Calculation of densities of states and spectral functions by Chebyshev recursion and maximum entropy

R. N. Silver and H. Röder
MS B262 Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
(Received 26 March 1997)

We present an efficient algorithm for calculating spectral properties of large sparse Hamiltonian matrices such as densities of states and spectral functions. The combination of Chebyshev recursion and maximum entropy achieves high-energy resolution without significant roundoff error, machine precision, or numerical instability limitations. If controlled statistical or systematic errors are acceptable, CPU and memory requirements scale linearly in the number of states. The inference of spectral properties from moments is much better conditioned for Chebyshev moments than for power moments. We adapt concepts from the kernel polynomial method, a linear Chebyshev approximation with optimized Gibbs damping, to control the accuracy of Fourier integrals of positive nonanalytic functions. We compare the performance of kernel polynomial and maximum entropy algorithms for an electronic structure example. [S1063-651X(97)10309-9]

PACS number(s): 02.70.−c, 71.15.−m, 71.20.−b

I. INTRODUCTION

Many computational physics problems involve very large sparse Hamiltonian matrices. If N is the number of states, finding all eigenvectors and eigenvalues requires CPU time scaling as N^3 and memory scaling as N^2. For individual eigenstates the preferred method is Lanczos diagonalization, which uses only matrix-vector-multiply operations and requires CPU and memory scaling as N. Densities of states and spectral functions for finite-dimensional Hamiltonians are sums of δ functions with positive amplitudes. In the thermodynamic limit of relevance to condensed-matter physics, these can extrapolate to singular structures such as isolated states, band edges, and Van Hove singularities. New linear scaling methods are needed for calculating such spectral properties which involve many eigenstates, and for quantities derived from them including thermodynamics, total energies for electronic structure and forces for molecular dynamics. Limited energy resolution and statistical accuracy are often acceptable provided uncertainties can be quantified.

The *maximum entropy method* [1–3] is a popular approach: maximize the information theoretic relative entropy of the spectrum subject to data constraints. The input data are usually power moments. Maximum entropy spectra are strictly positive. Maximum entropy spectra are the solution of a convex nonlinear optimization problem. Maximum entropy always yields broadened representations of the true spectra. The resolution function is nonuniform and unknown, with some parts of a spectrum converging more rapidly than others as the number of moments increases. Occasionally, maximum entropy yields a spurious structure; for example, it can ''ring'' in smooth regions of a spectrum near to a Van Hove singularity. Nonanalytic features are better approximated at higher-energy resolution, which is achieved in maximum entropy by fitting more moments. However, as moment order increases, the calculation of power moments is more sensitive to machine precision limits, and the optimization problem is more ill conditioned. Maximum entropy is difficult to implement for more than about 50 power moments.

The *kernel polynomial method* [4–6] is much easier to implement for high-energy resolution applications. It is a linear Chebyshev approximation to spectra using Chebyshev moment data. Abrupt truncation of Chebyshev series results in the Gibbs phenomenon: a lack of uniform convergence at nonanalytic (or singular) features in spectra. Instead, the moments of the kernel polynomial approximation are the data multiplied by Gibbs damping factors, which are chosen to ensure positive spectra with the highest-energy resolution. A kernel polynomial approximation is a convolution of the true spectrum with a known positive kernel polynomial function. It can be rapidly evaluated by fast Fourier transform without nonlinear optimization. In contrast to Lanczos methods, Chebyshev recursion is numerically stable without accumulation of roundoff error; thus there is no need for computationally expensive reorthogonalization [7]. For sparse Hamiltonians, the computational cost for generating Chebyshev moment data is linear scaling if controlled systematic or statistical errors are acceptable [4,8]. Chebyshev approximations have been applied recently to densities of states and spectral functions in diverse areas of condensed matter physics including the Heisenberg antiferromagnet [4], the Holstein t-J model [9], the dielectric constants of quantum dots [10], linear scaling algorithms for tight-binding molecular dynamics [11,8], nonorthogonal electronic structure [12], and so on. Chebyshev approximations have also been developed independently for scattering problems in quantum chemistry [13–18].

A comparison of the maximum entropy and kernel polynomial methods reveals advantages for each. (A comparison of Lanczos methods with kernel polynomial methods may be found in Ref. [5].) The maximum entropy method achieves significantly higher-energy resolution, requiring calculation of 4–10 times fewer moments for typical applications. However, the nonlinear optimization problem can be difficult to solve, the resolution is nonuniform and unknown, and there is a risk of artifacts. The kernel polynomial method has significantly poorer energy resolution. However, nonlinear optimization is not needed, the resolution function is uniform and known, and there is no risk of artifacts.

Our experience is that, in most cases, the computational cost of generating moment data limits the ability to do calculations. The practical necessity to use computational resources in the most efficient way motivates our development of a new maximum entropy algorithm based on Chebyshev moment data. Chebyshev moments have several advantages over power moments for a maximum entropy algorithm:

(i) *Machine precision limitations.* In a power moment information in low digits past the decimal point is redundant with information in low-order moments. New information is contained in higher digits whose cardinality increases with the order of the moment. Thus machine precision puts a limit on how many power moments are useful to calculate. In contrast, there is no redundancy in moments constructed from orthogonal polynomials [19], and machine precision is not limiting.

