JOURNAL OF COMPUTATIONAL PHYSICS **124,** 115–130 (1996)
ARTICLE NO. 0048

Kernel Polynomial Approximations for Densities of States and Spectral Functions‡

R. N. SILVER,* H. ROEDER,† A. F. VOTER,* AND J. D. KRESS*

*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 and †Lehrstuhl Theoretische Physik I, Universitaet Bayreuth, D-95440 Bayreuth, Germany

Received April 28, 1995; revised July 13, 1995

Chebyshev polynomial approximations are an efficient and numerically stable way to calculate properties of the very large Hamiltonians important in computational condensed matter physics. The present paper derives an optimal kernel polynomial which enforces positivity of density of states and spectral estimates, achieves the best energy resolution, and preserves normalization. This kernel polynomial method (KPM) is demonstrated for electronic structure and dynamic magnetic susceptibility calculations. For tight binding Hamiltonians of Si, we show how to achieve high precision and rapid convergence of the cohesive energy and vacancy formation energy by careful attention to the order of approximation. For disordered XXZ-magnets, we show that the KPM provides a simpler and more reliable procedure for calculating spectral functions than Lanczos recursion methods. Polynomial approximations to Fermi projection operators are also proposed.

I. INTRODUCTION

In a recent paper [1], we proposed a kernel polynomial method (KPM) to estimate properties of very large Hamiltonian matrices. The DOS (density of states) is approximated by an expansion in Chebyshev polynomials, using Chebyshev moments calculated by Hamiltonian matrix-onvector multiplications (MVM) according to the polynomial recurrence relation. The resulting estimate for a DOS is a convolution of a broadening function, or *kernel,* with the true DOS. The energy resolution is inversely proportional to the number of moments calculated. For a chosen energy resolution and statistical accuracy, the computer work and memory required for sparse Hamiltonians can scale linearly in the number of states. A similar polynomial method has been proposed independently [2].

The KPM is applicable to diverse problems in computational condensed matter physics. In [1], the KPM was demonstrated for the DOS and thermodynamic functions of Heisenberg antiferromagnets with the number of states as large as 2^{27}. Similar methods have already been applied to the DOS of the Holstein t-J model [3], to the dielectric constants of Si quantum dots [4], to linear scaling algorithms for tight-binding molecular dynamics [5], to projection methods for the electronic structure problem [6], and to provide alternatives to path integral methods for statistical mechanics [7]. Compared with other methods, the KPM is easy to implement, interpret, and manipulate. Unlike competing Lanczos recursion methods (LRM), the KPM avoids accumulation of numerical roundoff errors even for large numbers of MVMs.

In view of these recent applications to condensed matter physics and the prospect for expanded use in the future, it is important to optimize the KPM. Chebyshev polynomial approximations are isomorphic to truncated Fourier series. Abrupt truncation of such series will result in unwanted Gibbs oscillations at jump discontinuities in the DOS [8] and nonuniform convergence. To avoid the Gibbs phenomenon, the series must be truncated smoothly using an appropriate damping factor. Each damping factor corresponds to a specific kernel polynomial, which determines the energy resolution, the positivity, the bias of the DOS or spectral estimate, and the unwanted "leakage" of information from one energy to another. All of the applications of Chebyshev polynomial approximations to condensed matter physics so far [1–7] have used various nonoptimal kernels. These do not minimize the Gibbs phenomenon, produce positive DOS estimates, or achieve the best possible energy resolution.

In this paper we point out that an optimal kernel polynomial for DOS and spectral applications was proposed years ago in the mathematics literature on uniform approximation by polynomials [9–11]. However, the original derivation assumed specific continuity properties of the function being approximated, properties not obeyed by DOS, and spectral functions which are sums of delta functions for finite systems. Only cumulative distributions satisfy such requirements. Therefore, we propose alternative criteria for this choice of kernel which do not depend on continuity properties. Using representative condensed matter physics

‡ LA-UR-95-357.

0021-9991/96 $18.00

tasks, we compare the empirical performance of this optimal kernel to other kernels in the recent physics and applied mathematics literature. For the tight binding electronic structure of Si, we demonstrate that rapid convergence and high precision can be obtained for the cohesive energy and vacancy formation energy provided care is taken in the order of the approximations. For disordered XXZ-magnets, we demonstrate that the KPM provides a more reliable method for calculating spectral functions than LRM [12–15]. We also propose a polynomial approximation to a Fermi projection operator, and we demonstrate that another kernel [16] more effectively projects states above the Fermi energy to zero. Throughout, we compare results of the KPM with other calculational methods.

Section II reviews the KPM introduced in [1], Section III derives an optimal kernel for DOS and spectral applications, Section IV applies the KPM to the tight binding electronic structure of Si, Section V applies the KPM to the dynamical magnetic susceptibility of disordered XXZ-magnets including a comparison with Lanczos recursion methods (LRM), Section VI applies the KPM to Fermi projection operators, and Section VII concludes.

II. THE KERNEL POLYNOMIAL METHOD

We begin with a brief review of the KPM introduced in [1]. Consider the calculation of the DOS of an $N \times N$ Hamiltonian $\mathbf{H}$, defined as

$$D(\varepsilon) = \frac{1}{N}\sum_{n=1}^{N} \delta(\varepsilon - \varepsilon_n), \tag{1}$$

where ε_n are eigenenergies. The first step is to put lower, ε_l, and upper, ε_u, bounds on the energies in the DOS. A scaled Hamiltonian matrix, $\mathbf{X}$, is defined by $\mathbf{H} = a\mathbf{X} + b$, where $a \equiv (\varepsilon_u - \varepsilon_l)/2$ and $b \equiv (\varepsilon_u + \varepsilon_l)/2$. Eigenvalues of $\mathbf{X}$ satisfy $-1 \le x_n \le +1$. The scaled DOS can be represented by a polynomial expansion,

$$\begin{aligned} D(x) &= \frac{1}{N}\sum_{n=1}^{N} \delta(x - x_n) \\ &= \frac{1}{\pi\sqrt{1-x^2}}\left[\mu_0 + 2\sum_{m=1}^{\infty} \mu_m T_m(x)\right]. \end{aligned} \tag{2}$$

The $T_m(x)$ are Chebyshev polynomials of the first kind defined by recurrence relations,

$$\begin{aligned} T_o(x) &= 1, T_1(x) = x, \\ T_{m+1}(x) &= 2xT_m(x) - T_{m-1}(x). \end{aligned} \tag{3}$$

They are orthogonal satisfying

$$\begin{aligned} \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} T_m(x)T_n(x)\,dx &= \frac{\pi}{2}\delta_{m,n}\{m, n \ge 1\}, \\ \pi\{m = n = 0\}. \end{aligned} \tag{4}$$

This expansion may be reexpressed in terms of trigonometric functions using $x = \cos(\phi)$ and $T_m(x) = \cos(m\phi)$, so that Eq. (2) is analogous to a Fourier expansion. The Chebyshev polynomial moments of the DOS are $\mu_m \equiv \int_{-1}^{1} T_m(x)D(x)\,dx$. Reference [1] discusses efficient methods for calculating or estimating the μ_m using MVMs and the polynomial recurrence relations. As Lanczos observed [17], the accumulation of roundoff error in Chebyshev methods is negligible even for huge numbers of moments.

In this paper, we presume that a finite number of moments, M, are given, and we focus on how best to use them. Abrupt truncation of Eq. (2) would result in unwanted Gibbs oscillations. The KPM considers instead smooth truncations of the form

$$D_K(x) = \frac{1}{\pi\sqrt{1-x^2}}\left[\mu_0 g_0 + 2\sum_{m=1}^{M} \mu_m g_m T_m(x)\right]. \tag{5}$$

The g_m are *Gibbs damping factors* which depend implicitly on M. The relation of this estimate to the true DOS is

$$D_K(x) = \int_{-1}^{1} K(x, x_o)D(x_o)\,dx_o, \tag{6}$$

where the *kernel polynomial* is

$$K(x, x_o) = \frac{1}{\pi\sqrt{1-x^2}}\left[g_0 + 2\sum_{m=1}^{M} g_m T_m(x)T_m(x_o)\right]. \tag{7}$$

Generally, we define the *kernel polynomial approximation* to any distribution $F(x)$ defined on $-1 \le x \le$ as

$$F_K(x) \equiv \int_{-1}^{1} K(x, x_o)F(x_o)\,dx_o. \tag{8}$$

Such kernel polynomial approximations become convolutions if they are reexpressed in $\phi = \arccos(x)$ so that $0 \le \phi \le \pi$. The corresponding distribution in ϕ is $\tilde{F}(\phi) \equiv F(x)\sin(\phi)$. This domain is extended to $0 \le \phi \le 2\pi$ by invoking π-antiperiodic boundary conditions, $\tilde{F}(2\pi - \phi) = \pm\tilde{F}(\phi)$. Define the π-antiperiodic function

$$\delta_K(\phi) \equiv \frac{1}{2\pi}\left[g_0 + 2\sum_{m=1}^{N} g_m \cos(m\phi)\right]. \tag{9}$$

This should be regarded as a π-antiperiodic polynomial

approximation to a Dirac delta function. It is normalized to $\int_{-\pi}^{\pi} \delta_K(\phi)\, d\phi = g_0$ so that we choose $g_0 = 1$. It is peaked at $\phi = 2\pi n$, where n is integer. We shall choose g_m for $m \neq 1$ such that the width of this peak is $\Delta\phi \propto 1/M$. Then Eq. (8) is equivalent to

$$\tilde{F}_K(\phi) = \int_0^{2\pi} \delta_K(\phi - \phi_o)\tilde{F}(\phi_o)\, d\phi_o. \tag{10}$$

III. UNIFORM NORM KERNEL POLYNOMIALS

The choice of Gibbs damping factor, g_m, determines the quality of kernel polynomial approximations. Kernels may be characterized by their resolution, leakage, and positivity. Resolution is defined as the width of the peak around $x \approx x_o$. Leakage is defined by how rapidly the kernel approaches zero for large $|x - x_o|$. Positivity is defined by whether approximations to DOS are positive and whether approximations to cumulative DOS are monotonic. The importance of these properties depends on the scientific application. For example, a common task in electronic structure calculations is to estimate the Fermi energy, where the cumulative DOS equals the number of electrons. However, if the estimate for the cumulative DOS is not monotonic, there may be multiple solutions for the Fermi energy. Without Gibbs damping, i.e., $g_m = 1$, the kernel has unacceptable Gibbs oscillations at large $|x - x_o|$. If instead g_m is chosen to decrease smoothly with increasing m with termination condition $g_{M+1} = 0$, the resulting kernel will peak at $x = x_o$ with a width proportional to $1/M$ and will damp Gibbs oscillations at large $|x - x_o|$. Nevertheless, there remains considerable freedom to choose possible Gibbs damping factors, which in turn leads to varying estimates of the DOS and other quantities from knowledge of M moments. Reference [1] displays kernels corresponding to several Gibbs damping factors in the literature on polynomial approximations and Fourier analysis.

