Maximum Entropy Approach for Linear Scaling in the Electronic Structure Problem

David A. Drabold[a]
Department of Physics, University of Illinois, Urbana, Illinois 61801

Otto F. Sankey
Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287
(Received 24 February 1993)

We present a method for obtaining the band-structure energy and density of states for large, sparse Hamiltonian matrices as occur in electronic structure problems. The computation scales linearly with the dimension of the matrix. We use a statistical approach using random vectors along with the principle of maximum entropy to obtain highly accurate estimates for integrals over the density of states. We offer several tests of the approach using tight-binding Hamiltonians and apply the scheme to large icosahedral fullerenes with 720 and 2160 atoms.

PACS numbers: 71.10.+x, 71.25.−s

One of the current principal thrusts of solid state theory is to compute the energetics of very large systems using reliable electronic structure based methods. The importance of the problem is obvious: reliable electronic structure calculations for large (> 500 atom) systems are the key to understanding crucial aspects of biological molecules such as DNA, and the proper modeling of systems like amorphous materials and glasses, liquids, and surfaces. Electronic structure calculations in all of these areas are either currently impossible or plagued by finite size effects, which can be severe.

The ultimate bottleneck in these calculations is either the exact diagonalization of a local orbital matrix [1], which scales as N^3 (for N the dimension of the matrix), or the orthonormalization step in plane-wave [2,3] methods which also scales as N^3. In this Letter we offer a new *statistical* approach which scales linearly with N and estimates the entire eigenvalue spectrum. We anticipate that this statistical scheme will also be useful in plane-wave calculations, where only a small part of the spectrum is wanted.

There have recently been several "order N" [$O(N)$] methods proposed in the literature [4–7], all of which depend upon some assumption of electronic localization. Our proposal is very different from these. We adopt a statistical approach, an idea similar to importance sampling and the maximum entropy principle [8] to efficiently obtain incomplete (but sufficient) information to compute the total density of states (DOS), band-structure energy (BSE) integrals, and related quantities.

Let H be a large, sparse, Hermitian matrix of dimension N. Ideally one would like the eigenvalue spectrum ϵ_i, where $H\psi_i = \epsilon_i\psi_i$. In this work, we seek the DOS $\rho(E)$ and particularly the BSE obtained by integrating $E \times \rho(E)$ up to the Fermi level. We note that all the information carried by the density of states is contained in any one single vector ξ in the family of vectors of the form

$$\xi = \sum_{j=1}^{N} e^{i\phi_j}\psi_j/\sqrt{N}, \tag{1}$$

where ϕ_j specifies an arbitrary phase. Note that the expectation value of the DOS operator $\hat{\rho}(E) = \delta(E - \hat{H})$ between any ξ gives the exact density of states; $\rho(E) = \langle\xi|\hat{\rho}(E)|\xi\rangle$. The vectors ξ equally weigh all of the eigenvalues of the spectrum of H. For this reason we call such ξ *impartial vectors*. The Hamiltonian H has moments μ_n,

$$\mu_n = \int_{-\infty}^{\infty} dE E^n \rho(E) = \frac{1}{N}\mathrm{Tr}H^n, \tag{2}$$

and an impartial vector also generates exact moments through its expectation value,

$$\mu_n = \langle\xi|H^n|\xi\rangle. \tag{3}$$

However, as we discuss below, the expectation value is an $O(N)$ operation, while taking the trace in Eq. (2) is not.

In our method, there are two key steps. The first step is to effectively approximate ξ in some manner. We do this by selecting appropriate random vectors x, and use a penalty function method to find an improved vector x^* closer to an impartial vector. The second step involves transforming the information contained in ξ into $\rho(E)$ in an $O(N)$ way. We accomplish this by viewing $\rho(E)$ as a probability distribution and use the maximum entropy principle (Maxent) to determine the best estimate from partial information (a finite set of moments). The price we pay for this increased efficiency is that we do not compute the exact values of individual eigenvalues, but rather obtain an accurate, continuous representation of the electronic DOS.