(ii) *Conditioning.* The ill-posed inverse problem of inferring a spectrum from a limited number of Chebyshev moments is much better conditioned than from the same number of power moments. In particular, the Hessian for maximum entropy optimization has a much flatter eigenvalue spectrum for Chebyshev moments than for power moments.

(iii) *Computational efficiency and accuracy.* A simple coordinate transformation converts a Chebyshev series to a Fourier series, which enables use of fast Fourier transform methods.

In summary, the combination of Chebyshev recursion and maximum entropy should provide an efficient stable algorithm capable of reaching arbitrarily high-energy resolution.

There is an extensive literature [20–22] on convex nonlinear optimization applied to maximum entropy. For our applications, we find the principal new algorithmic difficulty to be control of the numerical accuracy of Fourier integrals when the true spectra have singular (or nonanalytic) features such as δ functions. We adapt concepts from the kernel polynomial method and the Shannon sampling theorem to solve this numerical accuracy problem. The resulting algorithm has no difficulty handling thousands of Chebyshev moments, if necessary.

In Sec. II we briefly review methods for generating Chebyshev moment data. In Sec. III we describe the kernel polynomial method. In Sec. IV we present our maximum entropy algorithm. In Sec. V we illustrate the method using an electronic structure example, comparing the performance of the maximum entropy and kernel polynomial methods. In Sec. VI we conclude.

II. GENERATION OF CHEBYSHEV MOMENT DATA

Consider a density of states as representative of the spectral properties of interest. The first step is to scale the Hamiltonian, $\mathbf{H}=a\mathbf{X}+b$ such that all eigenvalues x_n of $\mathbf{X}$ satisfy $-1\leqslant x_n\leqslant +1$. These end points are rapidly computed, for example, by Lanczos methods using the same matrix-vector-multiply operations required for generating Chebyshev moments. The only difference between the kernel polynomial method and the maximum entropy method is, in order to minimize endpoint corrections in fast Fourier transform evaluation of Fourier integrals, we recommend placing all x_n well inside -1 and $+1$, for example, $-0.99\leqslant x_n\leqslant +0.99$. This point will be discussed further in Sec. IV.

The density of states is then

$$D(x)=\frac{1}{N}\sum_{n=1}^{N}\delta(x-x_n). \tag{1}$$

Data about $D(x)$ consist of Chebyshev moments

$$\hat{\mu}_m=\mathrm{Tr}\{T_m(\mathbf{X})\}=\int_{-1}^{1}T_m(x)D(x)dx. \tag{2}$$

We use the notation $\hat{\mu}_m$ for a datum on a moment, even if this estimate is approximate. We calculate moments using the Chebyshev recursion relation

$$T_{m+1}(\mathbf{X})=2\mathbf{X}T_m(\mathbf{X})-T_{m-1}(\mathbf{X}), \tag{3}$$

which requires the same optimized matrix-vector-multiply algorithm used in Lanczos methods. Unlike Lanczos recursion, Chebyshev recursions are numerically stable to arbitrarily large numbers of recursions. We use rules for multiplying Chebyshev polynomials, e.g., $T_{2m}=2T_mT_m-1$, so that only $M/2$ matrix-vector multiplies are needed to generate M moments.

Exact evaluation of M moments requires CPU time proportional to $O(N^2M/2)$ for sparse matrices. Generate $T_m(\mathbf{X})|i\rangle$ for each basis state $|i\rangle$. The estimator for Chebyshev moments is then

$$\hat{\mu}_{2m}=\frac{2}{N}\sum_{i=1}^{N}\langle i|T_m(\mathbf{X})T_m(\mathbf{X})|i\rangle-1. \tag{4}$$

There is a similar expression for m odd.

Stochastic evaluation [4] requires CPU time scaling as $O(NMN_r)$. The estimator for Chebyshev moments is

$$\hat{\mu}_m\approx\frac{1}{N_r}\sum_{r}\langle r|T_m(\mathbf{X})|r\rangle, \tag{5}$$

where the $|r\rangle$ are N_r Gaussian random vectors. Such data have a statistical variance proportional to $(NN_r)^{-1}$, which may be expressed directly in terms of the density of states. An estimation of statistical errors was described by Silver and Roder [4]. More sophisticated choices of random vector appear to reduce statistical variance [3], but they introduce unwanted statistical bias and make error estimation difficult.

Local truncation evaluation of moments requires CPU time scaling as $O(NMJ)$. Here, moments are calculated with a locally truncated Hamiltonian $\mathbf{H}_i$, where J is the number of states included in the truncation range. The estimator for Chebyshev moments is

$$\hat{\mu}_m\approx\sum_{i}\langle i|T_m(\mathbf{X}_i)|i\rangle. \tag{6}$$

This generates data with a systematic error determined by the truncation range. ''Logical'' truncation [8] appears to converge more rapidly and smoothly than ''physical'' truncation schemes [11,23]. Local truncation may be applicable if the density matrix has only local off-diagonal elements, as in tight-binding Hamiltonians for insulators. Exact moment de-

rivatives needed to estimate forces for molecular dynamics can be calculated from a Chebyshev derivative formula.