This section considers the optimal choice for the kernel polynomial, $K(x, x_o)$. For applications to DOS and spectra we propose four reasonable criteria. First, the kernel should be a polynomial of degree M, if only M Chebyshev moments of the DOS are available. Second, the kernel estimates of DOS and spectra should be strictly positive as required by physics. Therefore, the kernel should be positive for all x and x_o. Third, the kernel should be normalized so that the total number of states is preserved under polynomial approximation. As discussed above, this condition is met by constraining $g_0 = 1$. And, fourth, the energy resolution should be the best achievable for M Chebyshev moments subject to these other constraints. That is, the kernel must have minimal width in $|x - x_o|$. We derive a kernel which meets these criteria. Further, we demonstrate the results are analogous to the mathematical definition of ''uniform approximation by polynomials.''

The first two criteria for our optimal kernel are equivalent to the requirement that $\delta_K(\phi)$ must be a polynomial of degree M and that it must be positive definite. These can be met uniquely by the representation

$$\delta_K(\phi) = \frac{1}{2\pi}\left|\sum_{\nu=0}^{M} a_\nu e^{i\nu\phi}\right|^2, \tag{11}$$

where the a_ν are real. Upon comparison with Eq. (9)

$$g_m = \sum_{\nu=0}^{M-m} a_\nu a_{\nu+m}. \tag{12}$$

The fourth criterion, that the energy resolution should be as good as possible, corresponds to minimizing the variance

$$\begin{aligned}\Delta\phi^2 &\equiv \int_{-\pi}^{\pi} \phi^2 \delta_K(\phi)\, d\phi \simeq \int_{-\pi}^{\pi} (2 - 2\cos(\phi))\delta_K(\phi)\, d\phi \\ &= 2g_0 - 2g_1.\end{aligned} \tag{13}$$

As M increases $\delta_K(\phi)$ becomes more sharply peaked about $\phi = 0$, and we can safely approximate $\phi^2 \approx 2 - 2\cos(\phi)$. Combining this with Eq. (12), the third criterion is equivalent to maximizing

$$Q = g_1 - \lambda g_0 = \sum_{\nu=0}^{M-1} a_\nu a_{\nu+1} - \lambda \sum_{\nu=0}^{M} a_\nu a_\nu, \tag{14}$$

where λ is a Lagrange multiplier to enforce the third criterion on normalization, $g_0 = 1$. The variational condition $\delta Q/\delta a_\nu = 0$ yields

$$a_{\nu+2} - 2\lambda a_{\nu+1} + a_\nu = 0, \quad 0 \leq \nu \leq M - 2, \tag{15}$$

with termination conditions

$$\begin{aligned} a_1 - 2\lambda a_0 &= 0, \\ -2\lambda a_M + a_{M-1} &= 0. \end{aligned}$$

But these are just the recurrence relations for Chebyshev polynomials of the second kind. Hence,

$$\begin{aligned} a_\nu &= \frac{U_\nu(\lambda)}{\sqrt{\sum_{\nu=0}^{M} U_\nu^2(\lambda)}}, \\ U_\nu(\lambda) &= \frac{\sin((\nu+1)\phi_\lambda)}{\sin(\phi_\lambda)}, \cos(\phi_\lambda) = \lambda. \end{aligned} \tag{16}$$

The last line in Eq. (16) corresponds to $U_{m+1}(\lambda) = 0$, which is equivalent to $\phi_\lambda = \pi n/(M+2)$. The maximum of Q, Eq. (14), is obtained for $n = 1$. This choice satisfies our

criteria for an optimal kernel to be used in the DOS and spectral applications of the KPM.

This kernel was originally derived by Jackson [9] in his development of the theory of uniform approximation by polynomials (see also, [10, 11]). For this reason, we shall refer to it as the Jackson (J) kernel. We briefly outline the essential elements of this theory as it relates to our problem. The uniform norm of a π-antiperiodic function $\tilde{C}(\phi)$ is defined by

$$\|\tilde{C}\| = \max_{0\le\phi\le\pi} |\tilde{C}(\phi)|. \tag{17}$$

The *best polynomial approximation of degree M,* $\tilde{C}_K$, is defined as the one which minimizes $\|\tilde{C} - \tilde{C}_K\|$. The *modulus of continuity* of $\tilde{C}(\phi)$ is defined by

$$\omega(\delta) = \sup_{\text{all } \phi_1,\phi_2;|\phi_1-\phi_2|\le\delta} |\tilde{C}(\phi_1) - \tilde{C}(\phi_2)|. \tag{18}$$

Note that $\lim_{\delta\to 0}\omega(\delta) = 0$ for continuous functions. Then, for kernel polynomial approximations, Eq. (10), one can prove the relation

$$\|\tilde{C}_K - \tilde{C}\| = \omega(\delta)\left(1 + \frac{\pi}{\delta\sqrt{2}}\sqrt{1-g_1}\right), \tag{19}$$

for some positive δ. Hence, the approximation which minimizes this norm can be achieved by maximizing g_1 under the constraint $g_0 = 1$. But this is the same variational condition as our Eq. (14). For our optimal kernel one can further prove

$$\|\tilde{C}_K - \tilde{C}\| \le \omega\left(\frac{1}{M}\right)\left(1 + \frac{\pi^2}{2}\right), \tag{20}$$

which relates the error of the uniform approximation by polynomials to the modulus of continuity. Provided $\tilde{C}$ is continuous, the error goes to zero as the number of moments goes to infinity. If the modulus of continuity is known, Eq. (20) provides error bounds on the uniform approximation by polynomials.

Although this theory provides the original derivation for this kernel, we are applying it to finite systems. The DOS and spectra of such systems are sums of Dirac delta functions, and they do not have a finite modulus of continuity. Only the cumulative DOS and cumulative spectra satisfy the conditions for the relevance of the theory of uniform approximation by polynomials; neither does Eq. (20) provide a practical procedure for estimating systematic errors bounds. By definition, the modulus of continuity of cumulative distributions is dominated by the most highly degenerate states. The corresponding error bounds are atypical and pessimistic. And, in practice, the modulus of continuity of cumulative distributions is usually unknown. For these reasons we prefer our derivation of the Jackson kernel, which does not refer to continuity properties of the DOS or spectra.

Figure 1 displays various Gibbs damping factors, g_m, calculated for $M = 100$. No Gibbs damping corresponds to $g_m^{\text{none}} = 1$. The uniform norm Gibbs damping factor is denoted by a g_m^J (J for Jackson). We also display some of the other Gibbs damping factors in the literature. Let $z \equiv m/(M+1)$. *VW* denotes a damping factor recently introduced [18] to smooth boundary conditions for many-body physics applications, but such arguments are equally applicable to kernel polynomials. It is

$$g_m^{VW} = \frac{1}{2}\left[1 - \tanh\left(\frac{z-0.5}{z(1-z)}\right)\right], \tag{21}$$

which has essential singularities at $z = 0, 1$. *L* denotes Lanczos' proposal to damp Gibbs phenomena in truncated Fourier series using "sigma factors" σ_m, which are the same as

$$g_m^{L^N} = (\sigma_m)^N = \left(\frac{\sin(\pi z)}{\pi z}\right)^N. \tag{22}$$

A very readable motivation for sigma factors can be found in Lanczos' book [16] on Fourier analysis. In [1], L^3 damping was touted as the best practical choice (although it was labeled *HO* for "higher order" sigma factor). At the time, the authors of [1] were not aware of the theory of uniform approximation by polynomials. Note that $g_m^{L^N}$ and g_m^J are remarkably similar, although only the latter will guarantee positivity of the kernel and the corresponding DOS and spectral estimates. *W* denotes the better of the two choices recommended by Wang [2],

$$g_m^W = \exp(-(z/0.8)^8). \tag{23}$$

It is far from optimal, and its practical performance is comparatively poor.

Figure 2 displays kernels corresponding to the various Gibbs damping factors in Fig. 1. The vertical axis is $2\delta_K(\phi)$ given by Eq. (9). It is a symmetric function of ϕ so only $\phi \ge 0$ need be shown. The normalization is chosen to make the area under each curve equal 1. Kernels *none, L* and *W* have comparatively better resolution near $\phi = 0$, but they exhibit significant Gibbs oscillations extending to large $|\phi|$. The deleterious practical consequence would be that information from one energy could "leak" into adjoining energies. This would be particularly severe for problems involving a few states near to a large cluster of uninteresting states, as is often the case in many-body physics. In

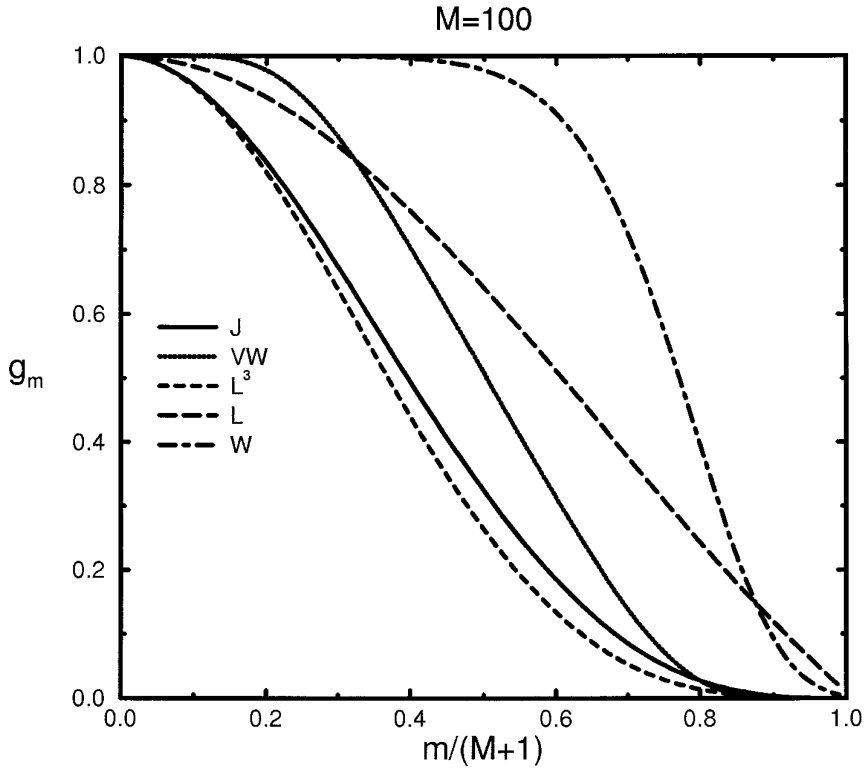

FIG. 1. *Gibbs damping factors*. The various Gibbs damping factors g_m discussed in Section III of the text. *J* denotes the uniform norm, or Jackson, function. No Gibbs damping corresponds to $g_m = 1$.