Skilling [9] was the first to note the possibility of extracting moment data from the operation of sparse matrices on random vectors. Silver, Roder, and Bruggeman [10] used random vectors to generate moment data and used an orthogonal polynomial fit for the DOS of the 2D 4×4 Heisenberg model. Here we greatly extend the practical value of this earlier work and investigate its appropriateness for electronic structure applications.

The first step of our technique is to approximate ξ. Let

x be an arbitrary random normalized vector in the space of H. Define the inner product $\langle x|y\rangle = \sum_i x_i^* y_i$. Now consider the following sequence ν_k, where

$$\nu_k = \langle x|H^k|x\rangle. \tag{4}$$

These objects are clearly the moments of a non-negative function $\rho_x(E)$ for any x as can be seen by inspecting Eq. (4) with x written in the eigenvector representation. Since $|\langle\psi_j|\xi\rangle|^2$ is uniformly distributed, the average of $\rho_x(E)$, $\langle\rho_x(E)\rangle$, over all random vectors x will reproduce $\rho(E)$. Similarly averaging Eq. (4) over random vectors x will reproduce the exact moments μ_n [Eq. (2)]. The ν_k are easily computed, as observed by Skilling [9] and Silver, Roder, and Bruggeman [10] for sparse H since ν_k may be computed recursively. If we define $y_k = H^k x$, then $y_{k+1} = Hy_k$, and $\nu_k = \langle x|y_k\rangle$. Thus, we may accumulate the ν_k by repeated operations of a sparse matrix on a vector. These calculations are $O(N)$.

A unique feature of our work is to use importance sampling in the process of accumulating moment data $\{\nu_k\}$ from H. As shown below, this dramatically improves the method for practical use. To implement this, we construct a penalty function of a vector x, subjecting it to three constraints (μ_0, μ_1, and μ_2),

$$P(x) = (\langle x|x\rangle - \mu_0)^2 + (\langle x|H|x\rangle - \mu_1)^2 + (\langle x|HH|x\rangle - \mu_2)^2. \tag{5}$$

The exact moments μ_1 and μ_2 are both $O(N)$ calculable, and by normalization, μ_0=1. We minimize the function [Eq. (5)] with a conjugate gradient (CG) method [11]. We find that it is straightforward to generate x^* such that $P(x^*) = 0$ from an initial random vector x. Note that such x^* is "closer" to an impartial vector ξ than an x merely chosen at random and normalized [9,10]. In practice, we select the initial components of x_i independently from the normal distribution. The larger the matrix, the fewer CG steps and the fewer the number of x^* vectors are required to obtain converged averages of the moments. It is of interest that this is true despite a well known tendency of CG to require more steps for a larger problem.

The second key step in our technique is to transform the information contained in x^* into $\rho(E)$ through moment data [Eq. (4)]. As demonstrated by a variety of workers [12–15], Maxent offers a very rapidly convergent approach to computing the density of states from its moments. Because of its information theoretic origin [8], Maxent introduces no artifacts stemming from *ad hoc* approximations. For numerical convenience, we scaled and shifted H so that the DOS has support only on (−1,1), and we used Tchebychev polynomials, T_n, instead of raw powers. It is these shifted and scaled units that are used in the figures of this paper. In practice, one can use low-order Maxent approximations to the DOS to obtain an approximate support, and modify H accordingly. Thus, no highly accurate guess is required *a priori* for the support. Given a set of moment data $\{\nu_k(x^*)\}$ from a given random vector x^*, the Maxent reconstruction is $\rho_M(E, x^*) = \exp[\sum \lambda_n(x^*)T_n(E)]$, where the Lagrange multipliers $\lambda_n(x^*)$ are determined by requiring that the Maxent DOS reproduce the input moments $\{\nu_k(x^*)\}$. See Turek [14] for a stable algorithm to solve the Maxent moment problem. After having generated a Maxent reconstruction of the DOS for a specific set of moments $\{\nu_k(x^*)\}$ for a given random vector, we average over such reconstructions for different vectors x^* to obtain $\langle\rho_{x^*}(E)\rangle$, our statistical approximation to the true DOS $\rho(E)$.

