, quantum forces the resulting molecular dynamics simulation can be performed with much larger time steps. Typically, it is possible to use steps an order of magnitude larger than in the Car-Parrinello method.

To illustrate the procedure, we consider a germanium cluster of seven atoms. With respect to the technical details for this example, the initial temperature of the simulation was taken to be 2800 K; the final temperature was taken to be 300 K. The annealing schedule lowered the temperature 500 K each 50 time steps. The time step was taken to be 7 fs. The friction coefficient in the Langevin equation was taken to be 6×10^{-4} a.u. After the clusters

reached a temperature of 300 K, they were quenched to 0 K. The ground state structure was found through a direct minimization by a steepest descent procedure.

Choosing an initial atomic configuration for the simulation takes some care. If the atoms are too far apart, they will exhibit Brownian motion and may not form a stable cluster as the simulation proceeds. If the atoms are too close together, they may form a metastable cluster from which the ground state may be kinetically inaccessible even at the initial high temperature. Often the initial cluster is formed by a random placement of the atoms with a constraint that any given atom must reside within 1.05 and 1.3 times the dimer bond length of at least one atom. The cluster in question is placed in a spherical domain. Outside of this domain, the wave function is required to vanish. The radius of the sphere is such that the outmost atom is at least 6 a.u. from the boundary. Initially, the grid spacing was 0.8 a.u. For the final quench to a ground state structure, the grid spacing was reduced to 0.5 a.u. As a rough estimate, one can compare this grid spacing with a plane wave cutoff of $(\pi/h)^2$ or about 40 Ry for h=0.5 a.u.

In Figure 4, we illustrate the simulated anneal for this Ge_7 cluster. While the initial cluster contains several of bonds, the structure is still somewhat removed from the ground state. After $\sim$200 time steps, the ground state structure is essentially formed. The ground state of Ge_7 is a bicapped pentagon, as is the corresponding structure for the Si_7 cluster. The binding energy shown is relative to the isolated Ge atom. We have not included gradient corrections, or spin polarization [20] in our work. Therefore, the values indicated are likely to overestimate the binding energies by $\sim$ 20% or so.

In Figure 5, we present the ground state structures for Ge_n for $n \leq 10$. The structures for Ge_n are very similar to Si_n. The primary difference resides in the bond lengths. The Si bond length in the crystal is 2.35 Å, whereas in Ge the bond length is 2.44 Å. This difference is reflected in the bond lengths for the corresponding clusters. Ge_n bond lengths are typically a few percent larger than the corresponding Si_n clusters.

It should be emphasized that this annealing simulation is an optimization procedure. As such, other optimization procedures may be used to extract the minimum energy struc-

tures. Recently, a genetic algorithm has been used to examine carbon clusters [21]. In this algorithm, an initial set of clusters is "mated" with the lowest energy offspring "surviving". By examining several thousand generations, it is possible to extract a reasonable structure for the ground state. The genetic algorithm has some advantages over a simulated anneal, especially for clusters which contain more than ∼20 atoms. One of these advantages is that kinetic barriers are more easily overcome. However, the implementation of the genetic algorithm is more involved than an annealing simulation, *e.g.*, in some cases "mutations," or *ad hoc* structural rearrangements, must be introduced to obtain the correct ground state.

B. Photoemission Spectra

A very useful probe of condensed matter involves the photoemission process. Incident photons are used to eject electrons from a solid. If the energy and spatial distributions of the electrons are known, then information can be obtained about the electronic structure of the materials of interest. For crystalline matter, the photoemission spectra can be related to the electronic density of states. For confined systems, the interpretation is not as straightforward. One of the earliest experiments performed to examine the electronic structures of small semiconductor clusters examined negatively charged Si_n and Ge_n ($n \leq 12$) clusters [22]. The photoemission spectra obtained in this work were used to gauge the energy gap between the highest occupied state and the lowest unoccupied state. Large gaps were assigned to the "magic number" clusters, while other clusters appeared to have vanishing gaps. Unfortunately, the first theoretical estimates [23] for these gaps showed substantial disagreements with the measured values. It was proposed by [22], that sophisticated calculations including transition cross sections and final states were necessary to identify the cluster geometry from the photoemission data. The data were first interpreted in terms of the gaps obtained for *neutral* clusters; it was later demonstrated that *atomic relaxations* within the *charged* cluster are important in analyzing the photoemission data [24]. In particular, atomic relaxations as a result of charging may change dramatically the electronic

spectra of certain clusters. These charge induced changes in the gap were found to yield very good agreement with the experiment.

The photoemission spectrum of Ge_{10}^- illustrates some of the key issues. Unlike Si_{10}^- , the experimental spectra for Ge_{10}^- does not exhibit a gap. Cheshnovsky *et al.* interpreted this to mean that Ge_{10}^- does not exist in the same structure as Si_{10}^-. This is a strange result. Si and Ge are chemically similar and the calculated structures for both *neutral* structures are similar. The lowest energy structure for both ten atom clusters is the tetracapped trigonal prism (labeled by I in Figure 5). The photoemission spectra for these clusters can be simulated by using Langevin dynamics. Within the Langevin framework, the clusters are immersed in a fictive heat bath, and as such, subjected to stochastic forces. If one maintains the temperature of the heat bath and averages over the eigenvalue spectra, a density of states for the cluster can be obtained. The heat bath resembles a buffer gas as in the experimental setup, but the time intervals for collisions are not similar to the true collision processes in the atomic beam. The simulated photoemission spectrum for Si_{10}^- is in very good agreement with the experimental results, reproducing both the threshold peak and other features in the spectrum. If a simulation is repeated for Ge_{10}^- using the tetracapped trigonal prism structure, the resulting photoemission spectrum is *not* in good agreement with experiment. Moreover, the calculated electron affinity is 2.0 eV in contrast to the experimental value of 2.6 eV. However, there is no reason to believe that the tetracapped trigonal prism structure is correct for Ge_{10} when charged. In fact, we find that the bicapped antiprism structure is lower in energy for Ge_{10}^- . The resulting spectra using both structures (I and II in Figure 5) are presented in Figure 6, and compare to the photoemission experiment. The calculated spectrum using the bicapped antiprism structure is in very good agreement with the photoemission. The presence of a gap is indicated by a small peak removed from the density of states [Figure 6(a)]. This feature is absent in the bicapped antiprism structure [Figure 6(b)] and consistent with experiment. For Ge_{10}, charging the structure reverses the relative stability of the two structures. This accounts for the major differences between the photoemission spectra.