contrast, kernels L^3, *VW*, and *J* have comparatively poorer resolution at small $|\phi|$, but their Gibbs oscillations are much more rapidly damped at large $|\phi|$. This is demonstrated on the right of the figure using a vertical scale expanded by four orders of magnitude. The leakage for L^3, *VW*, and *J* would be orders of magnitude smaller than for *none, L*, and *W*. The uniform norm kernel *J* is the only one which is nonnegative and, therefore, the only one to produce monotonic cumulative distributions. Note that [1] displays the Fejer kernel, corresponding to $g_m^F = 1 - z$, which is also nonnegative. However, leakage with the Fejer kernel is severe, resulting in poor performance in practical applications. Only the L^N kernels average to zero at large $|\phi|$, meeting Lanczos' criterion of zero "bias." And only the *VW* kernel has zero variance, $\int_{-\pi}^{\pi} \phi^2 \delta_K(\phi)\, d\phi = 0$, achieved at the expense of positivity of DOS estimates.

Which of these criteria are more important depends on the scientific application. We shall demonstrate in the next section that positivity of the DOS, and the corresponding montonicity of cumulative DOS, are critical for electronic structure applications. An important argument in favor of maximum entropy methods (MEM) for moment problems [22, 23] has been that they enforce such positivity. Introduction of the Jackson kernel enables the KPM to also enforce positivity. MEM satisfies all the known moment constraints within numerical accuracy, whereas the ratios of the moments of the KPM estimate to the known moments are given by the Gibbs damping factors. Therefore, MEM can be expected to converge more rapidly than the KPM with increasing *M*. Unfortunately finding the MEM solution requires a difficult convex nonlinear optimization, whereas finding the KPM solution is trivial. And manipulations of MEM estimates are more difficult because they are nonlinear functions of the moments, whereas manipulations of KPM estimates are much simpler because they are linear functions of the moments.

IV. APPLICATION TO TIGHT BINDING ELECTRONIC STRUCTURE

The total energy, tight-binding method is gaining popularity for atomistic simulations in materials science [24, 25]. This semiempirical electronic structure approach is based on a one-electron Hamiltonian with on-site energies and interatomic-distance-dependent hopping integrals. The atom-centered basis set is chosen to be appropriate for the valence orbitals of the system under study. The band (electronic) energy E_B, defined as the sum of the

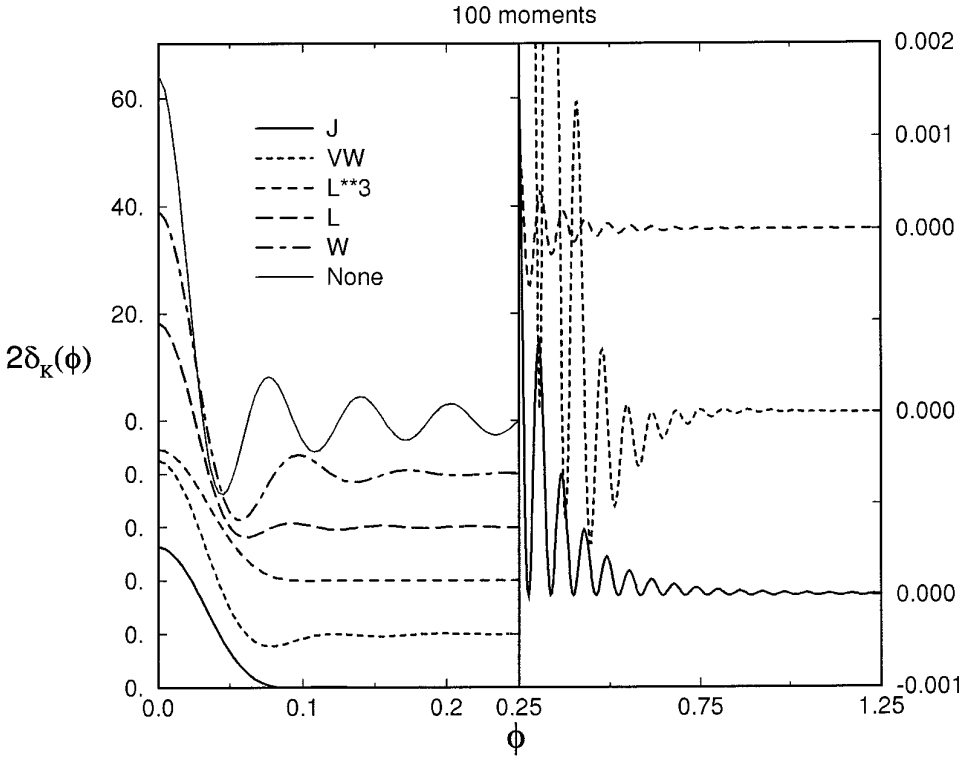

FIG. 2. *Kernel polynomials*. Various $\delta_K(\phi)$ corresponding to the Gibbs damping factors in Fig. 1. The kernels are symmetric functions of ϕ, so that only $\phi \geq 0$ is shown. The right side expands the vertical scale by four orders of magnitude expanded to exhibit the large $|\phi|$ behavior.

energies of all occupied states up to the Fermi energy, is augmented by a short-ranged pair potential. The computational bottleneck in tight-binding calculations is the diagonalization of $\mathbf{H}$, requiring $O(N^3)$ work. There has been much interest recently in approximations that improve on this scaling without undue loss of accuracy. In this section we test an $O(N^2)$ implementation of the KPM for this problem.

Silicon is chosen as a working example, and we employ the tight-binding parameters of Goodwin *et al.* [19]. The basis set consists of an *s* function and the three Cartesian *p* functions on each atom. For the present calculations, only the electronic energy is considered and the lattice is assumed to be fixed. To compute the DOS and cohesive energy of a system approximating bulk Si, a Hamiltonian matrix is constructed for a block of 216 Si atoms arranged in the diamond structure. This block is placed in a cubic box (supercell) with periodic boundary conditions. (Only the $\mathbf{k} = \mathbf{0}(\Gamma)$ point in the Brillouin zone is needed, since the 216 atom supercell system is equivalent to a nearly converged sample of $\mathbf{k}$-points for the primitive diamond-structure cell.) We also consider a 215-atom system, in which one atom has been removed while the remaining atoms are held fixed. This allows computation of the unrelaxed vacancy formation energy (E_V), defined as $E_V \equiv E_B^{215} - (\frac{215}{216})E_B^{216}$, which has been shown [20] to provide a much more stringent test of moment-based approximations than the cohesive energy.

Let $D(\varepsilon)$ be the DOS, $C(\varepsilon) = \int_{-\infty}^{\varepsilon} D(\varepsilon')\, d\varepsilon'$ be the cumulative DOS, and let $E(\varepsilon) = \int_{-\infty}^{\varepsilon} \varepsilon' D(\varepsilon')\, d\varepsilon'$ be the cumulative energy. The Fermi energy ε_F is defined by the condition that the number of occupied states corresponds to two electrons per atom per spin, $C(\varepsilon_F) = 2N_{\text{atom}}$. The band energy is then defined as $E_B \equiv E(\varepsilon_F)$. The Hamiltonian matrices considered here are sufficiently small that they may be exactly diagonalized by $O(N^3)$ methods for comparison.

Let us presume scaling factors *a* and *b* have been chosen so that all the eigenstates of the Hamiltonian have energies $\varepsilon_n = ax_n + b$ with $-1 \leq x_n \leq +1$. The scaled DOS, cumulative DOS, and cumulative energy are

$$
\begin{aligned}
D(x) &\equiv \sum_{n=1}^{N} \delta(x - x_n) = \frac{\tilde{D}(\phi)}{\sin(\phi)}, \\
C(x) &\equiv \int_{-1}^{x} D(x')\, dx' = \tilde{C}(\phi) = \int_{\phi}^{\pi} \tilde{D}(\phi')\, d\phi', \qquad (24) \\
E(x) &\equiv \int_{-1}^{x} x' D(x')\, dx' = \tilde{E}(\phi) = \int_{\phi}^{\pi} \cos(\phi') \tilde{D}(\phi')\, d\phi',
\end{aligned}
$$

respectively. Define $c_m(\phi) \equiv \cos(m\phi)/\pi$ and $s_m(\phi) \equiv \sin(m\phi)/\pi m$. Then expansions for these quantities in exact Chebyshev moments are

$$\begin{aligned}\tilde{D}(\phi) &= \frac{\mu_0}{\pi} + 2\sum_{m=1}^{\infty} \mu_m c_m(\phi),\\ \tilde{C}(\phi) &= \mu_0\left(1-\frac{\phi}{\pi}\right) - 2\sum_{m=1}^{\infty} \mu_m s_m(\phi),\\ \tilde{E}(\phi) &= \mu_1\left(1-\frac{\phi}{\pi}\right) - \sum_{m=1}^{\infty} (\mu_{m-1}+\mu_{m+1}) s_m(\phi).\end{aligned} \tag{25}$$

$\tilde{E}(\phi)$ may be continued to $0 \le \phi \le 2\pi$ using $\tilde{E}(2\pi - \phi) = -\tilde{E}(\phi)$.

