Green's function of the Holstein polaron

Glen L. Goodvin, Mona Berciu, and George A. Sawatzky

Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

We present a highly efficient yet accurate analytical approximation for the Green's function of a Holstein polaron. It is obtained by summing all the self-energy diagrams, but with each self-energy diagram averaged over the momenta of its free propagators. The result becomes exact for both zero bandwidth and for zero electron-phonon coupling and is accurate everywhere in the parameter space. The resulting Green's function satisfies exactly the first six spectral weight sum rules. All higher sum rules are satisfied with great accuracy, becoming asymptotically exact for coupling both much larger and much smaller than the free particle bandwidth. Comparison with existing numerical data also confirms this accuracy. We use this approximation to analyze in detail the redistribution of the spectral weight as the coupling strength varies.

I. INTRODUCTION

There is considerable interest in understanding the effects on the properties of a particle coming from interactions with an environment. Examples of such problems abound in condensed matter; the problem discussed here is that of an electron coupled to lattice vibrations, i.e., of electron-phonon coupling. For example, such coupling is believed to be relevant for understanding certain aspects of the high-temperature superconductors' behavior in the underdoped limit (where the "particle" coupling to phonons is the doping hole already dressed by interactions with the electrons in the lower Hubbard band),[1–4] but there are many examples of other materials characterized by strong-electron phonon coupling, including polymers like polyacetylene, nanotubes, C_{60} molecules, and other fullerenes.[5–7] Other problems of the same general type regard electrons coupled to spin waves of a magnetically ordered system,[8] or to orbitrons, for example in manganites,[9–11] or to some combination thereof.

In this work, we focus on the simplest Hamiltonian describing an electron on a lattice interacting with an optical phonon mode, namely the Holstein model:[12]

$$\mathcal{H} = \sum_{\mathbf{k}} (\varepsilon_{\mathbf{k}} c_{\mathbf{k}}^{\dagger} c_{\mathbf{k}} + \Omega b_{\mathbf{k}}^{\dagger} b_{\mathbf{k}}) + \frac{g}{\sqrt{N}} \sum_{\mathbf{k},\mathbf{q}} c_{\mathbf{k}-\mathbf{q}}^{\dagger} c_{\mathbf{k}} (b_{\mathbf{q}}^{\dagger} + b_{-\mathbf{q}}). \quad (1)$$

The first term is the kinetic energy of the electron, with $c_{\mathbf{k}}^{\dagger}$ and $c_{\mathbf{k}}$ being the electron creation and annihilation operators. For the single dressed particle (known as polaron, in this case) problem of interest to us, the spin of the particle is irrelevant and we suppress its index. $\varepsilon_{\mathbf{k}}$ is the free-particle dispersion. In all results shown here, we assume nearest-neighbor hopping on a d-dimensional simple cubic lattice of constant a and with a total of N sites with periodic boundary conditions, so that

$$\varepsilon_{\mathbf{k}} = -2t \sum_{i=1}^{d} \cos(k_i a), \quad (2)$$

but our results are valid for any other dispersion. The second term describes a branch of optical phonons of energy Ω. $b_{\mathbf{q}}^{\dagger}$ and $b_{\mathbf{q}}$ are the phonon creation and annihilation operators.

The last term is the on-site linear electron-phonon coupling $g \sum_i c_i^{\dagger} c_i (b_i^{\dagger} + b_i)$, written in $\mathbf{k}$ space. All sums over momenta are over the first Brillouin zone, $-\pi/a < k_i \leq \pi/a$. We set $\hbar = 1$ and $a = 1$ throughout this paper.

The quantity of interest to us is the Green's function of the single dressed particle, or polaron, defined as[13]

$$G(\mathbf{k}, \tau) = -i \langle 0 | T[c_{\mathbf{k}}^{\dagger}(\tau) c_{\mathbf{k}}(0)] | 0 \rangle, \quad (3)$$

where $|0\rangle$ is the ground state of the zero-particle system, which is the vacuum. T is the time ordering operator, and $c_{\mathbf{k}}(\tau) = e^{i\mathcal{H}\tau} c_{\mathbf{k}} e^{-i\mathcal{H}\tau}$. Since $\mathcal{H}|0\rangle = 0$ and $c_{\mathbf{k}}|0\rangle = 0$, Eq. (3) simplifies to

$$G(\mathbf{k}, \tau) = -i\Theta(\tau) \langle 0 | c_{\mathbf{k}} e^{-i\mathcal{H}\tau} c_{\mathbf{k}}^{\dagger} | 0 \rangle, \quad (4)$$

where $\Theta(\tau)$ is the Heaviside function. In other words, only the retarded part contributes. This Green's function gives the amplitude of probability that an electron introduced in the system at $\tau = 0$ and removed at a later time τ, leaves the system in its ground state. The usefulness of this quantity can be appreciated using its Lehmann representation:[13]