CPU time and memory limit the number of moments M and their statistical and systematic errors. Fortunately, both stochastic and local truncation methods provide means to estimate and control errors.

III. KERNEL POLYNOMIAL APPROXIMATION

The kernel polynomial method has two roles in this paper. First, it is a method to estimate spectra from Chebyshev moment data. Second, it provides our approach to control numerical accuracy in the evaluation of Fourier integrals of nonanalytic spectra in our maximum entropy algorithm.

An exact Chebyshev moment expansion of the density of states is

$$D(x)=\frac{1}{\pi\sqrt{1-x^2}}\left[\mu_0+2\sum_{m=1}^{\infty}\mu_m T_m(x)\right]. \tag{7}$$

The kernel polynomial method truncates Eq. (7) at M moments, introduces a factor g_m^M to damp Gibbs phenomenon, and substitutes (possibly inaccurate) data $\hat{\mu}_m$ for the moments. The kernel polynomial approximation to a density of states is then

$$D_K(x)=\frac{1}{\pi\sqrt{1-x^2}}\left[1+2\sum_{m=1}^{M}\hat{\mu}_m g_m^M T_m(x)\right]. \tag{8}$$

Let $\phi\equiv\cos^{-1}(x)$. Then $T_m(x)=\cos(m\phi)$. Define $D(\phi)\equiv\sin(\phi)D(X)$. The Chebyshev moments are then Fourier integrals

$$\mu_m=\int_{-1}^{1}T_m(x)D(x)dx=\int_0^{\pi}\cos(m\phi)D(\phi)d\phi. \tag{9}$$

If the data are exact, $D_K(\phi)$ can be represented as both a simple convolution and a truncated Fourier series,

$$D_K(\phi)=\int_0^{2\pi}\delta_K(\phi-\phi_o)D(\phi_o)d\phi_o,$$

$$\delta_K(\phi)=\frac{1}{2\pi}\left[g_0+2\sum_{m=1}^{M}g_m^M\cos(m\phi)\right]. \tag{10}$$

For later purposes, we emphasize that the Fourier integrals of $D_K(\phi)$ are $\mu_m=\hat{\mu}_m g_m^M$. Thus $D_K(\phi)$ does not fit the moment data. If the data are inexact, corresponding random variables should be added to Eq. (10).

The *kernel* $\delta_K(\phi)$ is a 2π-periodic polynomial approximation to a Dirac delta function, analogous to the resolution function of a spectrometer. Resolution is uniform in ϕ with width $\Delta\phi\propto M^{-1}$. If $g_m^M=1$, at large $|\phi|$ the kernel is oscillatory with period $\Delta\phi=\pi/M$ within an envelope function decreasing slowly as $1/\phi^2$. The result is the Gibbs phenomenon: a lack of uniform pointwise convergence of the cosine series at singular (or nonanalytic) structures in the density of states. An optimal g_m^M that minimizes the Gibbs phenomenon may be derived [5] by requiring the kernel to be a strictly positive normalized polynomial of degree M with minimal variance in ϕ. The result is

$$g_m^M=\sum_{\nu=0}^{M-m}a_\nu a_{\nu+m}, \tag{11}$$

where

$$a_\nu=\frac{U_\nu(\lambda)}{\sqrt{\Sigma_{\nu=0}^{M}U_\nu^2(\lambda)}}. \tag{12}$$

and

$$U_\nu(\lambda)=\frac{\sin[(\nu+1)\phi_\lambda]}{\sin(\phi_\lambda)},\quad \cos(\phi_\lambda)=\lambda. \tag{13}$$

Here U_ν are Chebyshev polynomials of the second kind, and $\phi_\lambda\equiv\pi/(M+2)$. g_m^M decrease smoothly and monotonically from 1 to 0 as m increases from 0 to M. This kernel was originally derived by minimizing the uniform norm [24]. Its envelope function decreases exponentially at large $|\phi|$.

The kernel polynomial method is also applicable to spectral functions [5],

$$A(\omega)=\lim_{\eta\to 0+}\frac{1}{\pi}\mathrm{Im}\left\{\left\langle\Psi_0\middle|\mathbf{O}^\dagger\frac{1}{\omega-\mathbf{H}-i\eta}\mathbf{O}\middle|\Psi_0\right\rangle\right\}, \tag{14}$$

where $\mathbf{O}$ is the appropriate Hermitian operator. The corresponding Chebyshev moments have the form $\hat{\mu}_m^{\mathbf{O}}=\langle\Psi_0|\mathbf{O}^\dagger T_m(\mathbf{X})\mathbf{O}|\Psi_0\rangle$. Silver *et al.* [5] compared the performance of the kernel polynomial method to Lanczos methods for spectral functions.

Applications of kernel polynomial approximations to thermodynamics use a rapidly converging Fourier-Bessel expansion of the partition function [4],

$$Z=e^{-\beta b}\left[I_0(\beta a)+2\sum_{m=1}^{\infty}I_m(\beta a)\hat{\mu}_m\right]. \tag{15}$$

$I_m(\beta a)$ are modified Bessel functions. The partition function involves integral rather than pointwise convergence, so the optimal choice is no Gibbs damping, $g_m^M=1$.