Exact Chebyshev moments may be generated using the polynomial recurrence relations and matrix-vector-multiplications (MVMs),

$$T_m(\mathbf{X})|i\rangle = 2\mathbf{X}T_{m-1}(\mathbf{X})|i\rangle - T_{m-2}(\mathbf{X})|i\rangle, \tag{26}$$

where $|i\rangle$ are basis states. Then moments are constructed using

$$\begin{aligned}\mu_{2m} &= \sum_{i=1}^{N} (2\langle i|T_m(\mathbf{X})T_m(\mathbf{X})|i\rangle - 1,\\ \mu_{2m-1} &= \sum_{i=1}^{N} (2\langle i|T_m(\mathbf{X})T_{m-1}(\mathbf{X})|i\rangle - \langle i|T_1(\mathbf{X})|i\rangle).\end{aligned} \tag{27}$$

Generating M moments requires $M/2$ MVMs. Provided the Hamiltonian is sparse, this procedure requires computational work scaling like N^2M and memory scaling like N. Reference [1] discusses a stochastic generalization of this procedure using random Gaussian vectors which requires computational work scaling like NM and which yields moments subject to statistical errors. Other possible linear-scaling KPMs will be discussed in a forthcoming paper. The treatment of systematic errors discussed in the present paper applies as well to such linear scaling methods.

Practical calculations will yield only a finite number of moments. At issue is how best to truncate the moment expansions, Eq. (25). There are two obvious choices which differ in the order of mathematical operations. One might take the kernel polynomial approximation to the DOS D_K using Eq. (6), and then calculate the cumulative energy from $E(\varepsilon) \approx \int_{-\infty}^{\varepsilon} \varepsilon' D_K(\varepsilon')\, d\varepsilon'$. We refer to this as the *approximate, then integrate* option (*AI*). Alternatively, one might calculate an exact moment expansion for the cumulative DOS and cumulative energy, e.g., $E(\varepsilon) = \int_{-\infty}^{\varepsilon} \varepsilon' D(\varepsilon')\, d\varepsilon'$, and then take its kernel polynomial approximation using Eq. (10). We refer to this as the *integrate, then approximate* option (*IA*).

Both *IA* and *AI* options use the same DOS,

$$\tilde{D}_K(\phi) = \frac{\mu_0 g_0}{\pi} + 2\sum_{m=1}^{N} \mu_m g_m c_m(\phi). \tag{28}$$

But the two choices yield different approximations to the cumulative DOS and to the cumulative energy. The *IA* expressions are

$$\begin{aligned}\tilde{C}^{IA}(\phi) &= 2\sum_{m=1}^{M} (\mu_o - \mu_m) g_m s_m(\phi),\\ \tilde{E}^{IA}(\phi) &= 2\mu_1 \sum_{m=1}^{M} g_m s_m(\phi) - \sum_{m=0}^{M} \mu_m g_{m+1} s_{m+1}(\phi)\\ &\quad - \sum_{m=2}^{M} \mu_m g_{m-1} s_{m-1}(\phi).\end{aligned} \tag{29}$$

The *AI* expressions are

$$\begin{aligned}\tilde{C}^{AI}(\phi) &= \mu_0 g_0\left(1-\frac{\phi}{\pi}\right) - 2\sum_{m=1}^{M} \mu_m g_m s_m(\phi),\\ \tilde{E}^{AI}(\phi) &= g_1\mu_1\left(1-\frac{\phi}{\pi}\right) - \sum_{m=0}^{M} \mu_m g_m s_{m+1}(\phi)\\ &\quad - \sum_{m=2}^{M} \mu_m g_m s_{m-1}(\phi).\end{aligned} \tag{30}$$

At any finite M, the DOS will be strictly positive and the cumulative DOS will be strictly monotonic only if the kernel polynomial is strictly positive, regardless of the choice between the *IA* and *AI* options. Both choices converge toward the exact answer as the number of moments is increased, so the fractional difference between them can be made arbitrarily small using enough moments. However, the goal should be to achieve the highest possible precision from the fewest possible moments. We shall demonstrate empirically that the *IA* option has a much faster asymptotic convergence rate with increasing number of moments than the *AI* option. Equivalently, *IA* requires many fewer moments to achieve a desired accuracy than does *AI*. Indeed, these two options may be related using integration by parts such that $\tilde{E}^{AI} = \tilde{E}^{IA} + R$, where the residuals satisfy $|R| \propto M^{-2}$ at large M.

We now proceed to the numerical results. Figure 3 shows a portion of the DOS for the 216 atom supercell for various kernel polynomials. Only the Jackson and Fejer kernels are guaranteed to yield positive DOS estimates. Therefore, only they yield monotonic cumulative DOS and unique estimates for the Fermi energy. The Jackson kernel is pre-

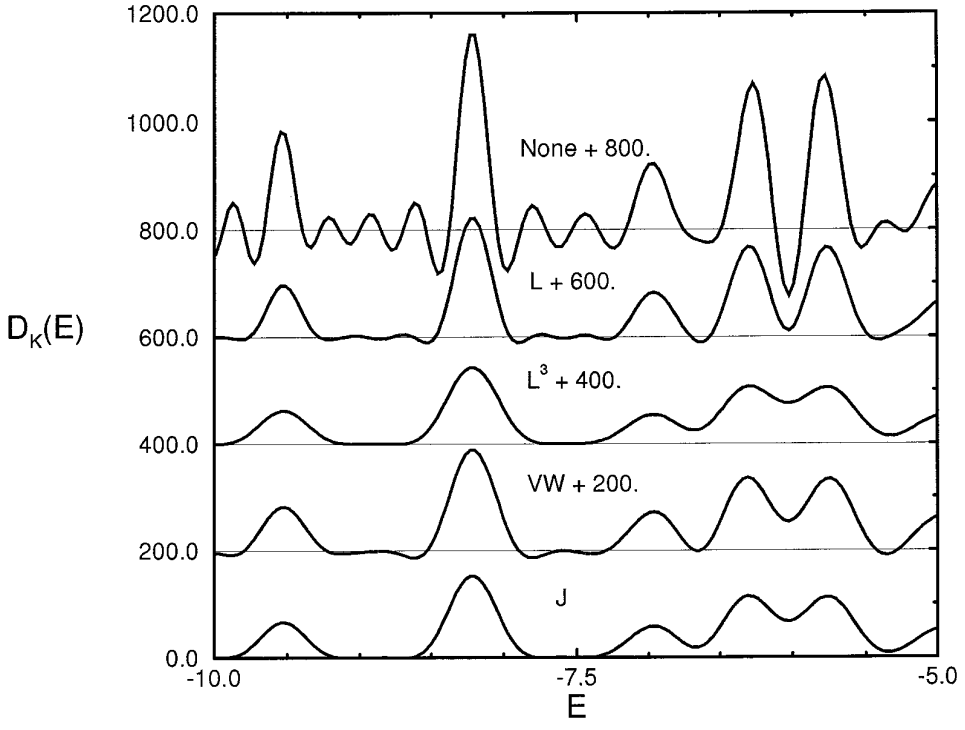

FIG. 3. *KPM DOS.* A portion of the kernel polynomial estimates of the DOS of the 216-atom Si supercell calculated for 200 Chebyshev moments for the various kernel polynomials displayed in Fig. 2.

ferred because of its much smaller leakage and better resolution. The L^3 kernel, first recommended in [1], produces results which are quite similar to the Jackson kernel, but the DOS is still not strictly positive.

Figure 4 demonstrates how the Fermi energy and the band energy are determined. Shown are the DOS, cumulative DOS, and cumulative energy for the 216 atom supercell obtained using the Jackson kernel, 200 moments, and the *IA* option. The Fermi energy is the point where the cumulative DOS equals the number of electrons. Then, the band energy is the cumulative energy evaluated at the Fermi energy. For comparison, results from $O(N^3)$ exact diagonalization are shown as vertical lines. They are positioned at eigenenergies of the Hamiltonian and their heights are equal to 20× their degeneracy.

Figure 5 compares the performance of the *IA* option, Eq. (29), to the *AI* option, Eq. (30). Shown on the left is the band energy, E_B, for the 216 atom supercell using a linear horizontal scale for M. Shown on the right is the vacancy formation energy, E_V, displayed using a logarithmic horizontal scale for M. The vertical scale is energy in electron volts. Also shown is the exact result obtained by diagonalization of the Hamiltonian, an $O(N^3)$ process. E_V^{IA} has converged on the exact tight binding E_V to within 1.0 eV for $M \geq 40$. Recall that the polynomial recurrence procedure requires only $M/2$ MVMs to yield M moments. So E_V^{IA} converges after only 20 MVMs. The *AI* option converges more slowly than the *IA* option, with the difference between E_V^{AI} and the exact E_V varying as M^{-2} at large M as predicted by theory. This figure makes clear that the *IA* option is preferable to the *AI* option for the vacancy formation energy calculation. However, for other calculations the KPM does not yet provide criteria for which delta function to approximate by a kernel polynomial, leading to different KPM estimates from the same moment data. The resolution of this ambiguity is a topic for future research.

The values for E_V^{IA} agree well with maximum entropy (MEM) results calculated for $M \leq 20$ [20]. However, the cohesive energy converges more slowly for KPM than for MEM; i.e., E_B^{IA} converges to within 0.1 eV/atom of exact tight binding by $M = 30$ and asymptotes to within 0.1 eV total at $M \geq 150$, while MEM requires only $M \geq 10$ to converge within 0.1 eV/atom. The faster convergence of MEM relative to KPM is offset by its much greater computational complexity. A detailed comparison of KPM and MEM will be presented elsewhere.