In the rest of this paper we give examples to demonstrate the method using a tight-binding model. The total energy is $E_{\rm TOT} = E_{\rm BS} + E_{\rm SR}$, where $E_{\rm BS}$ is the BSE given by a sum over occupied orbitals, and $E_{\rm SR}$ is a "short-ranged" repulsive two body potential which we do not concern ourselves with here.

We first consider perfect bulk GaAs in the 64 atom supercell using the model of Ref. [16]. This (320 × 320) matrix is very small for this method, but is easily exactly diagonalized for comparisons.

We will focus primarily on the BSE. There are three issues which relate to the convergence of the BSE. These issues are (i) the constraints incorporated through the penalty function on each random vector (importance sampling), (ii) the number $N_{\rm RV}$ of random vectors (RV) used, and (iii) the number of moments extracted from each vector which is then used in the Maxent DOS reconstruction.

We first demonstrate the effect of the constraints imposed on the random vectors [issue (i)] on the distribution of BSE we obtain. We choose $N_{\rm RV}$=1000 and use 30 moments in the Maxent reconstruction. For each random vector x, a density of states $\rho_x(E)$ for that vector is constructed, and the Fermi level $E_F(x)$ and BSE (per electron) $E_{\rm BS}(x)$ are determined as

$$N_e = 2N \int_{-1}^{E_F(x)} \rho_x(E)dE \tag{6}$$

and

$$E_{\rm BS}(x) = \frac{2N}{N_e} \int_{-1}^{E_F(x)} \rho_x(E)EdE. \tag{7}$$

Here N_e is the total number of electrons in the supercell and N is the total number of orbitals, and the factor of 2 reflects spin degeneracy.

By sampling 1000 RV's, we obtain a probability distribution $D(E_{\rm BS})$ which gives the likelihood that a single RV will give the value $E_{\rm BS}$ as its BSE. This is shown in Fig. 1. The width of $D(E_{\rm BS})$ decreases dramatically as more constraints are added. The full width at half maximum σ is empirically seen to be $\sigma_3 \approx (1/2^1)\sigma_2 \approx (1/2^3)\sigma_1$ where the subscript indicates the number of constraints (μ_i) used to construct the random vector x. The benefit of importance sampling is thus seen to be quite large.

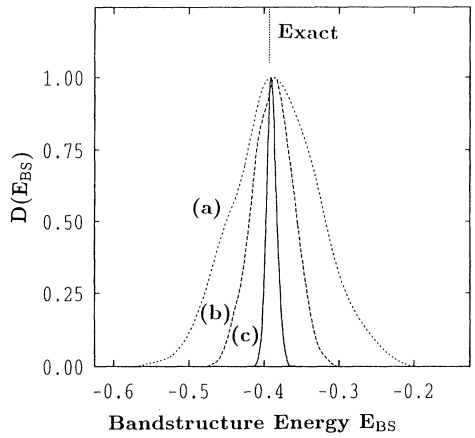

FIG. 1. Relative probability distribution $D(E_{\rm BS})$ that a random vector will yield $E_{\rm BS}$ as its band-structure energy. (In all figures, the BSE is determined from a shifted and scaled H with eigenvalues in $[-1, 1]$.) The random vectors are subjected to (a) one, (b) two, and (c) three constraints (μ_0, μ_1, and μ_2) in the penalty function.

Next we investigate issue (ii), the convergence of the BSE as a function of the number of random vectors, $N_{\rm RV}$. Here we find that the convergence is far superior if the *global* Fermi level is determined first by averaging over all RV's of the DOS, $N_e = N\int_{-1}^{E_F}\langle\rho_x(E)\rangle dE$, where $\langle\rho_x(E)\rangle$ is the average over all RV's. Similarly, $E_{\rm BS} = \int_{-1}^{E_F}\langle\rho_x(E)\rangle E dE$. The BSE using 30 moments in the Maxent reconstruction is shown in Fig. 2(a) using one constraint (μ_0 only), and in Fig. 2(b) using three constraints (μ_0, μ_1, and μ_2) in the penalty function for the random vector. The convergence to the exact result is much faster and has a much smaller fluctuation when three constraints are used. The final error in using 1000 RV's with μ_0, μ_1, and μ_2 constraints is 0.5 meV/atom—a very small error.