Our maximum entropy algorithm uses the kernel polynomial approximation in an interesting way. We employ fast Fourier transforms to evaluate Fourier integrals of $D(\phi)$ by summing over $L+1$ pixels equally spaced in ϕ, corresponding to sampling $D(\phi)$ at the $L+1$ zeros of $\cos[(L+1)\phi]$. The Shannon sampling theorem says that the only function which can be exactly evaluated by this procedure is a band limited function, a finite Lth-order cosine series. So, in fact, our algorithm exactly evaluates Fourier integrals of a Chebyshev approximation. The procedures in Sec. II generate moments of a function consisting of a sum of N δ functions with positive amplitudes, equivalent to an infinite-order Chebyshev series. Inasmuch as, typically, $N\gg L$, we should not expect to resolve all states. Our maximum entropy algorithm requires moments of an Lth-order positive Chebyshev approximation. The only Chebyshev approximation that can

satisfy the positivity constraint required by maximum entropy is a kernel polynomial approximation. Moments of the kernel polynomial approximation are related to moments of the true function by Gibbs damping factors. This subtle difference becomes important because we demand high numerical accuracy. We elaborate on these points in Sec. IV.

IV. MAXIMUM ENTROPY ALGORITHM

This section presents our maximum entropy algorithm. Although it may be regarded as an adaptation of previous maximum entropy algorithms [20–22], our problem has issues of numerical accuracy for $D(x)$ that contain singular (nonanalytic) structures such as δ functions.

Consider the case where the data are subject to Gaussian uncertainties,

$$\hat{\mu}_m = \mu_m + \eta_m, \quad \mathbf{E}\eta_m = 0, \quad \mathbf{E}\eta_m\eta_{m'} = \sigma_m^2\delta_{mm'}. \quad (16)$$

Here η is a random variable, and $\mathbf{E}$ denotes the statistical expectation value of the random variable following it. The χ^2 statistic for measuring quality of fit is then

$$\chi^2 = \sum_{m=0}^{M} \left(\frac{\hat{\mu}_m - \mu_m}{\sigma_m} \right)^2. \quad (17)$$

In the case of exact moment data, set σ_m to the numerical precision required, which can be very small. The $m=0$ term is included to constrain normalization, $\hat{\mu}_0 = 1$. Taking the limit $\sigma_0 \to 0$ strictly enforces normalization. In our applications using 32-bit computers, sixth or seventh digits past the decimal point of moments often contain important information. We typically drop χ^2 by 12–14 orders of magnitude below its starting values during the course of converging to a maximum entropy solution. Such high numerical accuracy can be critical to avoid spurious artifacts, and to yield the correct physics.

Therefore, very careful attention to numerical accuracy is required in evaluating Fourier integrals, Eq. (9). To have an efficient maximum entropy algorithm, we evaluate Fourier integrals by fast Fourier transform. This equals a sum over equally spaced points in ϕ,

$$\mu_m \approx \sum_{l=0}^{L} \cos(m\phi_l) D(\phi_l) \Delta\phi. \quad (18)$$

The $L+1$ ϕ_l satisfy $\cos[(L+1)\phi_l] = 0$, where $0 \leq l \leq L$. The Shannon sampling theorem says this approximation becomes exact only if $D(\phi)$ is a band limited function of degree L. But the exact $D(\phi)$ in our applications are sums of δ functions with positive amplitudes, so evaluation of Fourier integrals by fast Fourier transform with a finite number of pixels L is not exact. The maximum entropy $D(\phi)$ also correspond to infinite-order Fourier series, so evaluation of their Fourier integrals by this procedure is not exact either.

Our strategy to minimize numerical errors is to minimize high-frequency components of the maximum entropy solution. The goal of our algorithm is to find a kernel polynomial approximation for $M \times K$ moments instead of M moments, where K is some integer. Maximum entropy provides the criterion for extrapolating the moment series. But the moments of the kernel polynomial approximation of degree $M \times K$ are $\hat{\mu}_m g_m^{M\times K}$, so these are what we should use as data in our χ^2 criterion. Modifying the data in this way ensures that our target spectrum is positive and satisfies the Hausdorff conditions for the existence of a maximum entropy solution. Choosing K in our algorithm is equivalent to choosing the desired energy resolution. If our maximum entropy solution was in fact equivalent to a higher-resolution kernel polynomial approximation, choosing the number of pixels $L = M \times K$ would yield exact Fourier integrals. But inasmuch as our maximum entropy solution is not exactly band limited, we choose the number of pixels some integer factor I larger than $M \times K$, i.e., $L = M \times K \times I$. The extra factor I further reduces numerical errors.

Increasing L to improve numerical accuracy must be balanced against increased computational resources required for the fast Fourier transform. CPU time scales as $O(L \ln L)$ and memory scales as $O(L)$. Typically, we find maximum entropy improves resolution by factors of 4–10 over the kernel polynomial method, so most of the gain is obtained by choosing $4 \leqslant K \leqslant 10$. The corresponding $g_m^{M\times K}$'s for $0 \leqslant m \leqslant M$ are only slightly smaller than 1, but this difference is enough to determine whether our algorithm converges to the stopping criterion or stalls at high χ^2. Without the Gibbs damping correction, convergence may be very nonuniform, and in some regions approach an energy resolution that cannot be described with the number of pixels chosen. With the Gibbs damping correction, the dynamic range of the resolution improvement is limited, and can be handled with the number of pixels chosen. We also typically find the choice $I \geqslant 4$ to be sufficient to fit Chebyshev moments to seven-digit accuracy.