Finally, Fig. 6 compares the performance of various kernels for vacancy formation energy calculations. The *IA*

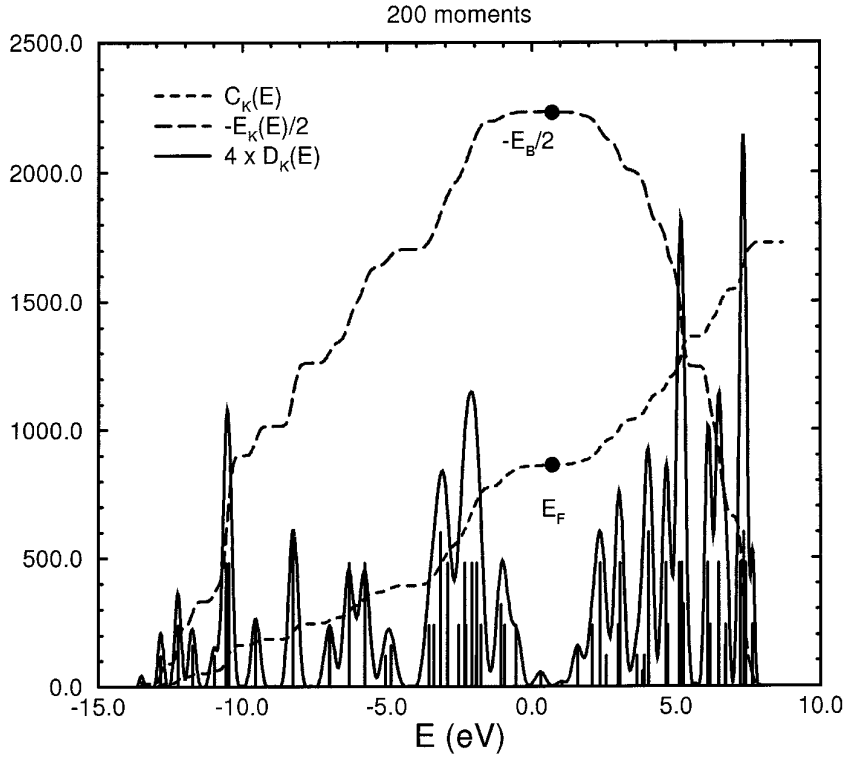

FIG. 4. *Calculating the band energy*. The solid line is the KPM DOS for the Si 216-atom supercell calculated with the Jackson kernel polynomial and 200 moments (100 MVMs). For comparison, the vertical lines depict exact diagonalization results. The lines are positioned at eigenenergies of the Hamiltonian and their height equals 20× their degeneracy. The dashed line is the cumulative DOS. The Fermi energy E_F is determined by $C_K(E_F) = 2N_{\text{atom}}$. The band energy is the cumulative energy at the Fermi energy, $E_B = E_K(E_F)$.

option is used throughout. The kernels for *none, L*, and L^3 Gibbs damping are not positive, resulting in possible multiple solutions for the Fermi energy found by a bisection procedure. This ambiguity is reflected in the scatter of results for E_V as a function of the number of moments. Only the Jackson kernel is positive definite, guaranteeing a unique solution for the Fermi energy and the most rapid convergence for E_V. The performance of the L^3 kernel is comparable.

V. APPLICATION TO DYNAMICAL PROPERTIES OF SPIN SYSTEMS

The most commonly used procedures for calculating properties of large Hamiltonians are Lanczos recursion methods (LRM). They require the same optimized MVM algorithm as the KPM. The LRM is the optimal way to calculate the ground state energy and wave functions, as well as static correlations as a function of the parameters of the Hamiltonian. To obtain more physical insight into the properties of the ground state, it is often desirable to calculate dynamical properties such as spectral functions,

$$A^O(\omega) = \frac{1}{\pi}\lim_{\delta\to 0^+} \Im m \left\{ \langle \Psi_0 | \mathbf{O}^\dagger \frac{1}{\omega - \mathbf{H} - i\delta} \mathbf{O} | \Psi_0 \rangle \right\}, \quad (31)$$

$$= \sum_{n=1}^{N} |\langle \Psi_0 | \mathbf{O} | n \rangle|^2 \delta(\omega + \varepsilon_0 - \varepsilon_n). \quad (32)$$

Here $\mathbf{O}$ is the matrix representation of an operator, e.g., the spin operator if one wants to calculate the dynamical magnetic susceptibility. Also, $|\Psi_0\rangle$ and ε_0 are the ground state wave function and the ground state energy, and $|n\rangle$ and ε_n define the excited states of the Hamiltonian $\mathbf{H}$. We shall argue that the KPM provides a more reliable method to calculate uniformly smoothed approximations to spectral functions than the LRM.

The KPM calculation of a spectral function proceeds as follows. Bounds on the energy spectrum and the ground state wave function are found using the LRM. The energies and Hamiltonian are rescaled as before, i.e., $\mathbf{H} \to \mathbf{X}$ and $\omega \to x$, so that all energies lie between $[-1, +1]$ and $0 \le x \le 2$. The delta function in Eq. (32) is approximated by a kernel polynomial

$$A_K^O(x) \approx \langle \Psi_o | \mathbf{O}^\dagger K(x-1, \mathbf{X}) \mathbf{O} | \Psi_o \rangle,$$
$$= \frac{1}{\pi\sqrt{1-(x-1)^2}} \left[g_o \mu_0^O + 2 \sum_{m=1}^{M} g_m T_m(x-1) \mu_m^O \right]. \quad (33)$$

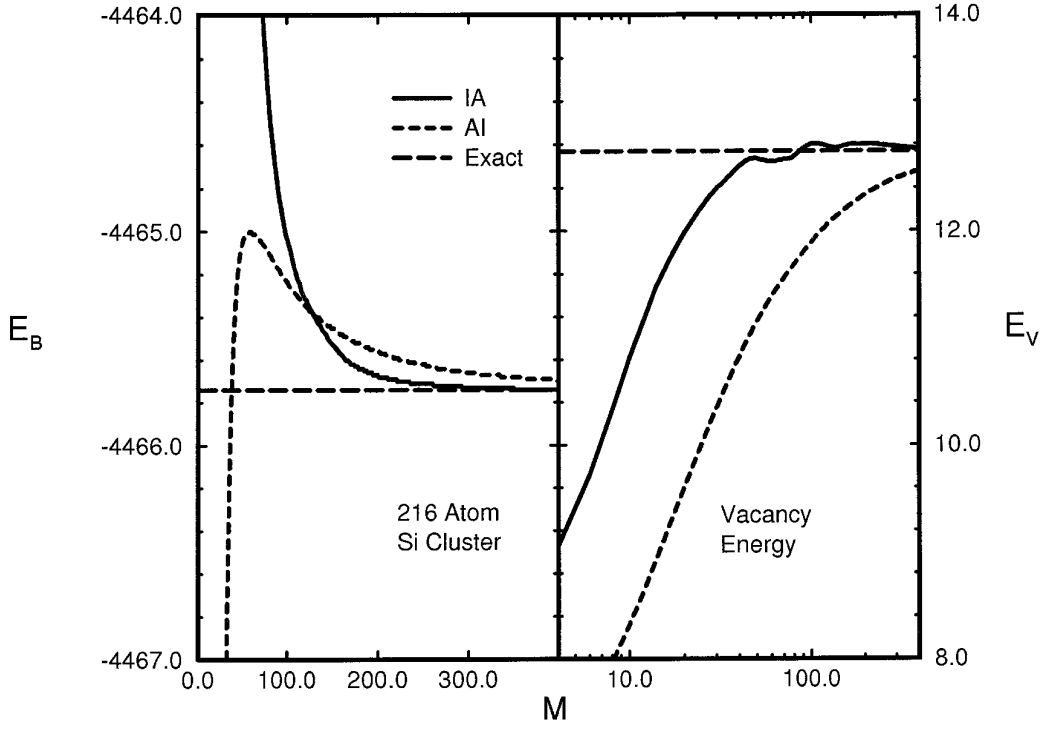

FIG. 5. *Calculating the vacancy formation energy*. The band energy, E_B, of the 216-atom supercell is shown on the left and the vacancy formation energy, E_V, is shown on the right as a function of the number of moments M. *IA* corresponds to the *integrate*, *then approximate* option described in Section IV. *AI* corresponds to the alternative *approximate*, *then integrate* option. More rapid convergence with increasing numbers of moments is obtained for the *IA* option.

Here the moments

$$\mu_m^O = \langle \Psi_o | \mathbf{O}^\dagger T_m(\mathbf{X}) \mathbf{O} | \Psi_o \rangle. \tag{34}$$

A vector $\mathbf{O}|\Psi_o\rangle$ is calculated. Then vectors $T_m(\mathbf{X})\mathbf{O}|\Psi_o\rangle$ are generated using the Chebyshev recurrence relations. Moments, Eq. (34), are again calculated most efficiently using relations for combining Chebyshev polynomials analogous to Eq. (27), so that M moments requires only $M/2 + 1$ MVMs. The resulting KPM spectral function is the convolution of a kernel polynomial (or resolution function) with the true spectral function, i.e.,

$$A_K^O(x) = \int_0^2 K(x-1, x'-1) A^O(x')\, dx', \quad 0 \le x \le 2. \tag{35}$$

The resolution is uniform in $\phi = \arccos(x-1)$, and the width is proportional to M^{-1}. With increasing M, $A_K^O(x)$ converges in a controlled manner to $A^O(x)$. We will compare KPM spectral functions to those obtained by the Lanczos recursion method (LRM). We use only the Jackson kernel for these comparisons.

The original LRM [12] for spectral functions yields a continued fraction expansion,

$$A_L^O(\omega) = \cfrac{\langle \Psi_0 | \mathbf{O}^\dagger \mathbf{O} | \Psi_0 \rangle}{\omega - a_o - \cfrac{b_1^2}{\omega - a_1 - \cfrac{b_2^2}{\omega - \cdots}}}, \tag{36}$$

The coefficients a_i, b_i are obtained using the recursion

$$\mathbf{f}_{i+1} = \mathbf{H}\mathbf{f}_i - a_i \mathbf{f}_i - b_i \mathbf{f}_{i-1}, \tag{37}$$

with

$$a_i = \mathbf{f}_i^{\mathrm{T}} \mathbf{H} \mathbf{f}_i / r_i, \quad b_{i+1}^2 = \mathbf{f}_{i+1}^{\mathrm{T}} \mathbf{H} \mathbf{f}_i, \quad r_i = \mathbf{f}_i^{\mathrm{T}} \mathbf{f}_i, \tag{38}$$

and the initial conditions

$$\mathbf{f}_0 = \mathbf{O}|\Psi_0\rangle, \quad b_0 = 0. \tag{39}$$

After M recursions, $A_L^O(\omega)$ consists of M delta function peaks which is difficult to display. In [12] a finite (small) δ was introduced in Eq. (31) to obtain a smooth spectral function. In [13, 14] this smoothing method was further refined in order to mimic an infinite system. The number of recursions in practical continued fraction applications