$$G(\mathbf{k}, \omega) = \sum_{\alpha} \frac{|\langle \alpha | c_{\mathbf{k}}^{\dagger} | 0 \rangle|^2}{\omega - E_{\alpha} + i\eta}, \quad (5)$$

where $\{|\alpha\rangle\}$ and $\{E_{\alpha}\}$ are the complete set of one-particle eigenstates and eigenenergies, $\mathcal{H}|\alpha\rangle = E_{\alpha}|\alpha\rangle$, and $\sum_{\mathbf{k}} c_{\mathbf{k}}^{\dagger} c_{\mathbf{k}}|\alpha\rangle = |\alpha\rangle$. The poles of the Green's function give the *whole one-particle spectrum*, while the associated residues, called quasiparticle (qp) weights, give partial information on the nature of the eigenstates. Moreover, the imaginary part of this Green's function, called the spectral weight, is directly measured experimentally through angle-resolved photoemission spectroscopy.[14]

There has already been a large amount of work devoted to understanding the properties of the Holstein polaron, and we briefly review some of it here. Most of it is numerically intensive work. Some examples are (i) exact diagonalization (ED) methods.[15–17] These are usually hampered by the fact that even for a finite lattice, the Hilbert space is infinite due to the infinite number of possible phonon configurations. Thus some truncation scheme is needed, but for small pho-

non frequencies and/or large couplings, the CPU resources needed and run times become prohibitive. This has led to a number of (ii) proposals based on variational approaches to decide which phonon configurations should be included,[18–24] as well as (iii) various quantum Monte Carlo (QMC) methods.[25–27] A brief review of these is provided in Refs. [28] and [29]. Of special interest are the so-called diagrammatic Monte Carlo simulations,[30–32] which calculate directly the Green's function in imaginary time by *numerically* summing all diagrams in the perturbational expansion. We make extensive use of comparisons with low-energy results of this method from Ref. [32]. While this method allows in principle the exact calculation of the Green's function, the requirement of convergence for the propagator series usually means that only low-energy properties are shown. Finally, there are methods suitable for some particular cases, such as density-matrix renormalization group for one-dimensional systems[33] and dynamic mean-field theory (DMFT) for infinite-dimensional systems.[34,35]

Of course, given the long history of this problem, many analytical techniques have been applied to it with varying degrees of success (for a review, see Ref. [36]). First of all, the Green's function is known exactly in two asymptotic limits. If there is no coupling, $g=0$, then the Green's function is that of the free electron. The ground state is at $\epsilon_0 = -2dt$ and the spectrum consists of a continuous band extending from $[-2dt, 2dt]$ (for the tight-binding model). The so-called impurity limit, with $t=0$, also has an exact solution, given by the Lang-Firsov formula[37]

$$G(\omega) = e^{-g^2/\Omega^2} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{g}{\Omega}\right)^{2n} \frac{1}{\omega + \frac{g^2}{\Omega} - n\Omega + i\eta}. \quad (6)$$

This can be viewed as the strong-coupling limit, since for $t=0$, g becomes the important energy scale in the system. In this limit the electron is localized at one site in real space, therefore it is fully delocalized in $\mathbf{k}$ space, and the Green's function is independent of $\mathbf{k}$. The spectrum has the ground state (GS) at $E_0 = -g^2/\Omega$ and an infinite sequence of equidistant levels spaced by Ω above it. This is extremely different from the free-particle spectrum, and it is of considerable interest to understand not only how the ground state evolves from $-2dt$ to $-g^2/\Omega$ as the ratio g/t is increased, but also the evolution of all the higher-energy spectral weight from a continuous, finite-width band to an infinite set of discrete levels. It is known that this crossover regime is analytic, that is, there are no abrupt changes in the polaron properties as the coupling is increased.[38]

Note that it is convenient to describe the effective coupling as the ratio of the two asymptotic ground-state energies, using as a new parameter

$$\lambda = \frac{g^2}{2dt\Omega}.$$

Most of the numerical methods reviewed above calculate only GS or low-energy properties, given the significant CPU time and numerical resources needed to calculate the whole spectrum. Very recently, several sets of whole-spectrum re-

TABLE I. Comparison between the total number of diagrams of a given order in the proper self-energy $\Sigma(\mathbf{k}, \omega)$ vs the number of diagrams of a given order kept within SCBA.

Order	1	2	3	4	5	6	7	8
Total	1	2	10	74	706	8162	110 410	1 708 394
SCBA	1	1	2	5	14	42	132	429

sults have become available,[17,23,28,39–41] however, only for a few points in the parameter space, and generally for low dimensions.

It is of obvious interest to find an analytical approximation for the Green's function that is simple to estimate, so that the whole parameter space can be studied easily, but also with high accuracy. This is precisely what we propose here (a short version of this work has been published in Ref. [42]). We call our approximation the momentum average (MA) approximation; its essence consists in *analytically* summing all the diagrams in the diagrammatic expansion, but with each diagram simplified in a certain way. Before introducing this method, we briefly review here the other two simple (in terms of computational effort) analytical approximations for the Green's function of the Holstein polaron, available in the literature.