C. Vibrational Modes

Experiments on the vibrational spectra of clusters can provide us with very important information about their physical properties. Recently, Raman experiments have been performed on clusters which have been deposited on inert substrates [25]. Since different structural configurations of a given cluster can possess different vibrational spectra, it is possible to compare the vibrational modes calculated for a particular structure with experiment. If the agreement between experiment and theory is good, this is a necessary condition for the validity of the theoretically predicted structure.

There are two common approaches for determining the vibrational spectra of clusters. One approach is to calculate the dynamical matrix for the ground state structure of the cluster:

$$M_{i\alpha,j\beta} = \frac{1}{m}\frac{\partial^2 E}{\partial R_i^\alpha \partial R_j^\alpha} = -\frac{1}{m}\frac{\partial F_i^\alpha}{\partial R_j^\alpha} \tag{19}$$

where m is the mass of the atom, E is the total energy of the system, F_i^α is the force on atom i in the direction α, R_i^α is the α component of coordinate for atom i. One can calculate the dynamical matrix elements by calculating the first order derivative of force versus atom displacement numerically. From the eigenvalues and eigenmodes of the dynamical matrix, one can obtain the vibrational frequencies and modes for the cluster of interest [26].

The other approach to determine the vibrational modes is to perform a molecular dynamics simulation. The cluster in question is excited by small random displacements. By recording the kinetic (or binding) energy of the cluster as a function of the simulation time, it is possible to extract the power spectrum of the cluster and determine the vibrational modes. This approach has an advantage for large clusters in that one never has to do a mode analysis explicitly. Another advantage is that anaharmonic modes and mode coupling can be examined. It has the disadvantage in that the simulation must be performed over a long time to extract all the modes.

As a specific example, consider the vibrational modes for a small silicon cluster: Si_4. The

starting geometry was taken to be a planar structure for this cluster as established from a higher order finite difference calculation [26].

It is straightforward to determine the dynamical matrix and eigenmodes for this cluster. In Figure 7, the fundamental vibrational modes are illustrated. In Table I, the frequency of these modes are presented. One can also determine the modes via a simulation. To initiate the simulation, one can perform a Langevin simulation [24] with a fixed temperature at 300K. After a few dozen time steps, the Langevin simulation is turned off, and the simulation proceeds following Newtonian dynamics with "quantum" forces. This procedure allows a stochastic element to be introduced and establish initial conditions for the simulation without bias toward a particular mode. For this example, time step in the MD simulation was taken to be 3.7 fs, or approximately 150 a.u. The simulation was allowed to proceed for 1000 time steps or roughly 4 ps. The variation of the kinetic and binding energies is given in Figure 8 as a function of the simulation time. Although some fluctuations of the total energy occurs, these fluctuations are relatively small, *i.e.*, less than $\sim$ 1 meV, and there is no noticeable drift of the total energy. Such fluctuations arise, in part, because of discretization errors. As the grid size is reduced, such errors are minimized [26]. Similar errors can occur in plane wave descriptions using supercells, *i.e.*, the artificial periodicity of the supercell can introduce erroneous forces on the cluster. By taking the power spectrum of either the KE or BE over this simulation time, the vibrational modes can be determined. These modes can be identified with the observed peaks in the power spectrum as illustrated in Figure 9.

A comparison of the calculated vibrational modes from the MD simulation and from a dynamical matrix calculation are listed in Table 2. Overall, the agreement between the two simulations and the dynamical matrix analysis is quite satisfactory. In particular, the softest mode, *i.e.*, the B_{3u} mode, and the splitting between the (A_g, B_{1u}) modes are well replicated in the power spectrum. The splitting of the (A_g, B_{1u}) modes is less than 10 cm^{-1}, or about 1 meV, which is probably at the resolution limit of any *ab initio* method.

The theoretical values are also compared to experiment. The predicted frequencies for the two A_g modes are surprisingly close to Raman experiments on silicon clusters [25]. The

other allowed Raman line of mode B_{3g} is expected to have a lower intensity and has not been observed experimentally.

The theoretical modes using the formalism outlined here are in good accord (except the lowest mode) with other theoretical calculations given in Table I: an LCAO calculation [27] and a Hartree-Fock (HF) calculation [28]. The calculated frequency of the lowest mode, *i.e.*, the B_{3u} mode, is problematic. The general agreement of the B_{3u} mode as calculated by the simulation and from the dynamical matrix is reassuring. Moreover, the real space calculations agree with the HF value to within $\sim$ 20-30 cm^{-1}. On the other hand, the LCAO method yields a value which is $50-70\%$ smaller than either the real space or HF calculations. The origin of this difference is not apparent. For a poorly converged basis, vibrational frequencies are often overestimated as opposed to the LCAO result which underestimates the value, at least when compared to other theoretical techniques. Setting aside the issue of the B_{3u} mode, the agreement between the measured Raman modes and theory for Si_4 suggests that Raman spectroscopy can provide a key test for the structures predicted by theory.

D. Polarizabilities

Recently polarizability measurements [29] have been performed for small semiconductor clusters. These measurements allow us to compare our computed values with experiment.

The polarizability tensor, α_{ij}, is defined as the second derivative of the energy with respect to electric field components. For a noninteracting quantum mechanical system, the expression for the polarizability can be easily obtained by using second order perturbation theory where the external electric field, $\mathcal{E}$, is treated as a weak perturbation.

Within the density functional theory, since the total energy is not the sum of individual eigenvalues, the calculation of polarizability becomes a nontrivial task. One approach is to use density functional perturbation theory which has been developed recently in Green's function and variational formulations [30,31].

Another approach, which is very convenient for handling the problem for *confined* systems, like clusters, is to solve the full problem exactly within the one electron approximation. In this approach, the external ionic potential $V_{\rm ion}(\mathbf{r})$ experienced by the electrons is modified to have an additional term given by $-e\mathcal{E}\cdot\mathbf{r}$. The Kohn-Sham equations are solved with the full external potential $V_{\rm ion}(\mathbf{r})-e\mathcal{E}\cdot\mathbf{r}$. For quantities like polarizability, which are derivatives of the total energy, one can compute the energy at a few field values, and differentiate numerically. Real space methods are very suitable for such calculations on confined systems, since the position operator $\mathbf{r}$ is not ill-defined, as is the case for supercell geometries in plane wave calculations.