End-point corrections are another concern in evaluating Fourier integrals. They are often essential to obtain reasonable convergence for high-order moments. Sophisticated approaches to this problem have been developed based on interpolation schemes [25]. However, within our algorithm we have the option of minimizing end-point corrections at the outset, by forcing the spectrum to be close to zero near endpoints $\phi = 0$ and $\phi = \pi$. The easy way to force $D(\phi)$ toward zero at end points is to scale the Hamiltonian $\mathbf{H} = a\mathbf{X} + b$ such that all eigenvalues x_n of $\mathbf{X}$ lie between, say, -0.99 and $+0.99$, rather than -1 and $+1$. This change has only a 1% impact on ϕ resolution, but it avoids the end-point corrections and the bias of interpolation schemes. This justifies the recommendations made in Sec. II for scaling the Hamiltonian. We often find this superficially small correction to be critical to achieving high accuracy fits without stalling.

Our optimization problem is to maximize the relative entropy

$$S \equiv \int_0^{\pi} \left[D(\phi) - D_o(\phi) - D(\phi) \ln\left(\frac{D(\phi)}{D_o(\phi)} \right) \right] d\phi \quad (19)$$

subject to data constraints. S is strictly negative and equals zero only when $D = D_o$. Here $D_o(\phi)$ is a default model for the density of states in the absence of data. We obtain faster convergence with less risk of spurious artifacts by using prior knowledge to choose a default model closer to the final answer. In the absence of prior knowledge, we usually use the kernel polynomial approximation as the default model,

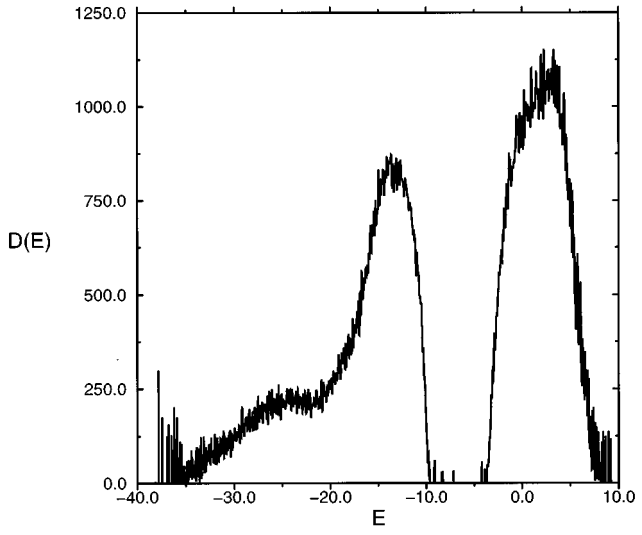

FIG. 1. Density of states of an orthogonal tight-binding model for amorphous diamond calculated with the maximum entropy method using 1024 exact Chebyshev moments.

inasmuch as the motivation for maximum entropy is to improve energy resolution beyond the kernel polynomial method.

More specifically, the *primal optimization problem* is to maximize

$$Q_p \equiv S - \frac{\chi^2}{2\alpha} \tag{20}$$

as a function of a continuous variable $D(\phi)$. The statistical regularization parameter α sets a balance between the fit, measured by χ^2, and an information measure $-S$ of distance between the inferred $D(\phi)$ and the default model $D_o(\phi)$. Alternatively, we regard $1/\alpha$ as a Lagrange multiplier enforcing a constraint on χ^2.

Our algorithm consists of three nested loops: the outer loop iteratively decreases α starting from a high value, until a stopping criterion is reached; the middle loop for each α iteratively solves for the $D(\phi)$ that maximizes Q_p; the inner loop at each α and $D(\phi)$ solves for the next update of $D(\phi)$ using linear equation solvers such as conjugate gradients. We discuss each of these loops in turn.

The outer loop typically starts at large $\alpha_1 \approx \chi_o^2$, the χ^2 of the default model D_o, then progress geometrically down in α, e.g. $\alpha_{k+1} = \alpha_k/2$. The corresponding χ^2 decreases and the information $-S$ increases. If the middle loop is unstable, as measured by a significant increase in χ^2, go back to conditions at the start of this loop, halve the step down in α, and iterate until stability is reached. Popular stopping criteria for α are $\chi^2 = M$ and $\chi^2 - 2\alpha S = M$, although many other criteria are discussed in the literature. Once the stopping criterion is passed, perform a golden search for the optimal α. We often find that the information $-S$ saturates at an α-independent value as α decreases, so that the outer loop may be stopped earlier.