The first is the self-consistent Born approximation (SCBA), which consists of summing exactly only the non-crossed diagrams. The percentage of diagrams kept decreases fast with increasing order (see Table I). If the coupling is small, the sum is dominated by the low order diagrams and SCBA works reasonably well. At strong coupling, the contribution of higher order diagrams becomes essential, and SCBA is expected to fail (see below). In this approximation, the Green's function is written in terms of a self-energy:

$$G_{\text{SCBA}}(\mathbf{k}, \omega) = \frac{1}{\omega - \epsilon_{\mathbf{k}} - \Sigma_{\text{SCBA}}(\omega) + i\eta},$$

with the self-consistency condition

$$\Sigma_{\text{SCBA}}(\omega) = \frac{g^2}{N} \sum_{\mathbf{q}} G_{\text{SCBA}}(\mathbf{k} - \mathbf{q}, \omega - \Omega).$$

Note that $\Sigma_{\text{SCBA}}(\omega)$ is independent of $\mathbf{k}$. This is a consequence of the simplicity of the Holstein model: If either the coupling g or the dispersion Ω were functions of the phonon momentum $\mathbf{q}$, the SCBA self-energy would depend explicitly on $\mathbf{k}$.[43] $\Sigma_{\text{SCBA}}(\omega)$ can be expressed[42] as a function of the average of the free propagator over the Brillouin zone (BZ) and can be evaluated very efficiently.

The other simple analytical approximation for the Green's function of the Holstein model is the generalized Lang-Firsov (LF) expression.[44,45] It is reminiscent of the Lang-Firsov expression of Eq. (6):

$$G_{LF}(\mathbf{k},\omega) = e^{-g^2/\Omega^2} \sum_{n=0}^{\infty} \frac{\frac{1}{n!}\left(\frac{g}{\Omega}\right)^{2n}}{\omega - e^{-g^2/\Omega^2}\varepsilon_{\mathbf{k}} + \frac{g^2}{\Omega} - n\Omega + i\eta}. \quad (7)$$

This expression is exact in both asymptotic limits $\lambda=0$ ($g=0$) and $\lambda \to \infty$ ($t=0$), but less accurate for finite t and g, as we show below.

The article is organized as follows. In Sec. II, we derive the momentum average approximation. Its diagrammatic meaning is discussed in Sec. III, where we also estimate its corresponding spectral weight sum rules. We show that MA satisfies exactly the first six sum rules, but, more importantly, it remains highly accurate for higher order sum rules. This is a strong argument in favor of its accuracy. The accuracy is gauged in more detail in Sec. IV, where we compare the MA predictions against the SCBA and generalized LF predictions, but also against a host of numerical results. This indeed demonstrates that the MA approximation is remarkably accurate for all parameter values, especially given its simplicity. In this section we also present some results regarding various properties of the Holstein polaron. Finally, Sec. V contains our summary and conclusions.

II. CALCULATING THE GREEN'S FUNCTION

A. Exact solution

As is always the case for Green's functions, one can use the equation of motion technique to generate an infinite hierarchy of coupled equations for an infinite number of related Green's functions. We derive it here for the Holstein polaron.

In the frequency domain, this approach is equivalent to using repeatedly Dyson's identity $\hat{G}(\omega)=\hat{G}_0(\omega) + \hat{G}(\omega)\hat{V}\hat{G}_0(\omega)$, which holds for any Green's operators $\hat{G}(\omega)=[\omega-\hat{\mathcal{H}}+i\eta]^{-1}$, $\hat{G}_0(\omega)=[\omega-\hat{\mathcal{H}}_0+i\eta]^{-1}$ and for any Hamiltonians $\hat{\mathcal{H}}=\hat{\mathcal{H}}_0+\hat{V}$. As is customary, we take $\hat{V}$ to be the electron-phonon interaction. Applying Dyson's identity once, we obtain

$$G(\mathbf{k},\omega) = G_0(\mathbf{k},\omega)\left[1 + \frac{g}{\sqrt{N}}\sum_{\mathbf{q}_1} F_1(\mathbf{k},\mathbf{q}_1,\omega)\right], \quad (8)$$

where

$$G_0(\mathbf{k},\omega) = \langle 0|c_{\mathbf{k}}\hat{G}_0(\omega)c_{\mathbf{k}}^{\dagger}|0\rangle = \frac{1}{\omega - \varepsilon_{\mathbf{k}} + i\eta} \quad (9)$$

is the free particle Green's function. We made use of the equality $\hat{V}c_{\mathbf{k}}^{\dagger}|0\rangle = (g/\sqrt{N})\sum_{\mathbf{q}}c_{\mathbf{k}-\mathbf{q}}^{\dagger}b_{\mathbf{q}}^{\dagger}|0\rangle$ and defined a new Green's function:

$$F_1(\mathbf{k},\mathbf{q}_1,\omega) = \langle 0|c_{\mathbf{k}}\hat{G}(\omega)c_{\mathbf{k}-\mathbf{q}_1}^{\dagger}b_{\mathbf{q}_1}^{\dagger}|0\rangle.$$

