

Kernel polynomial method for a nonorthogonal electronic-structure calculation of amorphous diamond

H. Röder and R. N. Silver

Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

D. A. Drabold and Jian Jun Dong

Department of Physics and Astronomy, Condensed Matter and Surface Science Program, Ohio University, Athens, Ohio 45701

(Received 16 September 1996)

The Kernel polynomial method (KPM) has been successfully applied to tight-binding electronic-structure calculations as an $O(N)$ method. Here we extend this method to nonorthogonal basis sets with a sparse overlap matrix \mathbf{S} and a sparse Hamiltonian \mathbf{H} . Since the KPM method utilizes matrix vector multiplications it is necessary to apply $\mathbf{S}^{-1}\mathbf{H}$ onto a vector. The multiplication of \mathbf{S}^{-1} is performed using a preconditioned conjugate-gradient method and does not involve the explicit inversion of \mathbf{S} . Hence the method scales the same way as the original KPM method, i.e., $O(N)$, although there is an overhead due to the additional conjugate-gradient part. We apply this method to a large scale electronic-structure calculation of amorphous diamond. [S0163-1829(97)04519-0]

Electronic-structure calculations are limited by the resources necessary to diagonalize large matrices. The demand on computer memory grows as N^2 and CPU time as N^3 , where N is the dimension of the matrices. A variety of methods at hand use iterative methods requiring only matrix-vector multiplications (MVM's), and therefore can take advantage of the sparsity of the matrices. Most of these methods use sophisticated variants of the Lanczos method (e.g., the lookahead Lanczos algorithm). It is, however, difficult to obtain more than a small percentage of the number of states with such a procedure, since their convergence rate depends on the relative spacing of eigenvalues, which can become extremely small in the center of bands. If one relaxes the constraint of requiring exact eigenvalues and by analogy with experiment is satisfied with the eigenvalue distribution convoluted with a well-defined resolution function, then the kernel polynomial method (KPM)^{1,2} and simple variants thereof³ have been shown to be successful in electronic-structure calculations.^{4,5} To date, all such applications used an orthogonal basis set. However, a large variety of quantum chemical and *ab initio* local-density approximation (LDA) calculations work with basis functions that are nonorthogonal, because this allows use of physical intuition in the choice and truncation of basis functions. It therefore becomes necessary to extend the KPM to approximately solve the system

$$\mathbf{H}|\Psi\rangle = \lambda \mathbf{S}|\Psi\rangle, \quad (1)$$

where \mathbf{H} and \mathbf{S} are sparse symmetric (Hermitian) matrices and S is semipositive definite. In order to proceed it is necessary to present a short summary of the KPM.

In the KPM the density of states $N(E) = \sum_{\lambda} \delta(E - E_{\lambda})$ is approximated by an expansion in orthogonal polynomials. Depending on the choice of the orthogonal polynomials used, the matrix \mathbf{H} needs to be rescaled to \mathbf{X} such that the eigenvalue spectrum of \mathbf{X} lies in the range in which the polynomials are defined. From now on we will use Chebychev

polynomials $T_m(x)$, which are defined in $[-1, 1]$. The recursion relation for the Chebychev polynomials is $T_m(x) = 2xT_{m-1}(x) - T_{m-2}(x)$ with $T_0(x) = 1$, and $T_1(x) = x$. The density of states can now be approximated as

$$N(x) = \frac{1}{\pi\sqrt{1-x^2}} \left[\mu_0 + 2 \sum_{m=1}^{\infty} \mu_m T_m(x) \right], \quad (2)$$

where the expansion coefficients (moments) are given by

$$\mu_m = \text{Tr}\{T_m(\mathbf{X})\} = \int_{-1}^1 T_m(x) N(x) dx. \quad (3)$$

In practice one can only calculate a finite number of terms in Eq. (2). The truncation effects result in typical Gibbs oscillation. These can be damped out by multiplying with appropriate Gibbs damping factors g_m^M leading to

$$N_K(x) = \frac{1}{\pi\sqrt{1-x^2}} \left[\mu_0 + 2 \sum_{m=1}^M \mu_m g_m^M T_m(x) \right]. \quad (4)$$