In principal, the middle loop solves the primal optimization problem

$$\frac{\delta Q_p}{\delta D(\phi)} = -\ln\left(\frac{D(\phi)}{D_o(\phi)}\right) + \sum_{m=0}^{M} \frac{\hat{\mu}_m - \mu_m}{\alpha \sigma_m^2} \cos(m\phi) = 0. \tag{21}$$

This is a convex optimization problem having a unique answer. Unfortunately, this approach is difficult because this problem statement is written in terms of a continuous positive variable $D(\phi)$ which varies typically by many orders of magnitude. In practice, the number of variables to optimize would be the number of pixels L chosen, which for reasons of numerical accuracy we discussed earlier is usually a quite large number. However, a *dual optimization problem* [20] solves the same problem, is more stable numerically, and is easier to implement than the primal problem. It requires only $M \ll L$ parameters $\vec{\lambda}$ defined by

$$\lambda_m \equiv \frac{\mu_m - \hat{\mu}_m}{\alpha \sigma_m^2}. \tag{22}$$

Then the maximum entropy $D(\phi)$ satisfying Eq. (21) is

$$D(\phi) = D_o(\phi) \exp\left(-\sum_{m=0}^{M} \lambda_m \cos(m\phi) \right). \tag{23}$$

This form is also obtained by maximizing entropy subject to Lagrange contraints on moments with Lagrange multipliers $\vec{\lambda}$. The dual optimization problem is to maximize

$$Q_d \equiv \ln\left(\int_0^\pi D(\phi) d\phi \right) + \sum_{m=0}^{M} \left[\hat{\mu}_m \lambda_m + \frac{\alpha \sigma_m^2 \lambda_m^2}{2} \right] \tag{24}$$

as a function of the set of λ_m. Mapping onto the dual space of Lagrange multipliers reduces an infinite-dimensional optimization problem to a feasible finite-dimensional problem. Away from the maximum, define

$$\xi_m \equiv \frac{\partial Q_d}{\partial \lambda_m} = \hat{\mu}_m - \mu_m + \alpha \sigma_m^2 \lambda_m . \tag{25}$$

Equation (22) is satisfied when $\xi_m = 0$.

The middle loop of our algorithm solves Eq. (22) by Newton-Raphson iteration. Beginning with some starting $\vec{\lambda}^0$, the $(n+1)$st step is

$$\mathcal{H}_n(\vec{\lambda}^{n+1} - \vec{\lambda}^n) = \vec{\xi}^n. \tag{26}$$

Here $\mathcal{H}$ is the Hessian of the dual problem, which is a positive definite $M \times M$ matrix and a simple function of the moments,

$$\mathcal{H}_{mm'} \equiv \frac{\partial^2 Q_d}{\partial \lambda_m \partial \lambda_{m'}'} = \frac{\mu_{m+m'} + \mu_{|m-m'|}}{2} + \alpha \sigma_m^2 \delta_{mm'} . \tag{27}$$

Then

$$Q_d = Q_p + \sum_{m=0}^{M} \frac{\xi_m^2}{2\alpha\sigma_m^2}. \tag{28}$$

We have $Q_d > Q_p$ in Eq. (28), and as the iteration proceeds Q_d (Q_p) approaches Q_∞ from above (below). Hence con-

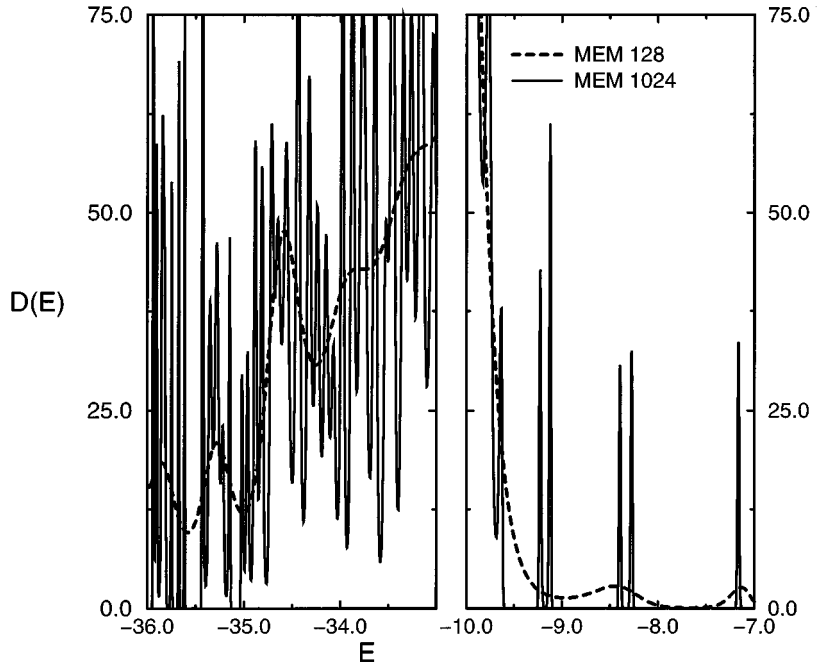

FIG. 2. Comparison of portions of the density of states in Fig. 1 calculated by maximum entropy for 128 and 1024 moments.

verging bounds at the nth iteration are $Q_d^n \geqslant Q^\infty \geqslant Q_p^n$, where $Q^\infty \equiv \lim_{n\to\infty}\{Q_d^n, Q_p^n\}$. These bounds provide stopping criteria for the middle loop. We typically stop at $Q_d = 1.02 Q_p$.