In Table II, we present some recent calculations for the polarizability of small Si and Ge clusters. (This procedure has recently been extended to heteropolar clusters such as $\mathrm{Ga}_m\mathrm{As}_n$, see [32]) It is interesting to note that some of these clusters have permanent dipoles. For example, Si_6 and Ge_6 both have nearly degenerate isomers. One of these isomers possesses a permanent dipole, the other does not. Hence, in principle, one might be able to separate the one isomer from the other via an inhomogeneous electric field.

E. Optical Spectra

While the theoretical background for calculating ground state properties of many-electron systems is now well established, excited state properties such as optical spectra present a challenge for computational methods. Recently developed linear response theory within the time-dependent density-functional formalism provides a new tool for calculating excited states properties [33]. This method, known as the time-dependent LDA (TDLDA), allows one to compute the true excitation energies from the conventional, time independent Kohn-Sham transition energies and wavefunctions.

Within the TDLDA, the electronic transition energies Ω_n are obtained from the solution of the following eigenvalue problem: [33]

$$\left[\omega_{ij\sigma}^2\delta_{ik}\delta_{jl}\delta_{\sigma\tau}+2\sqrt{f_{ij\sigma}\omega_{ij\sigma}}K_{ij\sigma,kl\tau}\sqrt{f_{kl\tau}\omega_{kl\tau}}\right]\mathbf{F}_n=\Omega_n^2\mathbf{F}_n \tag{20}$$

where $\omega_{ij\sigma} = \epsilon_{j\sigma} - \epsilon_{i\sigma}$ are the Kohn-Sham transition energies, $f_{ij\sigma} = n_{i\sigma} - n_{j\sigma}$ are the differences between the occupation numbers of the i-th and j-th states, the eigenvectors $\mathbf{F}_n$ are related to the transition oscillator strengths, and $K_{ij\sigma,kl\tau}$ is a coupling matrix given by:

$$K_{ij\sigma,kl\tau} = \iint \phi^*_{i\sigma}(\mathbf{r})\phi_{j\sigma}(\mathbf{r})\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \frac{\partial v^{\mathrm{xc}}_{\sigma}(\mathbf{r})}{\partial \rho_{\tau}(\mathbf{r}')}\right)\phi_{k\tau}(\mathbf{r}')\phi^*_{l\tau}(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \tag{21}$$

where i, j, σ are the occupied state, unoccupied state, and spin indices respectively, $\phi(\mathbf{r})$ are the Kohn-Sham wavefunctions, and $v^{\mathrm{xc}}(\mathbf{r})$ is the LDA exchange-correlation potential.

The TDLDA formalism is easy to implement in real space within the higher-order finite difference pseudopotential method [7]. The real-space pseudopotential code represents a natural choice for implementing TDLDA due to the real-space formulation of the general theory. With other methods, such as the plane wave approach, TDLDA calculations typically require an intermediate real-space basis. After the original plane wave calculation has been completed, all functions are transferred into that basis, and the TDLDA response is computed in real space [34]. The additional basis complicates calculations and introduces an extra error. The real-space approach simplifies implementation and allows us to perform the complete TDLDA response calculation in a single step.

We illustrate the TDLDA technique by calculating the absorption spectra of a sodium cluster. We chose sodium clusters as well-studied objects, for which accurate experimental measurements of the absorption spectra are available [35]. The ground-state structures of the clusters were determined by simulated annealing [24]. In all cases the obtained cluster geometries agreed well with the structures reported in other works [36]. Since the wavefunctions for the unoccupied electron states are very sensitive to the boundary conditions, TDLDA calculations need to be performed within a relatively large boundary domain. For sodium clusters we used a spherical domain with a radius of 25 a.u. and a grid spacing of 0.9 a.u. We carefully tested convergence of the calculated excitation energies with respect to these parameters.

The calculated absorption spectrum for Na_4 is shown in Figure 10 along with experiment. In addition, we illustrate the spectrum generated by considering transitions between the LDA

eigenvalues. The agreement between TDLDA and experiment is remarkable, especially when contrasted with the LDA spectrum. TDLDA correctly reproduces the experimental spectral shape, and the calculated peak positions agree with experiment within $0.1 - 0.2$ eV. The comparison with other theoretical work demonstrates that our TDLDA absorption spectrum is as accurate as the available CI spectra [37]. Furthermore, the TDLDA spectrum for the Na_4 cluster seems to be in better agreement with experiment than the GW absorption spectrum calculated in Ref. [38].

V. CONCLUSIONS

We have presented in this review chapter a real space method for describing the structural and electronic properties of materials and, in particular, confined systems. Real space methods offer a powerful approach to these systems. A few of the advantages of real space methods over "traditional" plane wave methods to the electronic structure problem are as follows: Real space methods are far easier to implement than plane wave codes with no loss of accuracy. This is especially true for parallel implementations where real space methods appear to be roughly an order of magnitude faster than comparable implementations with plane wave methods. They do not require the use of supercells for localized systems. No cell-cell interactions are present. Charged systems can be handled directly without artificial compensating backgrounds. Replication of vacuum is natural and minimized compared to extended basis sets. No Fast-Fourier Transforms are required and, consequently, global communications are minimized.

We have illustrated how this method can be applied to confined media. Specifically, we used real space method to calculate the photoemission spectra, Raman or vibrational spectra, polarizabilities and optical absorption spectra of clusters. By making comparisons with available experimental data, we have confirmed the accuracy and utility of real space methods.

While we focused in this review on small clusters, it is possible to apply these techniques

to quite large systems. For example, quantum dots with over 800 atoms have been examined with real space methods [40]. With increasingly efficient computer platforms and with new advances in algorithm developments, it is likely that larger systems will become routine in the near future.

REFERENCES

[1] J.R. Chelikowsky, and S.G. Louie, editors: *Quantum Theory of Real Materials*, (Kluwer Press, 1996) and references therein.

[2] W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).

[3] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[4] S. Lundqvist and N.H. March, *Theory of the Inhomogeneous Electron Gas*, (Plenum, NY), 1983 and refences therein.

[5] J.R. Chelikowsky, and M.L. Cohen: "Ab initio Pseudopotentials for Semiconductors," *Handbook on Semiconductors*, Editor: Peter Landsberg, (Elsevier, 1992), Vol. 1, p. 59.