F_1 is related to the amplitude of probability to start with the electron and a phonon at the initial time, and find only the electron in the system at the final time. Its own equation of motion relates back to $G(\mathbf{k},\omega)$ but also to a new Green's function with two phonons initially. In general, if we define

$$F_n(\mathbf{k},\mathbf{q}_1,\ldots,\mathbf{q}_n,\omega) = \langle 0|c_{\mathbf{k}}\hat{G}(\omega)c_{\mathbf{k}-\mathbf{q}_T}^{\dagger}b_{\mathbf{q}_1}^{\dagger}\cdots b_{\mathbf{q}_n}^{\dagger}|0\rangle,$$

where $\mathbf{q}_T=\sum_{i=1}^{n}\mathbf{q}_i$ is the total momentum of the n initial phonons, using Dyson's identity we find its equation of motion to be ($n \geq 1$)

$$F_n(\mathbf{k},\mathbf{q}_1,\ldots,\mathbf{q}_n,\omega) = \frac{g}{\sqrt{N}}G_0(\mathbf{k}-\mathbf{q}_T,\omega-n\Omega) \times \left[\sum_{i=1}^{n} F_{n-1}(\mathbf{k},\mathbf{q}_1,\ldots,\mathbf{q}_{i-1},\mathbf{q}_{i+1},\ldots,\mathbf{q}_n,\omega) + \sum_{\mathbf{q}_{n+1}} F_{n+1}(\mathbf{k},\mathbf{q}_1,\ldots,\mathbf{q}_{n+1},\omega)\right], \quad (10)$$

i.e., related to the Green's functions with $n-1$ and $n+1$ initial phonons. Equations (8) and (10) form the exact infinite hierarchy of coupled equations whose solution is the Holstein polaron Green's function $G(\mathbf{k},\omega)=F_0(\mathbf{k},\omega)$.

Obviously, this system of coupled equations can be solved trivially in the limit $\lambda=g=0$, in which case $G(\mathbf{k},\omega)=G_0(\mathbf{k},\omega)$ directly from Eq. (8). An exact solution equal to the Lang-Firsov result must also exist if $t=0$. Indeed, in this limit all Green's functions become independent of all momenta, and Eqs. (8) and (10) simplify to

$$G(\omega) = G_0(\omega)[1 + g\sqrt{N}F_1(\omega)],$$

$$F_n(\omega) = gG_0(\omega-n\Omega)\left[\frac{n}{\sqrt{N}}F_{n-1}(\omega) + \sqrt{N}F_{n+1}(\omega)\right],$$

where $G_0(\omega)=(\omega+i\eta)^{-1}$. These recurrence equations can be solved in terms of continued fractions. We briefly review the solution here. We suppress the functional notation and rewrite $F_n=\alpha_n F_{n-1}+\beta_n F_{n+1}$, where $\alpha_n \equiv (ng/\sqrt{N})G_0(\omega-n\Omega)$ and $\beta_n \equiv g\sqrt{N}G_0(\omega-n\Omega)$. On physical grounds we expect that F_{n+1} becomes vanishingly small for a sufficiently large n, since it describes physical processes which are less and less likely. This allows one to solve these equations iteratively starting from this sufficiently large n: $F_n \approx \alpha_n F_{n-1}$, to find after solving for F_{n-1}, F_{n-2}, etc., that
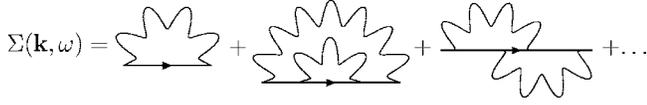
FIG. 1. The diagrammatic expansion for the self-energy.

$$F_1 = \frac{\alpha_1}{1 - \dfrac{\alpha_2\beta_1}{1 - \dfrac{\alpha_3\beta_2}{1 - \cdots}}} F_0. \qquad (11)$$

Allowing the continued fraction to be infinite instead of truncated after n steps gives the exact solution. Recalling that $F_0 \equiv G = G_0(1 + g\sqrt{N}F_1)$, one can now solve for G. With the original notation, we find

$$G(\omega) = \frac{G_0(\omega)}{1 - \dfrac{g^2 G_0(\omega)G_0(\omega-\Omega)}{1 - \dfrac{2g^2 G_0(\omega-\Omega)G_0(\omega-2\Omega)}{1 - \cdots}}}. \qquad (12)$$

After some further work, this can indeed be shown to equal the Lang-Firsov expression of Eq. (6).