The truncation together with the damping corresponds to a convolution of the correct density of states with a finite resolution function or "kernel," whose width decreases with the number of moments M . An alternative way to analyze the modified moment expansion (2) is to directly use an image reconstruction method like MaxEnt that can achieve significantly higher energy resolution for isolated states and band edges.^{6,7}

In order to generate the moments (3) in time and memory requirements of $O(N)$, one can utilize a stochastic method by approximating the trace in Eq. (3) by a sum over N_r Gaussian random vectors $|r\rangle$,

$$\mu_m = \lim_{N_r \rightarrow \infty} \left[\frac{1}{N_r} \sum_r \langle r | T_m(X) | r \rangle \right]. \quad (5)$$

The expectation value of $T_m(\mathbf{X})$ is calculated recursively by taking scalar products between $|r\rangle$ and the vectors $|m\rangle$ defined by

$$|m\rangle = 2X|m-1\rangle - |m-2\rangle, \quad (6)$$

with $|0\rangle = |r\rangle$ and $|1\rangle = \mathbf{X}|r\rangle$.

In order to extend the KPM to eigenvalue problems described by Eq. (1), the straightforward way is to multiply Eq. (1) from the left with \mathbf{S}^{-1} to obtain

$$\mathbf{S}^{-1}\mathbf{H}|\Psi\rangle = \tilde{\mathbf{H}}|\Psi\rangle = \lambda|\Psi\rangle, \quad (7)$$

and then to use the conventional KPM described above with $\tilde{\mathbf{H}} = \mathbf{S}^{-1}\mathbf{H}$ as the underlying matrix. However, the inverse of the overlap matrix is neither readily available nor is it in general sparse. But since within the KPM only MVMs are required to generate the moments there is a way out. A MVM in Eq. (6) can be written as a linear system of equations such as

$$\mathbf{S}|x\rangle = \mathbf{H}|m-1\rangle, \quad (8)$$

with unknown $|x\rangle$ and known right-hand side $\mathbf{H}|m-1\rangle$. There is a huge variety of methods available to solve such a problem. The typical method for large systems of linear equations with multiple right-hand sides is to generate a Cholesky factorization of \mathbf{S} , $\mathbf{S} = \mathbf{L}\mathbf{L}^H$. Again this is not practical since a Cholesky factorization takes $O(N^3)$ in CPU time and does not preserve sparsity. Therefore a strictly iterative method is necessary to solve Eq. (8). We choose the preconditioned conjugate-gradient (PCG) method, because it allows us to control convergence by the choice of the preconditioner.⁸ The same preconditioner will be used over and over again, so it pays to obtain a good sparse approximation to \mathbf{S}^{-1} . We use an incomplete stabilized Cholesky (ISCHO) factorization for this purpose. In the ISCHO one performs a Cholesky factorization in which one throws away all elements that are not within the sparsity pattern of \mathbf{S} , i.e., \mathbf{S} is approximated by

$$\mathbf{S} \approx \tilde{\mathbf{L}}\tilde{\mathbf{L}}^H = \tilde{\mathbf{S}}. \quad (9)$$

This renders the usually extremely stable Cholesky algorithm unstable due to the appearance of negative or zero diagonal elements in $\tilde{\mathbf{S}}$. We correct this effect by adding a diagonal matrix with small elements of equal size until the resulting $\tilde{\mathbf{S}}$ is positive definite. In order to obtain the best possible results from this preconditioner, one needs to choose the correction as small as possible to remain close to the original overlap matrix. An additional benefit of this method is that one can choose the solution of $\tilde{\mathbf{S}}|x\rangle = \mathbf{H}|m-1\rangle$ as a good initial vector for the PCG method. Using preconditioning typically reduces the number of MVMs necessary for the solution of Eq. (8) by at least a factor of 10. Combining conventional KPM with this PCG method allows a feasible calculation of the moments for the nonorthogonal eigenvalue problem (1). The memory requirement scales as the number of nonzero matrix elements of \mathbf{S} and \mathbf{H} and therefore as $O(N)$, if they are sparse. The number of MVMs needed as compared to orthogonal KPM is multiplied by the number of MVMs performed within the PCG to solve the system of Eq. (8). For the examples below this was a factor of about 30.