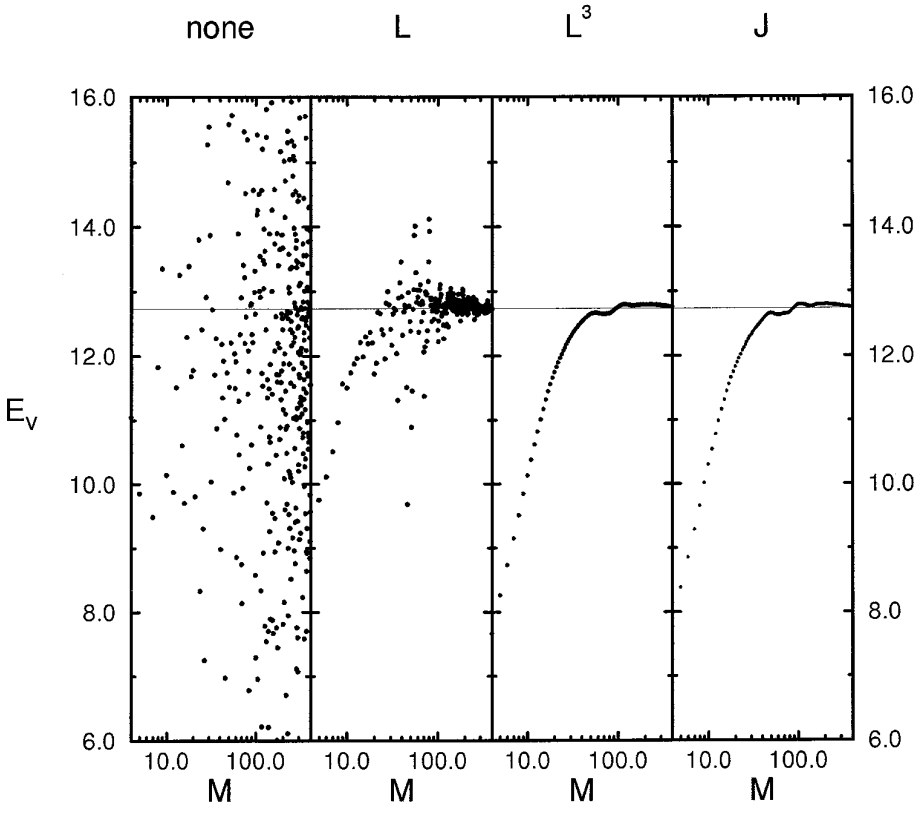

FIG. 6. *The effect of Gibbs damping factors on E_V.* The panels display scatter plots of the vacancy formation energy, E_V, as a function of the number of moments M on a logarithmic scale. Calculations used the *IA* option and the various kernels indicated. The E_V for nonpositive kernel polynomials is bouncing about, because there are multiple possible solutions for the Fermi energy. The Jackson kernel enforces positivity, so that there is a unique solution for the Fermi energy and rapid convergence. The L^3 kernel is comparable to the Jackson kernel for this application.

is limited to less than 100 by the high accuracy required for the coefficients.

A further refinement of LRM for spectral functions, the "spectral decoding method," was presented in [15]. They noticed that $A^O(\omega)$ can be approximated by

$$A^O(\omega) \approx \sum_{i=0}^{M} |\langle\phi_i|\mathbf{O}|\Psi_0\rangle|^2 \delta(\omega + \varepsilon_0 - \tilde{\varepsilon}_i), \qquad (40)$$

where $\tilde{\varepsilon}_i$ and $|\phi_i\rangle$ are the eigenvalues and eigenstates of $\mathbf{H}$ in the restricted basis set $\{\mathbf{f}_i,\ i = 1,\ ...,\ M\}$ generated by the recursion Eq. (37). In this basis set $\mathbf{H}$ is a tridiagonal matrix, $\mathbf{T}_M$, because the transformation matrix,

$$\mathbf{V} = (\mathbf{f}_0, \mathbf{f}_1, ..., \mathbf{f}_M), \qquad (41)$$

is an orthogonal $N \times (M + 1)$ matrix and $\mathbf{T}_M = \mathbf{V}^{\mathrm{T}}\mathbf{H}\mathbf{V}$.

Eigenvalues of the Hamiltonian appear as eigenvalues of T_M with increasing M, a behavior termed the Lanczos phenomenon. Using exact arithmetic one could obtain all N eigenvalues of the Hamiltonian by $M = N$ MVMs. For finite precision arithmetic a loss of orthogonality of the Lanczos vectors results in a requirement for $M \gg N$. If we denote the eigenvectors of $\mathbf{T}_M$ by $|\tilde{\phi}_i\rangle$, i.e., $\mathbf{T}_M|\tilde{\phi}_i\rangle = \tilde{\varepsilon}_i|\tilde{\phi}_i\rangle$, then $|\phi_i\rangle = \mathbf{V}^{\mathrm{T}}|\tilde{\phi}_i\rangle$ and the amplitudes $\langle\phi_i|\mathbf{O}|\Psi_0\rangle$ in Eq. (40) are given by the first components of the eigenvectors of $\mathbf{T}_M$:

$$\langle\phi_i|\mathbf{O}|\Psi_0\rangle = (|\tilde{\phi}_i\rangle)_1. \qquad (42)$$

The spectral decoding method eliminates the need to evaluate continued fraction coefficients. In the following, we use only results from the spectral decoding method to compare the LRM with the KPM, but our conclusions shall also apply to other LRM for spectral functions.

In both the KPM and LRM the computer work intensive steps are the MVMs. The LRM needs MVMs for the Lanczos recursion, and the KPM needs MVMs for the Chebyshev recursion. Thus both methods can benefit from optimization of the same MVM algorithm.

But the LRM and KPM differ in their numerical stability in the presence of numerical roundoff error. The Chebyshev recursion is numerically stable with no generation of

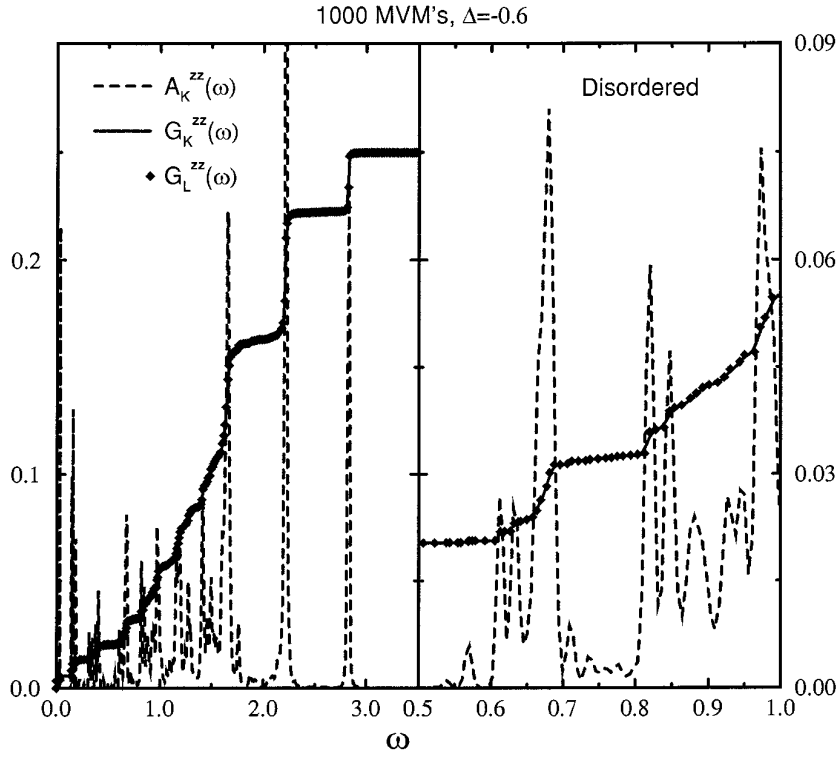

FIG. 7. *Longitudinal dynamical magnetic susceptibility*. Cumulative spectral functions G^{zz} from the KPM (solid line) and the LRM (diamonds) for the XXZ model. The spectral function A^{zz} from the KPM is also shown. All curves correspond to 1000 matrix-on-vector multiplications. The anisotropy strength is $\Delta = 0.6$ and the disorder strength is $\sigma = 0.5$.

spurious structure even at extremely large M, whereas the Lanczos recursion is numerically unstable. Note that the recursion Eq. (37) is very similar to the usual Lanczos recursion [21]. The important difference is that the vectors $\mathbf{f}_i$ in Eq. (37) are not normalized. The normalization coefficients r_i in Eq. (38) exponentially increase or decrease with recursion index i resulting in an eventual overflow or underflow at large M. (In the example below, we reached $r_{1000} = 3.3 \times 10^{2247}$.) Moreover, the Lanczos phenomenon results in a loss of orthogonality and a rapid accumulation of roundoff error. This can be overcome only by repeated reorthogonalization or by an elaborate error analysis, as we shall discuss further below.

As our example to compare various methods for calculation spectral functions, we consider the one-dimensional XXZ-model with exchange randomness described by the Hamiltonian

$$\mathbf{H}^{XXZ} = \sum_{i}^{N} (1 + J_i)(\mathbf{S}_i^x \mathbf{S}_{i+1}^x + \mathbf{S}_i^y \mathbf{S}_{i+1}^y) + \Delta \mathbf{S}_i^z \mathbf{S}_{i+1}^z. \qquad (43)$$

The sum goes over N sites with periodic boundary conditions. The random couplings J_i are drawn from a Gaussian distribution with width σ. Δ controls the anisotropy. Without random exchange the model can be solved exactly using the Bethe ansatz. The Hamiltonian has recently been of interest, because of the appearance of a random singlet phase [26]. We shall consider the longitudinal and the transverse dynamical spin susceptibilities, A^{zz} and A^{+-}. In the case of no disorder they consist of a series of well separated, coherent peaks. With disorder one obtains many more peaks, which may be closely spaced. Since our goal is to compare methods of calculation rather than to learn new physics, we do not average over the realizations of disorder. We consider just one realization of the J_i's. We present results for a nontrivial system size of $N = 22$ sites, which results in a Hilbert space dimension of 705432 in the $\mathbf{S}^z_{\text{total}} = \sum_i^N \mathbf{S}_i^z = 0$ subspace. As parameters we chose anisotropy parameter $\Delta = -0.6$ and a disorder strength $\sigma = 0.5$.