This exact hierarchy of coupled equations [Eqs. (8) and (10)] can also be solved in the general case of finite g and finite t by iteratively solving for F_1, then F_2, etc., and removing them from this coupled system. It is straightforward to verify that the solution obtained in this case is the diagrammatic expansion, which can be partially resummed to the expected form

$$G(\mathbf{k},\omega) = \frac{1}{[G_0(\mathbf{k},\omega)]^{-1} - \Sigma(\mathbf{k},\omega)}, \qquad (13)$$

where the self-energy $\Sigma(\mathbf{k},\omega)$ is the sum of all proper self-energy diagrams, the first few of which are shown in Fig. 1. While this solution as a sum of an infinite number of diagrams is exact, it is clearly not very useful if the sum cannot be performed. One typical strategy in such cases is to sum only a subset of these diagrams (e.g., the noncrossed ones, in the self-consistent Born approximation). This is reasonable when one can argue that the diagrams kept contribute much more than the neglected diagrams, which is not the case for SCBA in this problem (see below). We propose another strategy, explained in the next subsection, to find an approximate solution of these equations in the case of finite t and finite g.

B. Momentum average approximation

To obtain an approximate solution in the case of finite t and g, we proceed as follows. We first note that $G(\mathbf{k},\omega)$ depends on F_1 only through its average over the Brillouin zone $f_1(\mathbf{k},\omega) = (1/N)\sum_{\mathbf{q}_1} F_1(\mathbf{k},\mathbf{q}_1,\omega)$,

$$G(\mathbf{k},\omega) = G_0(\mathbf{k},\omega)[1 + g\sqrt{N}f_1(\mathbf{k},\omega)]. \qquad (14)$$

We define the *momentum averaged* Green's functions:

$$f_n(\mathbf{k},\omega) = \frac{1}{N^n} \sum_{\mathbf{q}_1,\ldots,\mathbf{q}_n} F_n(\mathbf{k},\mathbf{q}_1,\ldots,\mathbf{q}_n,\omega), \qquad (15)$$

where all momenta sums run over the BZ, and attempt to express Eq. (10) in terms of these simpler quantities, by performing the corresponding momenta averages on both sides. While the first term on the right-hand side can be written in terms of f_{n-1} exactly, the second term requires an approximation, which we choose to be

$$\frac{1}{N^n} \sum_{\mathbf{q}_1,\ldots,\mathbf{q}_{n+1}} F_{n+1}(\mathbf{k},\mathbf{q}_1,\ldots,\mathbf{q}_{n+1},\omega)G_0(\mathbf{k}-\mathbf{q}_T,\omega-n\Omega)$$
$$\approx N\bar{g}_0(\omega-n\Omega)f_{n+1}(\mathbf{k},\omega), \qquad (16)$$

where

$$\bar{g}_0(\omega) = \frac{1}{N}\sum_{\mathbf{k}} G_0(\mathbf{k},\omega) \qquad (17)$$

is the free propagator momentum-averaged over the BZ. One justification for replacing $G_0(\mathbf{k}-\mathbf{q}_T,\omega-n\Omega)$ by its momentum average $\bar{g}_0(\omega-n\Omega)$ in Eq. (16) is that $\mathbf{q}_T = \sum_{i=1}^n \mathbf{q}_i$ takes, with equal probability, any value in the BZ. Moreover, in the impurity limit $t=0$, where all Green's functions are momentum independent, Eq. (16) is exact. This suggests that our approach should be reasonable at least in the strong-coupling limit $t \ll g$. In a more practical sense, this approximation allows us to write $f_n(\mathbf{k},\omega)$ in terms of $f_{n-1}(\mathbf{k},\omega)$ and $f_{n+1}(\mathbf{k},\omega)$ only, which is what we require to be able to obtain an analytical expression for $G(\mathbf{k},\omega)$. We discuss the meaning and consequences of this approximation in more detail below.

After the approximation of Eq. (16), Eq. (10) becomes

$$f_n(\mathbf{k},\omega) = \frac{g\bar{g}_0(\omega-n\Omega)}{\sqrt{N}}[nf_{n-1}(\mathbf{k},\omega) + Nf_{n+1}(\mathbf{k},\omega)].$$

Together with Eq. (14) these simplified recurrence relations can be solved similarly to the $t=0$ case. We find

$$G(\mathbf{k},\omega) = \frac{1}{\omega - \varepsilon_{\mathbf{k}} - \Sigma_{MA}(\omega) + i\eta}, \qquad (18)$$

where the self-energy is, within the MA approximation,

$$\Sigma_{MA}(\omega) = \frac{g^2\bar{g}_0(\omega-\Omega)}{1 - \dfrac{2g^2\bar{g}_0(\omega-\Omega)\bar{g}_0(\omega-2\Omega)}{1 - \dfrac{3g^2\bar{g}_0(\omega-2\Omega)\bar{g}_0(\omega-3\Omega)}{1 - \cdots}}}. \qquad (19)$$

This is the main result of this work. As discussed, the MA approximation becomes exact in the limit of zero hopping ($t=0$), where $\bar{g}_0(\omega) \to G_0(\omega) = (\omega+i\eta)^{-1}$, but also for zero coupling, $g=0$. In the following sections, we show that the range of validity of this approximation extends well beyond these asymptotic limits, and that in fact the MA expression is reasonably accurate over the entire parameter space.