[6] See for examples: A. Briley, M.R. Pederson, K.A. Jackson, D.C. Patton, and D.V. Porezag, Phys. Rev. B **58**, 1786 (1997); K.A. Jackson, M.R. Pederson, D.V. Porezag, Z. Hajnal and Th. Fraunheim, Phys. Rev. B **55**, 2549 (1997); J. R. Chelikowsky and S.G. Louie, Phys. Rev. B **29**, 3470 (1984); R.W. Jansen and O.F. Sankey, Phys. Rev. B **36**, 6520 (1987) and references therein.

[7] 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**, 11355 (1994); J.R. Chelikowsky, N. Troullier, X. Jing, D. Dean, N. Binggeli, K. Wu and Y. Saad, Computer Physics Communications **85**,325 (1995); J.R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad: "Algorithms for Predicting Properties of Real Materials on High Performance Computers," Proceedings of the "Toward Teraflop Computing Conference and New Grand Challenge Applications," Baton Rouge, LA 1994, editors: R.K. Kalia and P. Vashista, (Nova, New York), 1995, p. 13; J.R. Chelikowsky, X. Jing, K. Wu and Y. Saad, Phys. Rev. B **53**, 12071 (1996).

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

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

[10] B. Fornberg and D. M. Sloan, *Acta Numerica 94*, Editor A. Iserles, Cambridge University Press, 1994.

[11] R.B. Morgan and D.S. Scott, SIAM J. Sci. Stat. Comput. **7**, 817 (1986).

[12] Y. Saad, A. Stathopoulos, J.R. Chelikowsky, K. Wu and S. Öğüt, BIT **36**, 563 (1996).

[13] C.H. Tong, T.F. Chan and C.C. J. Kuo, SIAM J. Sci. Stat. Comput. **13**, 227 (1992).

[14] B.N. Parlett, *The Symmetric Eigenvalue Problem*, (Prentice Hall, Englewood Cliffs), 1980.

[15] Y. Saad, *Iterative Methods for Sparse Linear Systems*, (PWS Publishing Company, Boston), 1196.

[16] G. Karypis and V. Kumar, "Parallel multilevel graph partitioning," *Proceedings of the 10th International Parallel Processing Symposium*, 1996, p. 314.

[17] Examples of similar approaches can be found in the following: E.L. Briggs, D.J. Sullivan and J. Bernholc, Phys. Rev. B **52**, 5471 (1995); F. Gygi and G. Galli, Phys. REv. B **52**, 2229 (1995); G. Zumbac, N. Modine and E. Kaxiras, Solid State Comm. **99**, 57 (1996); T. Hoshi and T. Fujiwra, J. Phys. Soc. of Japan **66**, 3710 (1997), and references therein.

[18] R. Kubo, Rep. Prog. Theor. Phys. **29**, 255, (1966).

[19] H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin),1984.

[20] F.W. Kutzler and G.S. Painter, Phys. Rev. B **45**, 3236 (1992).

[21] D. Deaven and K.M. Ho, Phys. Rev. Lett. **75**, 288 (1995).

[22] O. Cheshnovsky, S.H. Yang, C.L. Pettiett, M.J. Craycraft, Y. Liu, and R.E. Smalley, Chem. Phys. Lett. **138**, 119 (1987).

[23] D. Tomanek and M. Schlüter, Phys. Rev. Lett. **56**, 1055 (1986).

[24] N. Binggeli and J.R. Chelikowsky, Phys. Rev. B **50**, 11764(1994).

[25] E.C. Honea, A. Ogura, C.A. Murray, K.Raghavachari, O. Sprenger, M.F.,Jarrold, and W.L.Brown, Nature **366**, 42 (1993).

[26] X. Jing, N. Troullier, J.R. Chelikowsky, K. Wu and Y. Saad, Solid State Comm. **96**, 231 (1995).

[27] R. Fournier, S.B. Sinnott, and A.E. DePristo, J. Chem. Phys. **97**, 4149 (1992).

[28] C. Rohlfing and K. Raghavachari, J. Chem. Phys. **96**, 2114 (1992).

[29] R. Schäfer, S. Schlect, J., Woenckhaus and J.A. Becker, Phys. Rev. Lett. **76**, 471(1996).

[30] S. Baroni, P. Gianozzi and A. Testa, Phys. Rev. Lett. **58**, 1861 (1987).

[31] X. Gonze, D.C. Allan and M.P. Teter, Phys. Rev. Lett. **68**, 3603 (1992).

[32] I. Vasiliev, S. Öğüt and J.R. Chelikowsky, Phys. Rev. Lett. **78**, 4805 (1997).

[33] M. E. Casida, in *Recent Advances in Density-Functional Methods*, edited by D. P. Chong (World Scientific, Singapore, 1995), Part I, Chap. 5; in *Recent Developments and Applications of Modern Density Functional Theory*, edited by J. M. Seminario (Elsevier, Amsterdam, 1996).

[34] X. Blase, A. Rubio, S. G. Louie, and M. L. Cohen, Phys. Rev. B **52**, R2225 (1995).

[35] C.RC. Wang, S. Pollack, D. Cameron and M.M. Kappes, Chem. Phys. Lett. 93, 3787 (1990)

[36] I. Moullet, J. L. Martins, F. Reuse, and J. Buttet, Phys. Rev. B **42**, 11598 (1990); J. L. Martins, J. Buttet, and R. Car, Phys. Rev. B **31**, 1804 (1985).

[37] V. Bonacic-Koutecky, P. Fantucci, and J. Koutecky, J. Chem. Phys. **93**, 3802 (1990); Chem. Phys. Lett. **166**, 32 (1990).

[38] G. Onida, L. Reining, R.W. Godby, R. del Sole and W. Andreoni, Phys. Rev. Lett. **75**, 181 (1995).

[39] Igor Vasiliev, S. Öğüt, J.R. Chelikowsky, Phys. Rev. Lett. **82**, 1919 (1999).

[40] S. Öğüt, J.R. Chelikowsky, and S.G. Louie: "Quantum Confinement and Optical Gaps in Si Nanocrystals," *Phys. Rev. Lett.* **79**, 1770 (1997).

TABLES

TABLE I. Calculated and experimental vibrational frequencies in a Si_4 cluster. See Figure 7 for an illustration of the normal modes. The frequencies are given in cm^{-1}.