The inner loop is the solution of Eq. (26) by linear equation solvers such as conjugate gradients. This task can be handled by standard packages such as EISPACK. An advantage of Chebyshev moments is that the spectrum of eigenvalues of the Hessian in Eq. (27) is almost flat, whereas the eigenvalue spectrum for power moments is steeply decreasing. Thus this problem is well conditioned for Chebyshev moments, but can easily become ill conditioned for power moments as M increases.

We find the CPU time required by our algorithm scales as $O(ML\ \ln(L))$. But it remains negligible compared to the CPU time required to generate moment data for most problems. As we stated before, use of the maximum entropy method usually cuts overall CPU requirements by at least a factor of 4 over the kernel polynomial method. Isolated features in spectra, such as individual states and band edges, usually converge much faster, up to a factor of 10 or more.

A few words should be said about data generated by stochastic methods in Eq. (5). Calculation of the covariance matrix **C** for such data is described in an earlier paper [4]. Its structure is the same as the Hessian in Eq. (27). The appropriate generalized χ^2 statistic is

$$\chi^2 = (\hat{\mu} - \mu)^\dagger \mathbf{C}^{-1} (\hat{\mu} - \mu). \tag{29}$$

There is a cancellation, leading to an effective Hessian essentially proportional to a unit matrix and independent of D. This property further facilitates finding maximum entropy solutions for data generated by the stochastic method.

Energy derivatives needed for molecular dynamics can be derived for maximum entropy using the same expressions for exact derivatives of moments. The statistical error for stochastic methods using Gaussian random vectors can easily be accommodated, because the covariance of moments is proportional to the Hessian. Details of these extensions will be presented elsewhere.

V. AN EXAMPLE

We illustrate the performance of our maximum entropy method (MEM) algorithm with the example of an orthogonal tight-binding calculation of the density of states of amorphous diamond by Dong and Drabold [26]. The purpose of their study was to examine the localized to extended transition in band-tail states in an amorphous semiconductor. We refer readers to their paper for a discussion of the physics. Here we are only interested in a comparison of methods for estimating densities of states. We note that their maximum entropy calculations used Chebyshev moments, but were limited to about 90 moments because of the difficulties in implementing a maximum entropy algorithm. With the algorithmic improvements presented in this paper, we demonstrate that maximum entropy calculations with thousands of moments are now feasible.

The Hamiltonian considered has 16 384 states. We show maximum entropy results for the goal of improving resolution by a factor of 8 over the kernel polynomial method. Hence we choose $K = 8$. We set $I = 4$ to minimize numerical errors in the evaluation of Fourier integrals. We use up to $M = 1024$ exact Chebyshev moments. Hence the number of pixels $L = M \times K \times I \leqslant 32\,768$.

Figure 1 shows the full density of states obtained using maximum entropy for $M = 1024$. Note the optical gap in the density of states. Figure 2 shows a comparison of maximum entropy results for $M = 128$ and 1024 for two portions of this spectrum. On the left is a region where the states are dense, and on the right is a region inside the gap where the states are sparse. $M = 128$ is a factor of 2 or 3 larger than the maximum number of moments that can be handled by maximum entropy algorithms using power moments. Our algo-

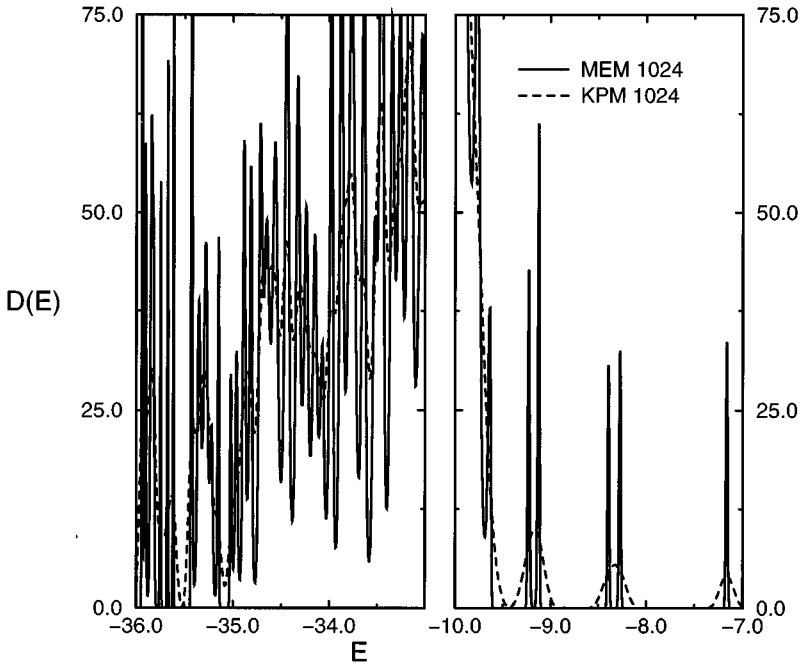

FIG. 3. Comparisons of portions of the density of states in Fig. 1 calculated by the kernel polynomial method (KPM) for 1024 moments and by the maximum entropy method (MEM) for 1024 moments.

rithm clearly is able to handle many more moments and to achieve much higher-energy resolution. Figure 3 shows the same comparison between kernel polynomial method and maximum entropy with $M=1024$ moments. Again maximum entropy provides a dramatic improvement in energy resolution over the kernel polynomial method.