Note that although the MA self-energy looks similar to the DMFT self-energy,[34,35] this is in fact a very different approximation. This issue is discussed in the Appendix.

III. MEANING OF THE MA APPROXIMATION

A. Diagrammatics

To understand the diagrammatical meaning of the MA approximation, we expand Eq. (19) in powers of g^2:

$$\Sigma_{MA}(\omega) = g^2 \bar{g}_0(\omega - \Omega) + g^4 [2\bar{g}_0^2(\omega - \Omega)\bar{g}_0(\omega - 2\Omega)]$$
$$+ g^6 [4\bar{g}_0^3(\omega - \Omega)\bar{g}_0^2(\omega - 2\Omega) + 6\bar{g}_0^2(\omega - \Omega)$$
$$\times \bar{g}_0^2(\omega - 2\Omega)\bar{g}_0(\omega - 3\Omega)] + O(g^8) \qquad (20)$$

and analyze the various terms. First, one can verify that the MA approximation generates the correct number of proper self-energy diagrams to all orders. Indeed, there is one term of order g^2, two terms of order g^4, $4 + 6 = 10$ terms of order g^6, and so on (see Table I). Moreover, to each of these terms we can associate an MA diagram. These have the same topology as the exact proper self-energy diagrams. The difference is that each free propagator $G_0(\mathbf{k}, \omega)$ in the exact self-energy diagrams is replaced with a momentum averaged free propagator $\bar{g}_0(\omega)$ (with the correct frequency) in each MA diagram.

Using Eq. (17), the first-order self-energy diagram is (see Fig. 1)

$$\Sigma^{(1)}(\mathbf{k}, \omega) = \frac{g^2}{N} \sum_{\mathbf{q}} G_0(\mathbf{k} - \mathbf{q}, \omega - \Omega) = g^2 \bar{g}_0(\omega - \Omega),$$

i.e., $\Sigma^{(1)}(\mathbf{k}, \omega) = \Sigma_{MA}^{(1)}(\omega)$, and thus MA is exact to first order. Differences appear from the second order diagrams, where the two exact contributions (see Fig. 1)

$$\frac{g^4}{N^2} \sum_{\mathbf{q}_1, \mathbf{q}_2} G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega) G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \omega - 2\Omega)$$
$$\times [G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega) + G_0(\mathbf{k} - \mathbf{q}_2, \omega - \Omega)] \qquad (21)$$

are replaced, within MA, by two equal contributions:

$$2g^4 \bar{g}_0^2(\omega - \Omega)\bar{g}_0(\omega - 2\Omega) = \frac{2g^4}{N^3} \sum_{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3} G_0(\mathbf{q}_1, \omega - \Omega)$$
$$\times G_0(\mathbf{q}_2, \omega - 2\Omega) G_0(\mathbf{q}_3, \omega - \Omega). \qquad (22)$$

Comparing Eq. (21) to Eq. (22), we see that the MA self-energy diagrams have the correct number of free propagators with the correct frequencies; however, the momenta of the free propagators are uncorrelated and individually averaged over. It is as if there is no connection between the momentum carried by a cloud phonon when it is emitted and when it is reabsorbed by the electron. Precisely the same holds for all higher order self-energy diagrams.

To gain a better understanding of the difference between the exact and the MA diagrams, let us further analyze the dependence on t of Eqs. (21) and (22). For $t = 0$ the expressions are identical, because the free propagators become in-

dependent of momenta and, as already discussed, MA becomes exact. For finite t, we expand each free propagator as

$$G_0(\mathbf{k}, \omega) = G_0(\omega)[1 + \varepsilon_\mathbf{k} G_0(\omega) + (\varepsilon_\mathbf{k} G_0(\omega))^2 + \cdots],$$

where $G_0(\omega) = (\omega + i\eta)^{-1}$. Inserting this expansion into Eqs. (21) and (22) and collecting powers of t yields the following. All $O(t)$ terms vanish in both the exact and the MA diagrams because they are proportional to a $\Sigma_\mathbf{q} \varepsilon_\mathbf{q} = 0$. In fact, all odd-order powers in t vanish because $\Sigma_\mathbf{q} \varepsilon_\mathbf{q}^{2n+1} = 0$. Next, consider terms of order t^2. Such terms arise either from expanding one free propagator to $O(t^2)$ or from expanding two different free propagators to $O(t)$. The former case leads to the same result for both the exact and the MA diagrams, and we obtain six contributions proportional to $(1/N)\Sigma_\mathbf{q} \epsilon_\mathbf{q}^2 = 2dt^2$. The latter case, however, reveals a difference. Five of the six $O(t^2)$ such contributions from the exact diagrams vanish because they involve propagators carrying different momenta and, for example, $\Sigma_{\mathbf{q}_1, \mathbf{q}_2} \varepsilon_{\mathbf{k} - \mathbf{q}_1} \varepsilon_{\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2} = 0$. The exception comes from the two outside free propagators of the noncrossed diagram, which carry the same momentum and result in another $(1/N)\Sigma_{\mathbf{q}_1} \varepsilon_{\mathbf{k} - \mathbf{q}_1}^2 = 2dt^2$ contribution. This is absent in the MA approximation, where different propagators always carry different momenta. It follows that the MA second order self-energy diagrams capture six out of the seven finite $O(t^2)$ contributions correctly. Similar considerations apply for higher order t powers and for higher order diagrams, differences between the MA and the exact self-energy diagrams coming only from terms involving free propagators carrying equal momenta in noncrossed diagrams. However, the error from such missed terms becomes smaller and smaller as one goes to higher order diagrams because the percentage of self-energy diagrams with one or more pairs of free propagators of equal momenta decreases exponentially.