Although this appears large the method is intrinsically $O(N)$ (for sparse matrices). The matrix dimension at which the KPM becomes more efficient is $N \approx 1000$ as compared to a conventional $O(N^3)$ matrix diagonalization, if one wants the total energy at the Fermi level accurate up to four digits.

The particular physical problem that we address is to compute the density of electronic states of a large and realistic model of an amorphous semiconductor. Issues include: (1) the existence and origin of gap states, (2) the nature of the band tail states and the structure of the tailing in a Harris functional LDA calculation, and (3) total band energies computed by appropriately integrating the density of states (DOS). The structural models of this paper are 4096 atom cubic supercells of *a*-diamond (*a*-D) provided by Djordjevic *et al.*¹² with and without fourfold rings. We note that *a*-D is a hypothetical, entirely fourfold material at a density of 3.52 gm/cm³ as crystalline diamond, but with topological (primarily bond angle) disorder. The closest experimental analogues are probably *a*-Si and *a*-Ge. That the cell of Ref. 12 is a structurally credible model of amorphous diamond can be inferred from our LDA relaxations of smaller versions (216 and 512 atom models) constructed in an analogous fashion with the Wooten-Weaire-Winer method.¹³ We found that these smaller supercells of *a*-D were practically unchanged upon relaxation.¹⁰

To illustrate the method, we employed the Harris functional LDA Hamiltonian of Sankey and Niklewski⁹ to compute the overlap and Hamiltonian matrix. The essential approximations are (1) the LDA, (2) Bachelet-Hamann-Schlüter pseudopotentials, (3) a minimal basis set of confined pseudoatomic orbitals (one *s* and three *p*'s per site), a confinement radius of $4.1a_B$ was used, and (4) angular momentum (multipole) expansions for the three center matrix elements. The transferability of this basis set, as gauged by its phase diagram, is discussed in detail in Sec. II of Ref. 10. A fuller discussion of the compact basis functions can be found in Ref. 9. While the method is approximate, it provides a rather reliable description of carbon systems. At a practical level, exact diagonalization of the nonorthogonal eigenvalue problem becomes inconvenient for ≈ 500 atom systems. Ordejoń *et al.*¹¹ have made important extensions of the method to yield linear system size scaling for atomistic force calculations. To the extent that the present method provides an accurate density of electronic states, and the atomistic force methods based upon truncation do not (since they work in the occupied subspace), the schemes are complementary. To our knowledge, this is the largest *ab initio* spectral calculation undertaken to date.

In this type of calculation, convergence to the exact density of states is determined by the number of moments used in the KPM or maximum entropy reconstruction and the number of random vectors used. In large spectral density estimation problems, a very small number of random vectors is adequate (four were used in Figs. 1 and 2); the number of moments is the prime determinant of convergence, and 512 moments clearly provide a highly resolved density of states.

Our work reveals that fourfold rings produce a clear electronic signature of gap states as seen in Fig. 2. As fourfold rings imply substantial strain in a generally tetrahedral network, this result is not surprising. It is, however, gratifying

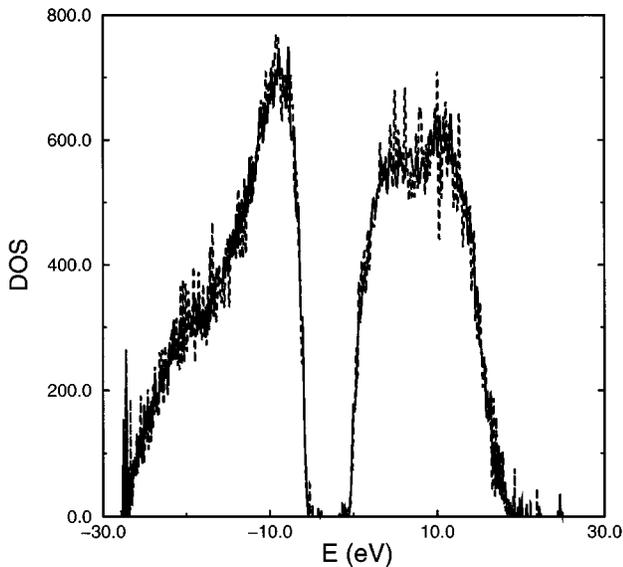


FIG. 1. The density of states for 4096 atom models of amorphous diamond with (solid line) and without (dashed line) four member rings. The DOS was obtained using MaxEnt from 512 moments averaged over four-random vectors.