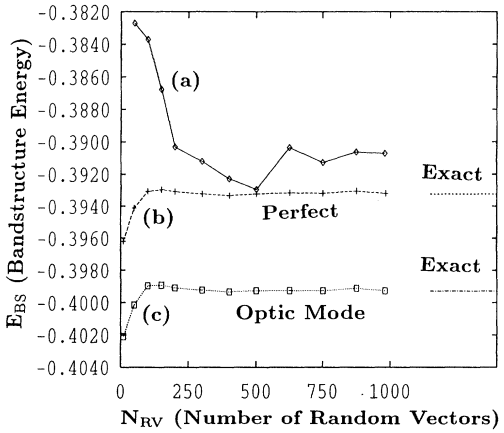

FIG. 2. Convergence of the BSE vs the number of random vectors. (a) and (b) for perfect GaAs, and (c) is GaAs with an optic phonon. Only the normalization constraint (μ_0) is imposed in (a), while three constraints are used for (b) and (c).

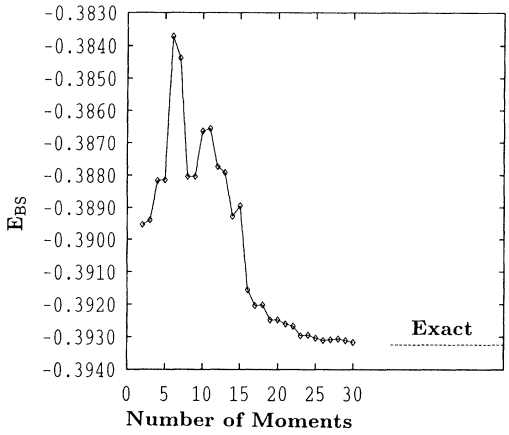

FIG. 3. Convergence of BSE for perfect GaAs vs number of exact moments used.

In practical applications, one is most interested in energy differences between one system and another. To study this, we gently disturb the perfect GaAs lattice by moving the atoms in a **k=0** optic mode displacement pattern. In the perfect crystal, the vector separating Ga and As is $d_0(111)/\sqrt{(3)}$. This vector is changed to $d_0(111)/\sqrt{(3)} - \delta(111)$ where we choose 0.05 Å for δ. The Hamiltonian matrix elements are altered according to standard direction cosine formulas and the interactions are scaled as $1/d^2$ where d is the distance between the atoms. The BSE for different numbers of random vectors is given in Fig. 2(c), and the exact result from matrix diagonalization is shown for comparison. The small difference in BSE between the perfect crystal and the crystal with an optic mode distortion is almost perfectly reproduced.

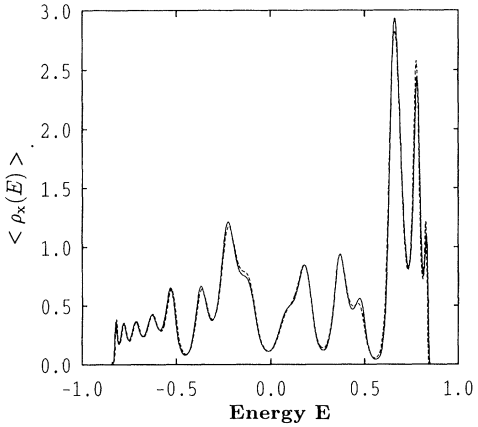

FIG. 4. Density of states for C_{720} (solid line) and C_{2160} (dotted line). The Fermi energy is near 0.25.