We conclude that the MA approximation captures most of the t dependence of each self-energy diagram, while summing over all diagrams. This analysis suggests that MA should be quite accurate for any finite g and t values. In the next section, we reinforce this conclusion by analyzing the sum rules of the spectral weight.

B. Sum rules

For an even better idea of the accuracy of the MA approximation, we consider the sum rules for the spectral weight $A(\mathbf{k}, \omega) = -(1/\pi) \text{Im} \, G(\mathbf{k}, \omega)$, defined as

$$M_n(\mathbf{k}) = \int_{-\infty}^{\infty} d\omega \, \omega^n A(\mathbf{k}, \omega). \qquad (23)$$

For a problem of this type (single dressed particle), the sum rules can be calculated exactly to arbitrary order. The usual approach is based on the equation of motion technique.[45] We review it briefly here in order to make a few useful observations. The key step is to rewrite $\omega^n = (i(d/d\tau))^n e^{-i\omega\tau}|_{\tau=0}$, so that we have

$$M_n(\mathbf{k}) = -\frac{1}{\pi} \operatorname{Im} \left(i \frac{d}{d\tau} \right)^n \int_{-\infty}^{\infty} d\omega \, e^{-i\omega\tau} G(\mathbf{k}, \omega) \Bigg|_{\tau=0}.$$

The integral is now simply $G(\mathbf{k}, \tau \to 0^+)$. Using the definition of Eq. (4), we find, for any $\tau > 0$;

$$\left(i \frac{d}{d\tau} \right)^n G(\mathbf{k}, \tau) = -i\Theta(\tau)\langle 0| c_{\mathbf{k}} \mathcal{H}^n e^{-i\mathcal{H}\tau} c_{\mathbf{k}}^\dagger |0\rangle,$$

so that the sum rules simplify to

$$M_n(\mathbf{k}) = \langle 0| c_{\mathbf{k}} \mathcal{H}^n c_{\mathbf{k}}^\dagger |0\rangle. \tag{24}$$

These vacuum expectation values can be evaluated directly with some effort. We find $M_0(\mathbf{k}) = 1$, $M_1(\mathbf{k}) = \epsilon_{\mathbf{k}}$, $M_2(\mathbf{k}) = \epsilon_{\mathbf{k}}^2 + g^2$, $M_3(\mathbf{k}) = \varepsilon_{\mathbf{k}}^3 + 2g^2 \varepsilon_{\mathbf{k}} + g^2 \Omega$, etc.

One very important conclusion that can be drawn from this derivation is that these sum rules have the same functional dependence on the energy scales t, Ω, g anywhere in the parameter space. Of course, in various asymptotic regimes, different terms dominate the overall value [e.g., $M_2(\mathbf{k}) \approx g^2$ if $g \gg t$, while $M_2(\mathbf{k}) \approx \epsilon_{\mathbf{k}}^2$ if $g \ll t$]. However, this shows that if one can evaluate the sum rules *exactly* in any asymptotic regime, for instance by using perturbation theory, the results hold true *everywhere* in the parameter space, even where perturbation fails.

The second important conclusion one can draw from Eq. (24) is that each term in $M_n(\mathbf{k})$ is proportional to $t^p \Omega^m g^{n-m-p}$, where $0 \le p, m, n-m-p \le n$ are integers, i.e., the sum rule M_n is a polynomial of total order n in the energy scales of the problem. More complicated dependence on t, Ω, g, for example through $\exp(-g^2/\Omega^2)$, simply cannot appear (see below).