	B_{3u}	B_{2u}	A_g	B_{3g}	A_g	B_{1u}
Experiment [25]			345		470	
Dynamical Matrix (This work)	160	280	340	460	480	500
MD simulation (This work)	150	250	340	440	490	500
HF [28]	117	305	357	465	489	529
LCAO [27]	55	248	348	436	464	495

TABLE II. Static dipole moments and average polarizabilities of small silicon and germanium clusters.

Silicon			Germanium		
cluster	$\|\mu\|$ (D)	$\langle\alpha\rangle$ ($Å^3/atom$)	cluster	$\|\mu\|$ (D)	$\langle\alpha\rangle$ ($Å^3/atom$)
Si_2	0	6.29	Ge_2	0	6.67
Si_3	0.33	5.22	Ge_3	0.43	5.89
Si_4	0	5.07	Ge_4	0	5.45
Si_5	0	4.81	Ge_5	0	5.15
Si_6 (I)	0	4.46	Ge_6 (I)	0	4.87
Si_6 (II)	0.19	4.48	Ge_6 (II)	0.14	4.88
Si_7	0	4.37	Ge_7	0	4.70

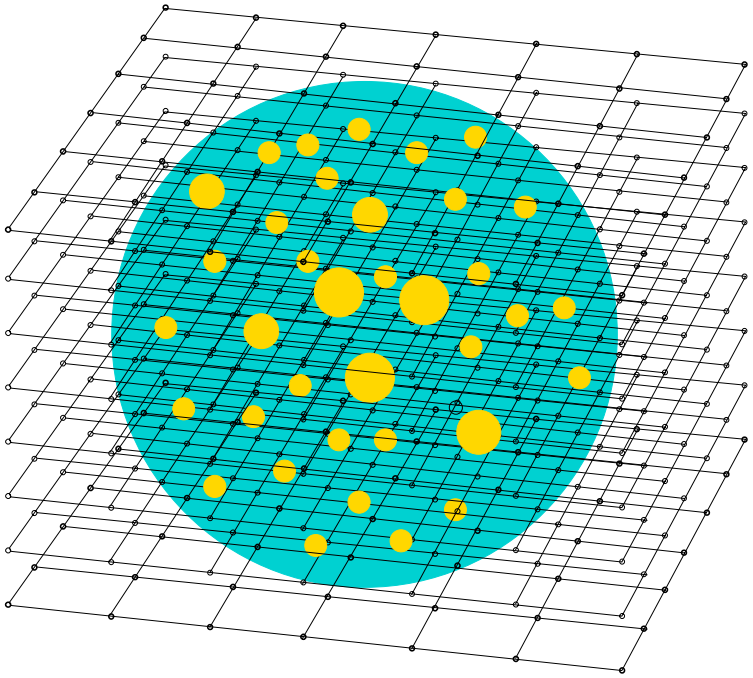

FIG. 1. Uniform grid illustrating a typical configuration for examining the electronic structure of a localized system. The gray sphere represents the domain where the wave functions are allowed to be nonzero. The light spheres within the domain are atoms.

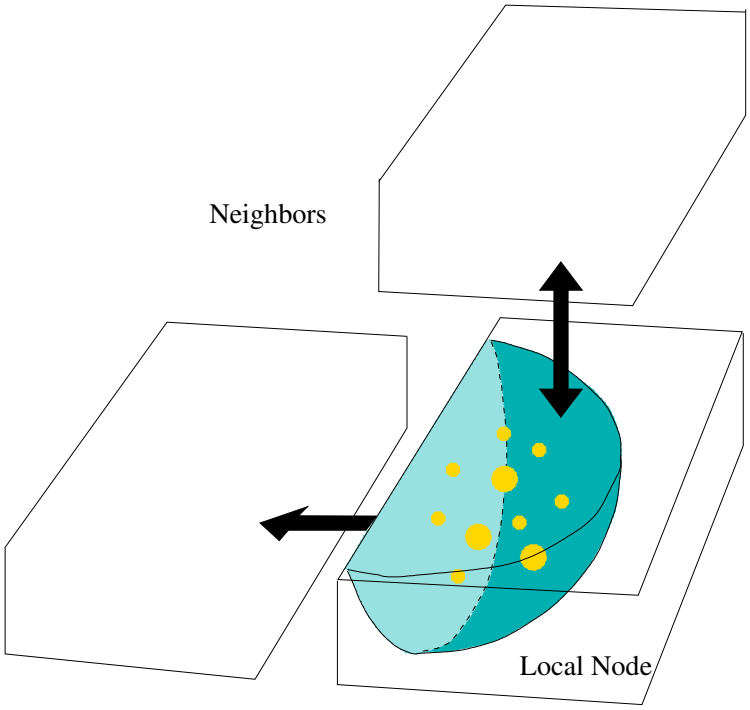

FIG. 2. An example of a possible decomposition. The subdomains illustrated are assigned to a particular processor. Although the subdomains are shown as cubic, they can be chosen to be an arbitrary configuration. See Figure 1.

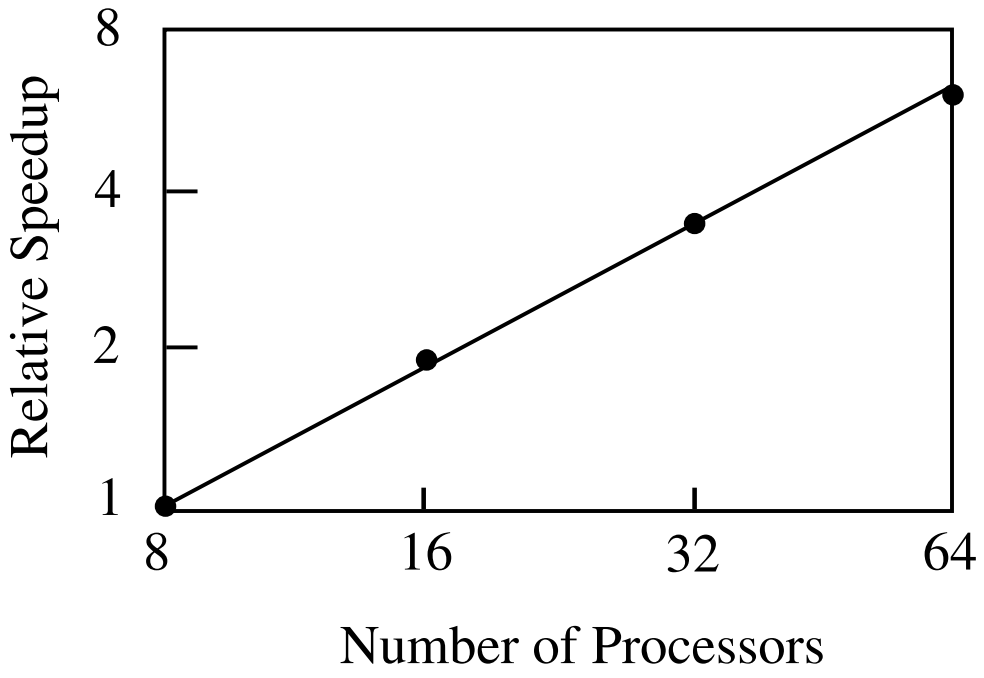

FIG. 3. Speedup efficiency for large silicon cluster on a massively parallel platform.