For finite systems the true spectral functions will consist of a sum of delta functions with amplitudes $|\langle \Psi_0 | \mathbf{O} | n \rangle|^2$, which are difficult to display graphically. In a practical calculation, the number of recursions M may be orders of magnitude smaller than the number of states. A Lanczos estimate for the spectral density, $A_L^Q(\omega)$, will consist of M

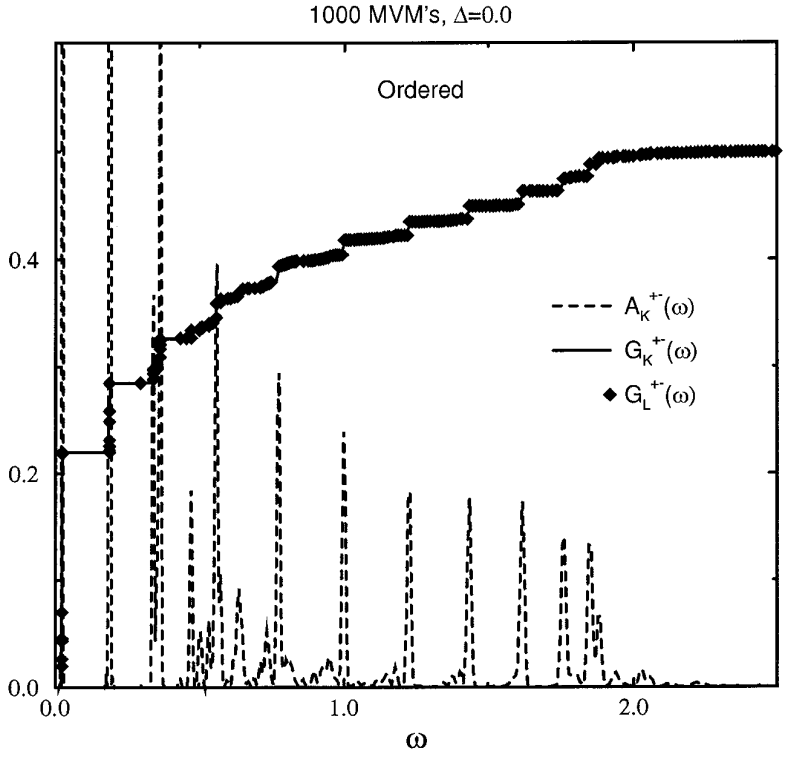

FIG. 8. *Transverse dynamical magnetic susceptibility*. Cumulative spectral functions G^{+-} from the KPM (solid line) and the LRM (diamonds) for the XXZ model. The spectral function A^{+-} from the KPM is also shown. All curves correspond to 1000 matrix-on-vector multiplications. The anisotropy strength is $\Delta = 0.0$ and the disorder strength is $\sigma = 0.0$.

delta-function peaks. A kernel polynomial estimate for the spectral density, $A_K^O(\omega)$, will be a naturally smoothed version of $A^O(\omega)$. In order to avoid *ad hoc* smoothing of $A_L^O(\omega)$, we compare the LRM and KPM on the basis of their cumulative spectral functions,

$$G^O(\omega) = \int_{-\infty}^{\omega} d\omega' A^O(\omega'). \tag{44}$$

$G^O(\omega)$ is a superposition of step functions, with steps at the poles of the spectral function and the step-height equal to the amplitude of the delta function peaks. A plot of $G^O(\omega)$, instead of $A^O(\omega)$, displays information about both the position of the poles and their amplitudes. In addition we shall show $A_K^O(\omega)$.

We compare $G_L^O(\omega)$ and $G_K^O(\omega)$ as a function of M. $G_L^O(\omega)$ is a superposition of step functions at the positions of the eigenvalues of $\mathbf{T}_M$. $G_K^O(\omega)$ is a convolution of a kernel function with the true $G^O(\omega)$. As a reference we use results from 1000 MVMs, for which LRM and KPM give practically identical spectra for the parameters chosen. Figure 7 shows the results for the disordered case and $\Delta = 0.6$, and Fig. 8 shows the results for the ordered case. The results from the LRM (the diamonds) lie right on top of the results of the KPM (the solid lines). The position of the diamonds corresponds to the eigenvalues $\tilde{\varepsilon}_i$ from Eq. (40). For one given peak (e.g., the first one in Fig. 8) there may be more than one $\tilde{\varepsilon}_i$, whose weights then add up to the true $G^O(\omega)$. To illustrate the behaviour of the LRM as a function of the number of recursion steps M, we show in Fig. 9 an enlarged portion of a spectral density. The eigenvalues move about as M is increased. There are also spurious eigenvalues which disappear as the iteration proceeds. Fortunately, spurious eigenvalues appear to have a small weight and little influence on the cumulative spectral function.

In a practical Lanczos iteration, one can categorize the eigenvalues of $\mathbf{T}_M$ by their behavior as a function of M:

- Eigenvalues may correspond to one peak in the true spectral function, even though they may be degenerate. (This degeneracy is not exact but within numerical accuracy, because a symmetric tridiagonal matrix only has simple eigenvalues or it is block diagonal). Such eigenvalues are labeled as *converged*. Multiple copies of the same eigenvalue are often labeled *clones*.
- Eigenvalues may appear and disappear as the iteration

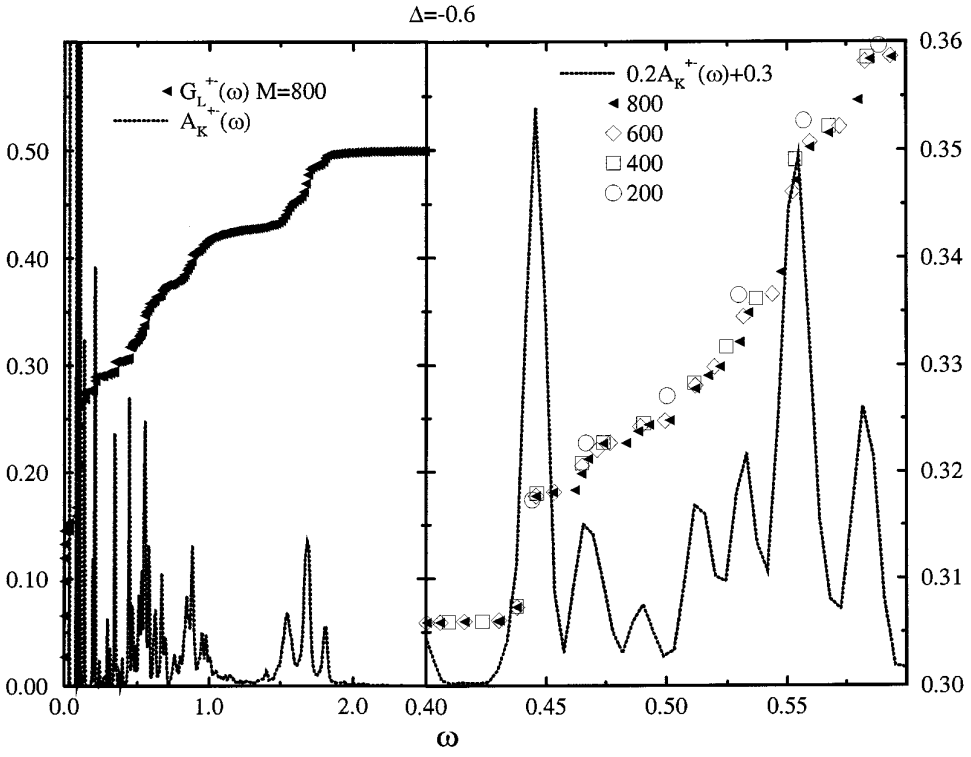

FIG. 9. *The Lanczos phenomenon*. The right side shows the cumulative function G^{+-} for the XXZ model as a function of recursion steps M. The parameter values are the same as in Fig. 7. Shown is an enlarged section. As a guide to the eye $A^{+-} + 0.3$ from the KPM is drawn to indicate the position of poles. The left side shows the full scale for 800 recursions.

proceeds. They are labelled as *spurious* and should not be used. Spurious eigenvalues are often labeled *ghosts*.

• Eigenvalues may change slowly as the iteration proceeds. They are labeled as *approximations* to real eigenvalues. The rate of convergence of approximate eigenvalues depends on their position in the spectrum. Those which lie on the edges of the true eigenvalue spectrum and which are well separated converge faster than those associated with closely spaced clusters of true eigenvalues.

The convergence of one of the $\tilde{\varepsilon}_i$ causes the loss of orthogonality of the vectors $\mathbf{f}_i$ in the presence of numerical roundoff error. For small enough matrices it is possible to reorthogonalize the vectors. For large problems Cullum and Willoughby [21] have proposed a procedure that does the labeling as described above and produces error bars.

In view of this well-established behavior of LRM, the popular procedure to yield a smooth spectral estimate from the LRM (i.e., uniformly broaden the peaks defined by eigenvalues of $\mathbf{T}_M$ by introducing a nonzero width δ in Eq. (31)) appears to make little sense. Smoothing of converged eigenvalues, which appear first for well-separated or band edge states, throws away valid information from the LRM. Smoothing of approximate eigenvalues, which generally correspond to clusters of true states, yields spectral estimates with an unknown relation to the true spectrum. Smoothing of spurious eigenvalues is even more absurd, when there exist procedures to identify and remove them. The amount of smoothing used in such LRM methods is *ad hoc*, and there is no reason to expect the optimal smoothing to be uniform across the entire spectrum. In contrast, in the KPM smoothing is uniquely determined by the number of Chebyshev recursions, and it is naturally uniform.

In summary, the LRM and the KPM give qualitatively similar results for the cumulative function $G^O(\omega)$. However, the KPM provides a more direct route to obtaining a spectral function with a uniform resolution. Natural uniform smoothing is built into the KPM, and no spurious structure due to numerical instability is generated even at extremely large M. In contrast, the relation of the LRM to spectral functions is more complicated due to the Lanczos phenomena. Uniform smoothing of the LRM poles is misleading because of the nonuniform convergence rates and spurious peaks as a function of M. The number of recursions in the LRM is limited by an eventual over(under)-flow of the r_i in Eq. (38). If the scientific interest centers on only a few well separated peaks or band edges in a spectral function, the LRM might be cheaper in terms of MVMs. If the number of poles becomes large or they become closely spaced, the KPM is more reliable. We find