VI. CONCLUSION

We have described an efficient algorithm to calculate densities of states and spectral functions of large sparse Hamiltonians using Chebyshev recursion and maximum entropy. It is a nonlinear extension of the kernel polynomial method. It is capable of handling large numbers of moments and nonanalytic (singular) structures in densities of states and spectral functions to achieve high-energy resolution. The choice of Chebyshev recursion overcomes problems of machine precision and ill-conditioning found in maximum entropy algorithms for power moments. It also circumvents the accumulation of numerical roundoff errors in Lanczos recursion. It controls the numerical accuracy of Fourier integrals by multiplying the moment data by Gibbs damping factors appropriate to the number of pixels chosen. It avoids endpoint corrections to the fast Fourier transforms by appropriate scaling of the Hamiltonian. Our algorithm achieves significant resolution gains over the kernel polynomial method for practical physics examples. The CPU time we need to find the maximum entropy solution scales approximately as the number of pixels times the number of moments. For most applications, this time is small compared to the CPU time we need to generate the Chebyshev moment data. Overall CPU time can scale linearly in the number of states if controlled statistical or systematic errors are acceptable.

The resulting maximum entropy algorithm has some tuning parameters to control the balance between numerical accuracy, convergence, energy resolution, and computational resources. A FORTRAN 77 program implementing our algorithms for the Kernel polynomial method and MEM is available from the authors by sending e-mail to rns@loke.lanl.gov. It uses publically available libraries including DFFTPACK for fast Fourier transforms and EISPACK for solving systems of linear equations.

ACKNOWLEDGMENTS

Research was supported by the U.S. Department of Energy. We thank D. Drabold and J. J. Dong for their kind permission to use the example reported here. We thank D. Drabold and J. Kress for helpful comments on the manuscript.

[1] L. R. Mead and N. Papanicolaou, J. Math. Phys. (N.Y.) **25**, 2404 (1984).
[2] R. H. Brown and A. E. Carlsson, Phys. Rev. B **32**, 6125 (1985).
[3] D. A. Drabold and O. F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).
[4] R. N. Silver and H. Röder, Int. J. Mod. Phys. C **5**, 735 (1994).
[5] R. N. Silver, H. Röder, A. F. Voter, and J. D. Kress, J. Comput. Phys. **124**, 115 (1996).
[6] L. W. Wang, Phys. Rev. B **49**, 10 154 (1994).
[7] B. N. Parlett, *The Symmetric Eigenvalue Problem* (Prentice-Hall, New York, 1976).
[8] A. F. Voter, J. D. Kress, and R. N. Silver, Phys. Rev. B **53**, 12 733 (1996).

[9] H. Roder, H. Fehshke, and R. N. Silver, Europhys. Lett. **28**, 257 (1994).
[10] L. W. Wang and A. Zunger, Phys. Rev. Lett. **73**, 1039 (1994).
[11] S. Goedecker and L. Colombo, Phys. Rev. Lett. **73**, 122 (1994).
[12] H. Roder, R. N. Silver, D. Drabold, and J. J. Dong, Phys. Rev. B (to be published).
[13] R. Kosloff, J. Chem. Phys. **92**, 2087 (1988).
[14] Y. Huang, W. Zhu, D. J. Kouri, and D. K. Hoffman, Chem. Phys. Lett. **206**, 96 (1993).
[15] W. Zhu, Y. Huang, D. J. Kouri, C. Chandler, and D. K. Hoffman, Chem. Phys. Lett. **217**, 73 (1994).
[16] D. J. Kouri, W. Zhu, G. A. Parker, and D. K. Hoffman, Chem. Phys. Lett. **238**, 395 (1995).
[17] G. A. Parker, W. Zhu, Y. Huang, D. K. Hoffman, and D. J. Kouri, Comput. Phys. Commun. **96**, 27 (1996).
[18] D. J. Kouri, Y. Huang, and D. K. Hoffman, in *Dynamics of Molecules and Chemical Reactions*, edited by R. E. Wyatt and J. Z. H. Zhang (Marcel Dekker, New York, 1996), p. 307.
[19] J. C. Wheeler, M. G. Prais, and C. Blumstein, Phys. Rev. B **10**, 2429 (1974).
[20] C. Auyeng and R. M. Mersereau, in *Digital Image Reconstruction*, edited by A. K. Katsaggelos (Springer, New York, 1991), pp. 21–55.
[21] J. Skilling, in *Physics & Probability: Essays in Honor of Edwin T. Jaynes*, edited by W. T. Grandy, Jr. and P. W. Milonni (Cambridge University Press, Cambridge, 1993), pp. 207–221.
[22] I. Turek, J. Phys. C **21**, 3251 (1988).
[23] S. Goedecker and M. Teter, Phys. Rev. B **51**, 9455 (1995).
[24] D. Jackson, *Theory of Approximation* (American Mathematical Society Colloquium Publications, Providence, 1930), Vol. XI.
[25] See, e.g., W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992), Sec. 13.9.
[26] J. J. Dong and D. A. Drabold, Phys. Rev. B **54**, 10 284 (1996).