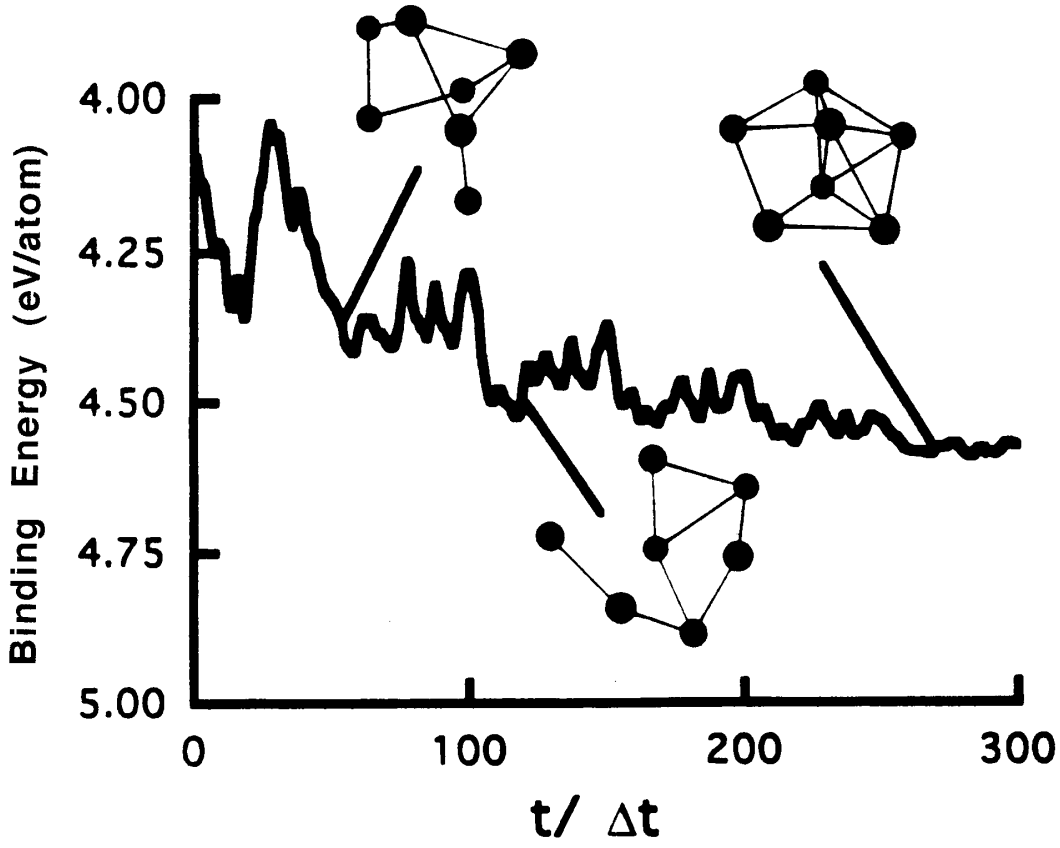

FIG. 4. Binding energy of Ge_7 during a Langevin simulation. The initial temperature is 2800 K; the final temperature is 300 K. Bonds are drawn for interatomic distances of less than 2.5Å. The time step is 7 fs.

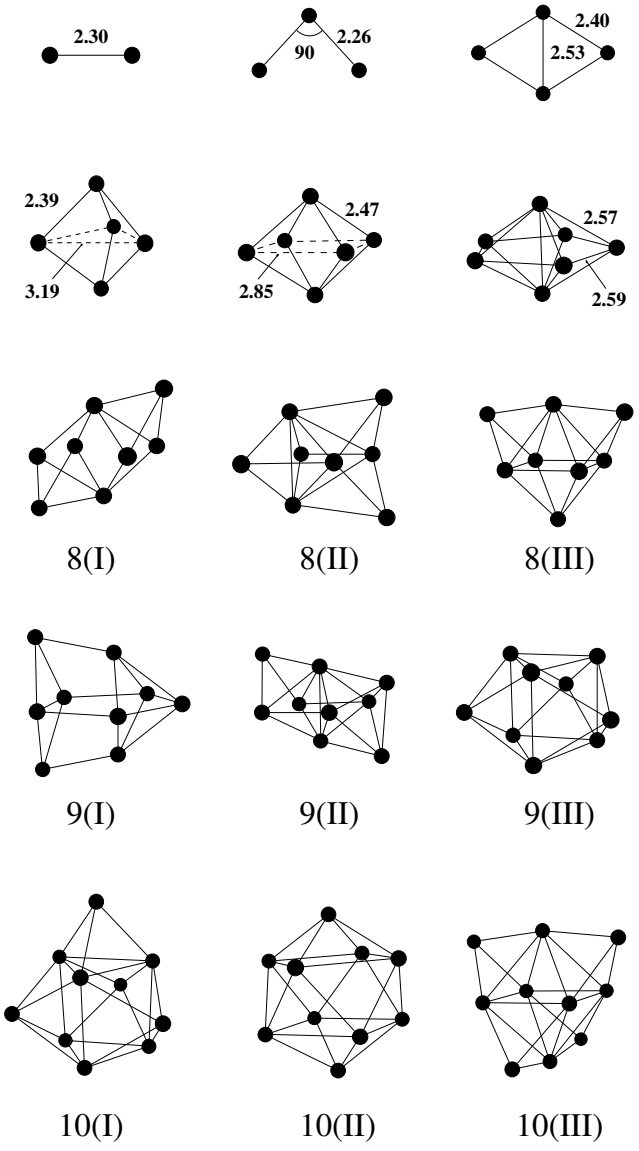

FIG. 5. Ground state geometries and some low-energy isomers of Ge_n ($n \leq 10$) clusters. Interatomic distances (in Å) are given for clusters with $n \leq 7$. For $n > 8$, the lowest energy isomer is given by (I).