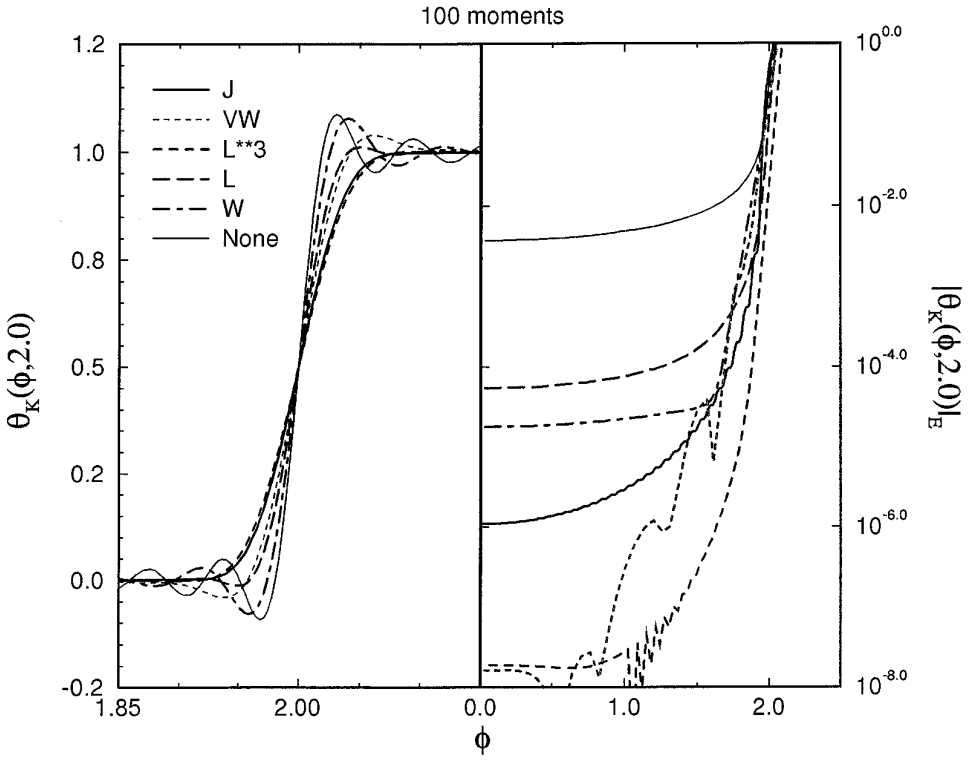

FIG. 10. *Fermi projection operators*. The left side shows kernel polynomial approximations to step functions and Fermi projection operators for various kernels on a linear vertical scale. The right-hand side shows the envelope function of the polynomial approximation to step functions on a logarithmic vertical scale and an expanded linear ϕ scale. Calculations are shown for 100 moments. The L^3 kernel and VW kernels project unoccupied states to zero by at least eight orders of magnitude, limited only by machine precision. The Jackson kernel provides the next most effective Fermi projection operator, projecting to zero by about six orders of magnitude. Other kernels provide even less effective projection operators.

KPM spectral functions to be easier to implement and interpret than the LRM.

VI. APPLICATION TO FERMI PROJECTION OPERATORS

Although the Jackson kernel has the best performance for spectra and DOS, other kernels may be preferable for other scientific applications. Indeed, as shown in [1], the no Gibbs damping choice $g_m = 1$ is best for statistical mechanics calculations in which partition functions may be expanded in Chebyshev moments and modified Bessel functions. In this section, we demonstrate that the Lanczos kernel and its variants can be more effective than the Jackson kernel in constructing polynomial approximations to Heaviside projection operators. Such operators are used in electronic structure calculations, for example, to project unoccupied states above the Fermi energy to zero, in which case we label them *Fermi projection operators*.

An ideal Fermi projection operator $\mathbf{P}^F$ is defined by its action on a random vector $|\psi\rangle$. Consider the expansion in eigenstates $|n\rangle$ of $\mathbf{H}$,

$$|\psi\rangle = \sum_{n=1}^{N} b_n |n\rangle, \quad \mathbf{P}^F|\psi\rangle = \sum_{n=1}^{N} \Theta(E_F - \varepsilon_n) b_n |n\rangle, \quad (45)$$

where Θ is a Heaviside step function. That is, $\mathbf{P}^F$ is the operator equivalent of a step function, $\mathbf{P}_F = \Theta(E_F - \mathbf{H})$. Ideal Fermi projection operators would require an infinite number of MVMs to construct, but polynomial approximations to them are feasible. For example, [5, 6] construct projection operators using polynomial approximations to finite temperature Fermi–Dirac functions with temperature a free tuning parameter. However, in our context, a step function is the integral of a Dirac delta function, $\delta(x - x_o)$. Therefore, a polynomial approximation to a step function can be constructed from the integral of the polynomial approximation to a delta function, the kernel $K(x, x_o)$. And a polynomial approximation to a projection operator can be constructed from the integral of the operator kernel, $K(x, \mathbf{X})$.

Specifically, this corresponds to the choice $\tilde{F}(\phi) = \Theta(\phi - \phi_F)$ in Eq. (10), where ϕ_F is defined by $\varepsilon_F = a\cos(\phi_F) + b$. Then

$$\Theta_K(\phi, \phi_F) = 1 - \frac{\phi_F}{\pi} - 2 \sum_{m=1}^{M} g_m s_m(\phi_F) \cos(m\phi). \quad (46)$$

Similarly, the kernel polynomial approximation to a Fermi projection operator is

$$\mathbf{P}_K^F = \left(1 - \frac{\phi_F}{\pi}\right) \mathbf{1} - 2 \sum_{m=1}^{M} g_m s_m(\phi_F) T_m(\mathbf{X}). \quad (47)$$

In practice, this operator is applied by using the Chebyshev recurrence relations to generate the $T_m(\mathbf{X})|\psi\rangle$.

Figure 10 demonstrates the dependence of the quality of Fermi projection operators on the choice of Gibbs damping factor. It plots Eq. (46) for the same Gibbs damping factors as in Figs. 1 and 2. The left side shows $\Theta_K(\phi, \phi_F)$ for $\phi_F = 2.0$ and $M = 100$. What is important for the applications of Fermi projection operators is how closely this function approaches zero for $\phi \leq \phi_F$. These functions are oscillatory due to the Gibbs phenomenon, so we display on the right the envelope function of their absolute value. In other words, the curves on the right are upper bounds on the true $|\Theta_K(\phi, \phi_F)|$. The most efficient projection of high energy ($\varepsilon > \varepsilon_F$ or $\phi < \phi_F$) states to zero is given by the L^3 Gibbs damping, with *VW* damping a close second. Both these functions project to zero by more than eight orders of magnitude, limited only by machine double precision on a 32-bit computer. The third best choice for a Fermi projection operator is provided by the Jackson Gibbs damping factor J, which projects to zero by six orders of magnitude in this example. The other kernels provide orders of magnitude less effective Fermi projection operators.

Because projection to zero improves as the number of moments is increased, the higher efficiency of the L^3 Gibbs damping corresponds to a computer work requirement for fewer MVMs. In practice one can achieve even more effective projection to zero at $\phi \leq \phi_F$ than L^3 by using even higher power Lanczos damping L^N. However, this would be achieved at the expense of the slope of the function for $\phi \approx \phi_F$. Ultimately, a decision about which power of the Lanczos Gibbs damping factor to use depends on the application.

VII. CONCLUSION

This paper, coupled with others in the recent literature [2–7], demonstrates the versatility of the kernel polynomial method [1] for diverse applications in computational condensed matter physics. The introduction of the Jackson kernel polynomial improves the KPM by enforcing positivity of DOS and spectral estimates, along with monotonicity of cumulative DOS. Such positivity is crucial for rapid convergence in electronic structure applications. The Jackson kernel also provides the highest resolution uniformly smoothed estimates of dynamical spectral functions. Relative to other methods for DOS and spectra such as Lanczos recursion and maximum entropy, the KPM is much easier to implement and manipulate with small risk of numerical instability. No reorthogonalization of vectors or nonlinear optimization is required. The KPM requires the same optimized matrix–vector multiplication algorithm used in Lanczos recursion methods. We believe the KPM will prove to be a useful complement to the arsenal of methods for large matrix computations in condensed matter physics.

ACKNOWLEDGMENTS

This research was supported in part by the Office of Basic Energy Sciences, Division of Materials Research, of the U.S. Department of Energy.

REFERENCES

1. R. N. Silver and H. Röder, *Int. J. Mod. Phys. C* **5,** 735 (1994).
2. L. W. Wang, *Phys. Rev. B* **49,** 10154 (1994).
3. H. Röder, H. Fehske, and R. N. Silver, *Europhys. Lett.* **28,** 257 (1994).
4. L. W. Wang and A. Zunger, *Phys. Rev. Lett.* **73,** 1039 (1994).
5. S. Goedecker and L. Colombo, *Phys. Rev. Lett.* **73,** 122 (1994).
6. O. F. Sankey, D. A. Drabold, and A. Gibson, *Phys. Rev. B* **50,** 1376 (1994).
7. R. N. Silver, to be published.
8. G. Arfken, *Mathematical Methods for Physicists* (Academic Press, New York, 1985), Section 14.5.
9. D. Jackson, *Theory of Approximation*, Amer. Math. Soc. Colloq. Publ., Vol. XI (Am. Math. Soc., Providence, RI, (1930).
10. G. Meinardus, *Approximation of Functions: Theory and Numerical Methods* (Springer-Verlag, New York, 1967).
11. T. J. Rivlin, *An Introduction to the Approximation of Functions* (Blaisdell, Waltham, MA, 1969).
12. E. R. Gagliano and C. A. Balseiro, *Phys. Rev. Lett.* **59,** 2999 (1987).
13. E. Dagotto *et al.*, *Phys. Rev. B* **41,** 9049 (1990).
14. D. Poilblanc *et al.*, *Phys. Rev. B* **47,** 3268 (1993).
15. Q. F. Zhong, S. Sorella, and A. Parola, *Phys. Rev. B* **49,** 6408 (1994).
16. C. Lanczos, *Discourse on Fourier Series* (Hafner, New York, 1956).
17. C. Lanczos, *Applied Analysis* (Prentice–Hall, Englewood Cliffs, NJ, 1956), p. 187.
18. M. Vekić and S. R. White, *Phys. Rev. Lett.* **71,** 4283 (1993).
19. L. Goodwin, A. J. Skinner, and D. G. Pettifor, *Europhys. Lett.* **9,** 701 (1989).
20. J. D. Kress and A. F. Voter, *Phys. Rev. B*, submitted.
21. J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations. Vol. I. Theory* (Birkhaeuser, Boston, 1985).
22. L. R. Mead and N. Papanicolaou, *J. Math. Phys.* **25,** 2404 (1984).
23. D. A. Drabold and O. F. Sankey, *Phys. Rev. Lett.* **70,** 3631 (1993).
24. O. F. Sankey and R. E. Allen, *Phys. Rev. B* **33,** 7164 (1986).
25. C. Z. Wang, C. T. Chan, and K. M. Ho, *Phys. Rev. B* **39,** 8586 (1989).
26. C. A. Doty and D. S. Fisher, *Phys. Rev. B* **45,** 2167 (1992).