to see such an unambiguous electronic manifestation of a structural distortion. We find it encouraging that the location of the states is so precisely characterized from (global) moment information, and also that this accuracy is possible even though the integrated weight of the gap states is tiny ($\approx 0.1\%$ of the total DOS). The discreteness of the spectrum is, of course, due to the finite, albeit large system size. We have not attempted to “extrapolate” the DOS to the thermodynamic limit in which case the spectrum would be quasi-continuous. We note that the difficult component of the total (system) energy is the band structure term (an incomplete integral over the DOS), which can be directly computed

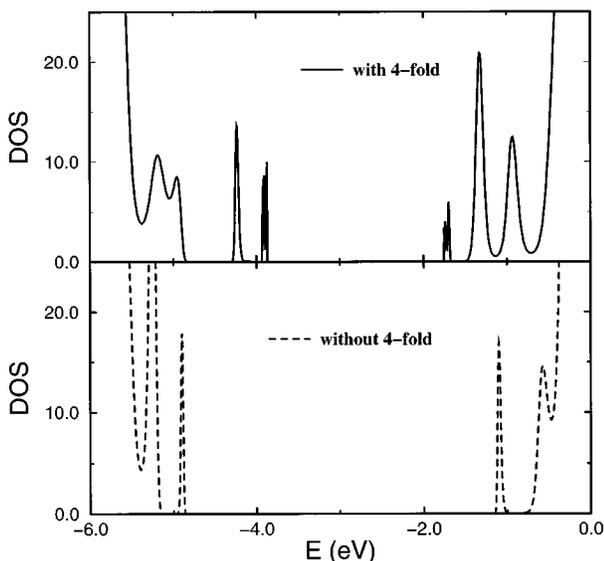


FIG. 2. An enlargement of Fig. 1 for the region around the band gap depicting a small number of intragap states showing the difference due to varying topologies.

from the KPM or maximum entropy DOS and would be immediately useful for methods requiring only total energies (like Monte Carlo) and is a possible starting point for atomistic force calculations. We do not develop these ideas here.

In an earlier paper, Dong and Drabold¹⁴ used a simpler orthogonal tight-binding Hamiltonian to study the evolution of electron states from midgap (localized by the topological disorder) to valence (extended), and also noted the existence of approximately exponential band tails according to a maximum entropy analysis of the DOS. There are no direct experimental measurements of the DOS of *a*-D, but because of its tetrahedral topology we might expect the DOS to qualitatively resemble that of *a*-Si or *a*-Ge. We note that the width of the valence states is narrower than one would expect on these grounds, particularly in connection with the lower energy “*s*-like” part of the valence states. Perhaps fortuitously, this is rather well reproduced in the empirical tight-binding calculation. The band tails in this calculation do not appear to be exponential; they fall off more rapidly, and it is not clear what the structure of the tails would be in the thermodynamic limit. It is worth noting for completeness that there is no rigorous reason to expect total energy methods (empirical or LDA) to provide quantitatively accurate band structure information, though there is a wealth of empirical evidence that such an approach is often informative.

From the point of view of experimental work on amorphous semiconductors there is interest in the microscopic origin of band tail and deep states. The density of states clearly reveals a variety of band tail and deep states (we do not attempt to differentiate between the two here). Since the present calculation provides no information about eigenvectors conjugate to a specific energy eigenvalue, we use results from exact diagonalization on the 512 atom cell also constructed by Djordjevic, Thorpe, and Wooten¹² to infer the likely structural origins of the deep levels that appear in Fig. 1. We found in Ref. 15 that states near the valence edge originate primarily in bond angle distortions. It is well known on elementary grounds that an ideal sp^3 dangling bond produces a state midgap in a column IV material. Near the conduction edge, we have seen that badly strained bonds (with bondlengths greater than 1.8 Å) manifest themselves, although one must be careful about over generalization since there are bond angle distortions present as well. Thus, with some degree of oversimplification, a picture emerges in which deep states near the valence edge originate in bond angle strains, midgap states from sp^3 dangling bonds (shifted somewhat depending upon the details of their local environment), and deep donor states from conformations with severe bond length and bond angle distortions. We observe that large numbers of sp^2 sites modify the picture considerably as we discuss in detail elsewhere.¹⁵