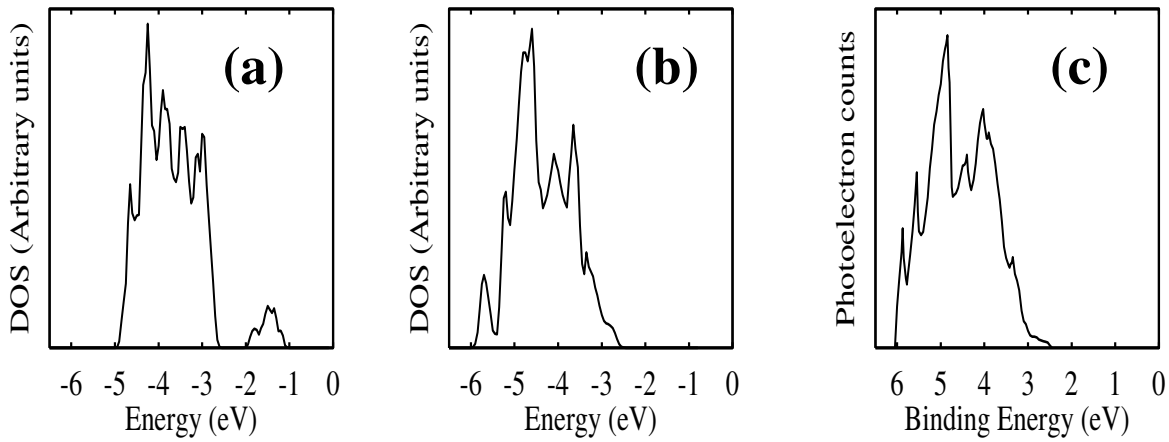

FIG. 6. (a) Calculated density of states for Ge_{10}^{-}. (b) Experimental photoemission spectra from Ref. 22.

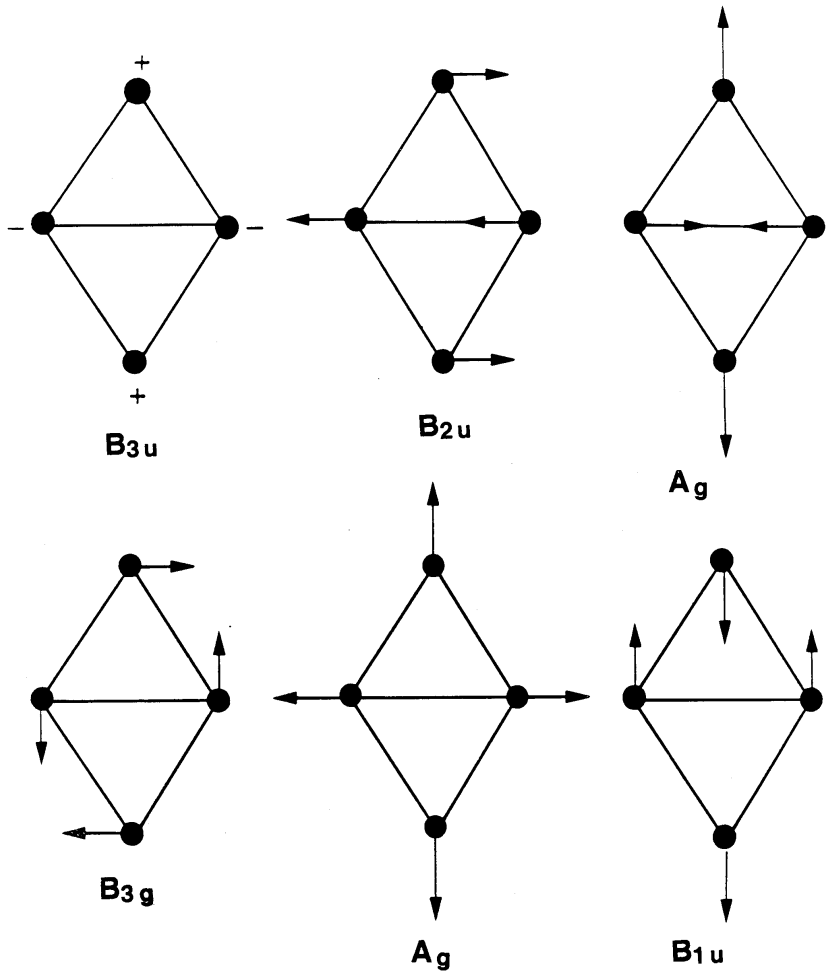

FIG. 7. Normal modes for a Si_4 cluster. The + and − signs indicate motion in and out of the plane, respectively.