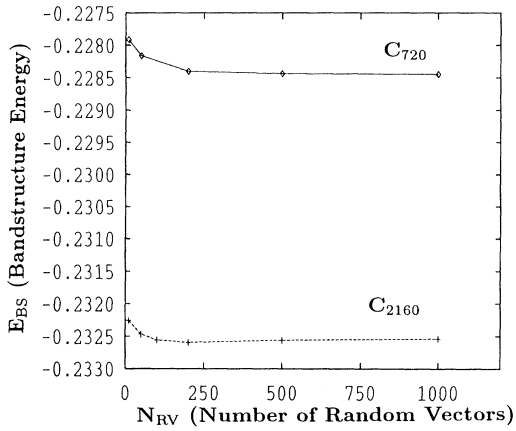

FIG. 5. Convergence of BSE vs number of importance sampled vectors N_{RV} for C_{720} (solid lines) and C_{2160} (dotted lines).

The final factor influencing the convergence of the BSE is the total number of moments used in the Maxent reconstruction [issue (iii)]. To investigate this we form an exact impartial vector ξ by diagonalizing the Hamiltonian, adding all the eigenvectors, and renormalizing the result. We then compute from this impartial vector ξ the Maxent reconstructed DOS and BSE. In Fig. 3 we show the convergence of the BSE as a function of the number of moments extracted from ξ and then used in the Maxent reconstruction. Maxent does an exceedingly good job in determining E_{BS} even for very few moments. The maximum excursion of E_{BS} from the exact results (occurring near 6 moments) is only ~ 0.009/electron (0.11 eV/electron). This should be contrasted to the energy uncertainty inherent to the RV's of $\sim$0.06 (0.8 eV/electron) by studying $D(E_{BS})$.

To explore systems of unprecedented size, we computed the electronic DOS and BSE of two large icosahedral fullerenes: C_{720} and C_{2160}. We follow exactly the procedure detailed above for GaAs, but we use the ETB model of Ref. [17] and 35 moments. In Fig. 4 we show the DOS for these fullerenes relaxed to their equilibrium geometry with an empirical potential [18]. It is perhaps not surprising that the DOS's are very close: these large fullerenes are becoming locally very similar to graphite. In Fig. 5 we show the convergence of the BSE versus the number of importance sampled vectors used. We find convergence at the level of 1 meV/electron using a few hundred vectors. These calculations required a few hours of CPU time on an IBM RS-6000 work station. If only the DOS is required, only a very few importance sampled vectors are needed. Such calculations require only a few minutes for the fullerenes described here.

In conclusion we have discussed a new *statistical* approach to the sparse, Hermitian matrix eigenvalue problem, and showed it to be useful in electronic structure calculations.

We would like to acknowledge informative discussions with S. Yang, R. M. Martin, P. Ordejon, Jose Ortega, and Gary Adams. This work was supported in part by the ONR under Contract No. ONR N00014-90-J-1304, the Federation of Advanced Materials Industries, and the Materials Research Laboratory at the University of Illinois under Contract No. NSF DMR 89 20538.

[a] Permanent address: Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701-2979.

[1] O.F. Sankey and D.J. Niklewski, Phys. Rev. B **40**, 3979 (1989); O. F. Sankey, D. A. Drabold, and G. B. Adams, Bull. Am. Phys. Soc. **36**, 924 (1991).
[2] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[3] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12 255 (1989).
[4] P. Ordejon, D. A. Drabold, M. Grumbach, and R. M. Martin (unpublished).
[5] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1992).
[6] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).
[7] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992).
[8] E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics* (Kluwer, Dordrecht, Holland, 1983).
[9] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer, Dordrecht, 1989), p. 455.
[10] R. N. Silver, H. Roder, and J. A. Bruggeman, Int. J. Mod. Phys. A (to be published).
[11] W. H. Press *et al.*, *Numerical Recipes, The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1986).
[12] L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
[13] A. E. Carlsson and P. A. Fedders, Phys. Rev. B **34**, 3567 (1986).
[14] I. Turek, J. Phys. C **21**, 3251 (1988).
[15] D. A. Drabold and G. L. Jones, J. Phys. A **24**, 4705 (1991).
[16] P. Vogl *et al.*, J. Phys. Chem. Solids **44**, 365 (1983).
[17] C. H. Xu *et al.*, J. Phys. Condens. Matter **4**, 6047 (1992).
[18] J. Tersoff, Phys. Rev. Lett. **56**, 632 (1986).