In this presentation we have described how to extend the KPM method to nonorthogonal basis sets. Although the overhead per MVM is considerably larger than for orthogonal sets, the method remains $O(N)$ and allows for the approximate diagonalization of system sizes on a workstation that were previously extremely extensive calculations even on large mainframes. The preconditioner used in this work is very general and gives reasonable convergence rates for a wide range of problems. If the overlap matrix becomes very singular, it may be necessary to specialize the preconditioner

for this class of problem. For other problems a different preconditioner, e.g., symmetric successive overrelation⁸ could prove to be more efficient. If a large number of modified moments $M > 1000$ is required and the overlap matrix is very stiff, repeated application of $\tilde{\mathbf{H}}$ from Eq. (7) could result in numerically non-Hermitian behavior. In this case one needs to resort to using $\mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2}$ as $\tilde{\mathbf{H}}$, and an iterative, $O(N)$ solution of $|y\rangle = \mathbf{S}^{-1/2}|x\rangle$ is necessary. If it is possible to perform a fast wavelet transform, the algorithm described in Ref. 16 solves this problem.

The above calculations are easily extended to the calculation of operator expectation values such as the charge density and projected occupation numbers. The calculation of forces

can also be performed within this framework. Since we use only MVMs this type of calculation is easily parallelized; hence, the diagonalization of truly huge electronic structure problems in the order of over 10^5 atoms may be within reach.

The relative ease of this calculation opens many additional avenues for study, including atomistic force calculations, studies of defects and localization in cells large enough to be truly free of finite size artifacts, calculation of the vibrational states,¹⁷ and others.

Supported in part by the Office of Basic Energy Sciences of the U.S. Department of Energy and under NSF Grant No. DMR 93-22412.

¹R. N. Silver and H. Roeder, *Int. J. Mod. Phys. C* **5**, 735 (1994).

²L. W. Wang, *Phys. Rev. B* **49**, 10 154 (1994).

³S. Goedecker and L. Colombo, *Phys. Rev. Lett.* **73**, 122 (1994).

⁴R. N. Silver *et al.*, *J. Comput. Phys.* **124**, 115 (1996).

⁵A. F. Voter *et al.*, *Phys. Rev. B* **53**, 12 733 (1996).

⁶R. N. Silver *et al.*, *Simulation Multiconference '96 Proceedings, High Performance Computing 96* (The Society for Computer Simulation International, San Diego, 1996), Vol. 130.

⁷D. A. Drabold and O. F. Sankey, *Phys. Rev. Lett.* **93**, 3631 (1993); D. A. Drabold *et al.*, *Solid State Commun.* **96**, 833 (1995).

⁸W. Hackbusch, in *Iterative Solution of Large Sparse Systems of Equations* (Springer-Verlag, New York, 1994).

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

¹⁰D. A. Drabold *et al.*, *Phys. Rev. B* **49**, 16 415 (1994).

¹¹P. Ordejoń *et al.*, *Phys. Rev. B* **51**, 1456 (1995), and references therein.

¹²B. R. Djordjevic, M. F. Thorpe, and F. Wooten, *Phys. Rev. B* **52**, 5685 (1995).

¹³F. Wooten *et al.*, *Phys. Rev. Lett.* **54**, 1392 (1985).

¹⁴J. Dong and D. A. Drabold, *Phys. Rev. B* **54**, 10 284 (1996).

¹⁵D. A. Drabold *et al.*, *Phys. Rev. B* **54**, 5480 (1996).

¹⁶G. Beylkin *et al.*, in *Wavelets and their Applications*, edited by Ruskai *et al.*, (Jones and Bartlett, 1992).

¹⁷P. Ordejoń *et al.*, *Phys. Rev. Lett.* **75**, 1324 (1995).