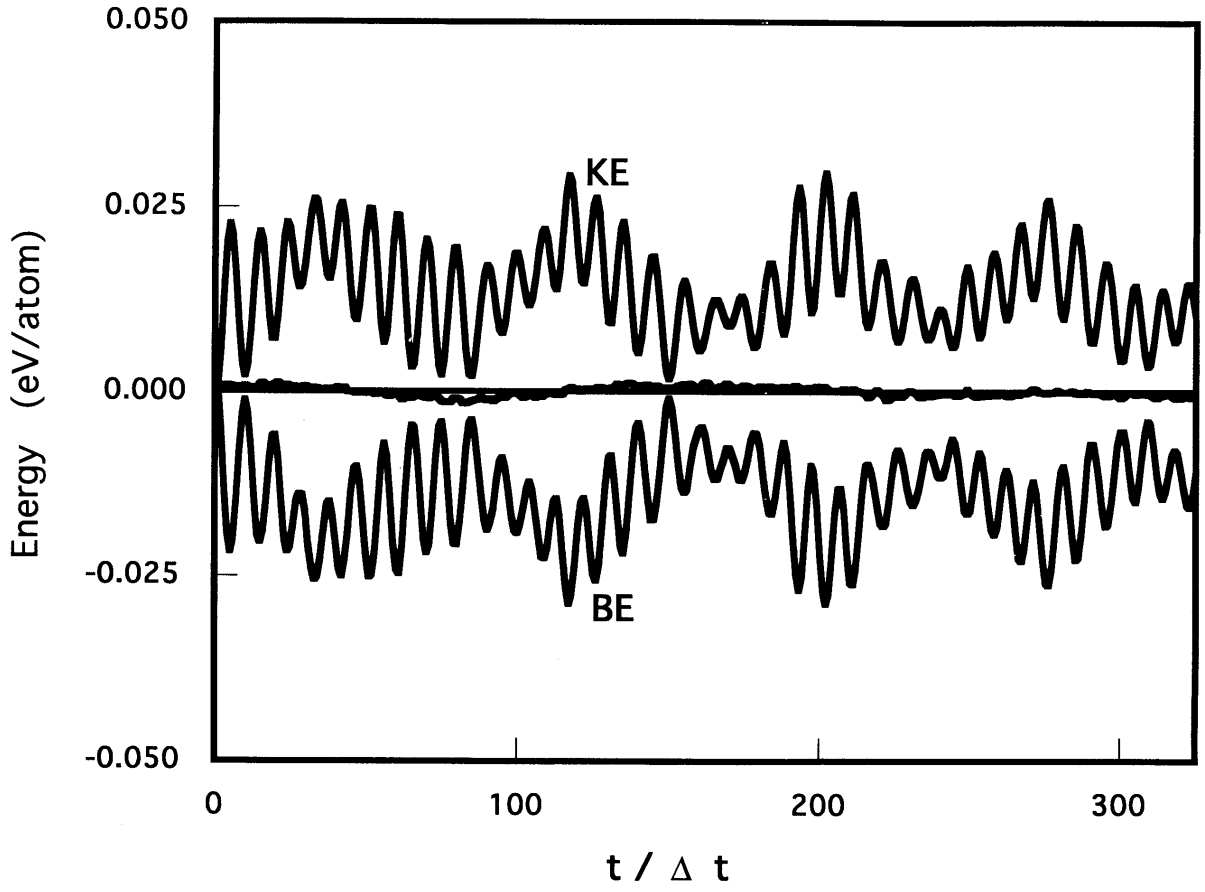

FIG. 8. Simulation for a Si_4 cluster. The kinetic energy (KE) and binding energy (BE) are shown as a function of simulation time. The total energy (KE+BE) is also shown with the zero of energy taken as the average of the total energy. The time step, Δt, is 7.4fs.

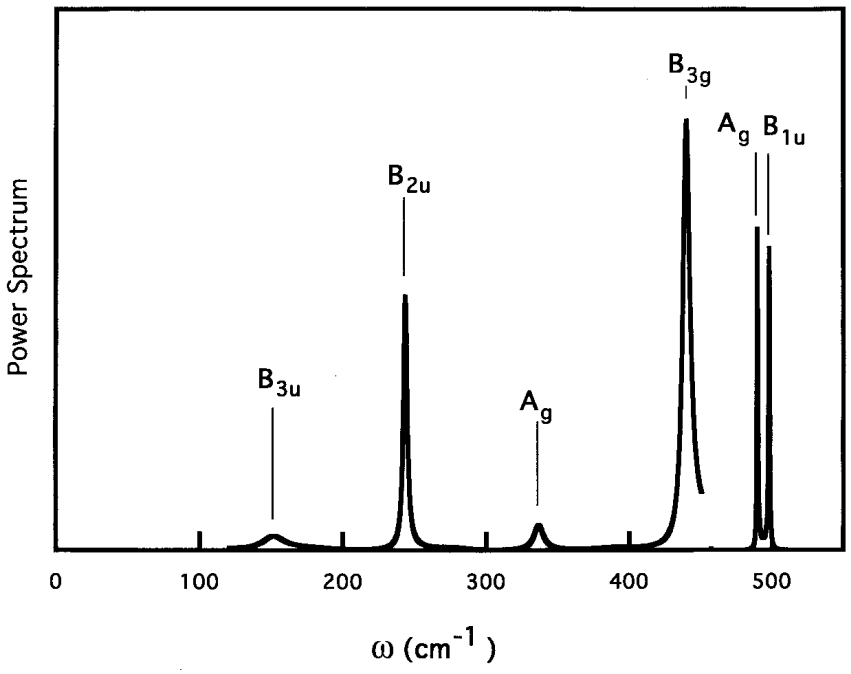

FIG. 9. Power spectrum of the vibrational modes of the Si_4 cluster. The simulation time was taken to be 4 ps. The intensity of the B_{3g} and (A_g,B_{1u}) peaks has been scaled by 10^{-2}.

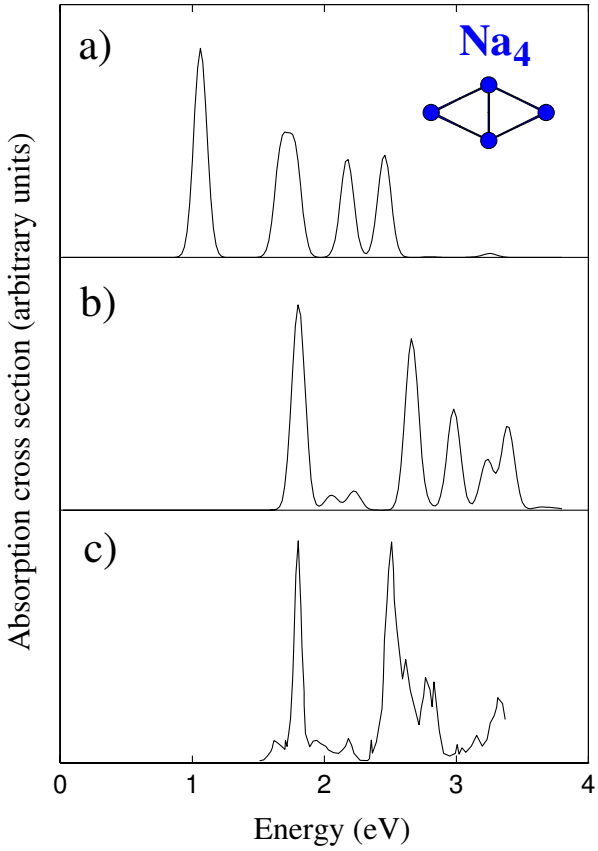

FIG. 10. The calculated and experimental absorption spectrum for Na_4. (a) shows a local density approximation to the spectrum using Kohn-Sham eigenvalues. (b) shows a TDLDA calculation. Technical details of the calculation can be found in [39]. (c) panel is experiment from [35].