The first conclusion suggests an alternative derivation of the sum rules, which can also be used for the MA and SCBA sum rules. Namely, we use the diagrammatic perturbational expansion of the Green's function valid for $g \ll t$ to evaluate directly the integrals $\int_{-\infty}^{\infty} d\omega \, \omega^n G(\mathbf{k}, \omega)$ and retain the imaginary part. In this case, $G(\mathbf{k}, \omega) = \sum_{n=0}^{\infty} \sum_{i=1}^{s_n} D_{n,i}(\mathbf{k}, \omega)$, where $D_{n,i}(\mathbf{k}, \omega)$, $i = 1, s_n$ are all Green's function diagrams of order n, i.e., containing n phonon lines. The multiplicity $s_n = (2n-1)!! = 1 \times 3 \dots (2n-1)$. Each diagram $D_{n,i}(\mathbf{k}, \omega)$ is a product of $2n+1$ free propagators, summed over internal phonon momenta. Since for large frequency each $G_0(\mathbf{k}, \omega) \to (\omega + i\eta)^{-1}$, it follows that for $|\omega| \to \infty$, each $D_{n,i}(\mathbf{k}, \omega) \to g^{2n}/(\omega + i\eta)^{2n+1}$. Since any integrand that decreases faster than $1/\omega^2$ has a vanishing contribution to Eq. (23), it follows that the diagrams of order n only contribute to the sum rules $M_p(\mathbf{k})$ with $p \ge 2n$. Thus, even though $G(\mathbf{k}, \omega)$ is the sum of an infinite number of diagrams, only a finite number of them, of low order, contribute to any given sum rule and the calculation can be done. The same holds true for the MA sum rules, the only difference being that the self-energy parts in the Green's functions diagrams are replaced with the corresponding MA self-energy parts.

Let us analyze the differences between contributions of the exact and of the MA diagrams to the sum rules. It is straightforward to verify that

$$-\frac{1}{\pi} \operatorname{Im} \int_{-\infty}^{\infty} d\omega \, \omega^{2n} g^{2n} \prod_{i=1}^{2n+1} G_0(\mathbf{q}_i, \omega - \Omega_i) = g^{2n},$$

$$-\frac{1}{\pi} \operatorname{Im} \int_{-\infty}^{\infty} d\omega \, \omega^{2n+1} g^{2n} \prod_{i=1}^{2n+1} G_0(\mathbf{q}_i, \omega - \Omega_i)$$

$$= g^{2n} \sum_{i=1}^{2n+1} (\Omega_i + \epsilon_{\mathbf{q}_i})$$

and

$$-\frac{1}{\pi} \operatorname{Im} \int_{-\infty}^{\infty} d\omega \, \omega^{2n+2} g^{2n} \prod_{i=1}^{2n+1} G_0(\mathbf{q}_i, \omega - \Omega_i)$$

$$= g^{2n} \left[\sum_{i=1}^{2n+1} (\Omega_i + \epsilon_{\mathbf{q}_i})^2 + \sum_{i<j} (\Omega_i + \epsilon_{\mathbf{q}_i})(\Omega_j + \epsilon_{\mathbf{q}_j}) \right].$$

Both the exact and the MA diagrams of order n contain products of the general form $g^{2n} \Pi_{i=1}^{2n+1} G_0(\mathbf{q}_i, \omega - \Omega_i)$. Some of the free propagators are actually $G_0(\mathbf{k}, \omega)$ (always the first and the last one, but there can also be intermediary ones connecting proper self-energy parts). All other free propagators have momenta dependent on the phonon momenta, which are summed over (in the exact diagrams), or are individually averaged over (in the MA diagrams). Since there is one-to-one correspondence between the number of exact vs MA diagrams and their topologies, and since $\sum_{\mathbf{q}} \epsilon_{\mathbf{q}} = 0$, it follows that the exact and the MA diagrams of order n give precisely the same contributions to $M_{2n}(\mathbf{k})$ and $M_{2n+1}(\mathbf{k})$. Differences appear in the contribution to $M_{2n+2}(\mathbf{k})$ if there is at least one pair of propagators in any of the self-energy parts of the exact diagram that carry the same momenta. In this case, the corresponding $\epsilon_{\mathbf{q}_i} \epsilon_{\mathbf{q}_j}$ averages to $2dt^2$ when the sums over phonon momenta are carried for the exact diagrams; whereas these terms always average to zero for the MA diagrams. Since most free propagators in the self-energy parts have different momenta, such differences are quantitatively small. This is especially true for large phonon frequencies Ω, where the contributions proportional to Ω captured correctly by MA scale like some power of $2n+1$. This analysis can be continued for higher sum rules, with similar conclusions.

We can now summarize our findings. Only the zeroth order Green's function diagram [the free propagator $G_0(\mathbf{k}, \omega)$] contributes to $M_0(\mathbf{k})$ and $M_1(\mathbf{k})$. Since this is included correctly in the MA and the SCBA cases, both give the correct $M_0(\mathbf{k})$ and $M_1(\mathbf{k})$. In fact, this diagram contributes an $\epsilon_{\mathbf{k}}^n$ to $M_n(\mathbf{k})$, which is always the leading power in t contribution. The first order diagram is also exact in both the MA and SCBA cases, therefore both give the correct $M_2(\mathbf{k})$ and $M_3(\mathbf{k})$ sum rules as well. Differences appear from M_4 onwards. Because SCBA only keeps two out of the three second order Green's function diagrams, and five out of 15 third order diagrams, etc., the leading terms in g are $2g^4$ instead of
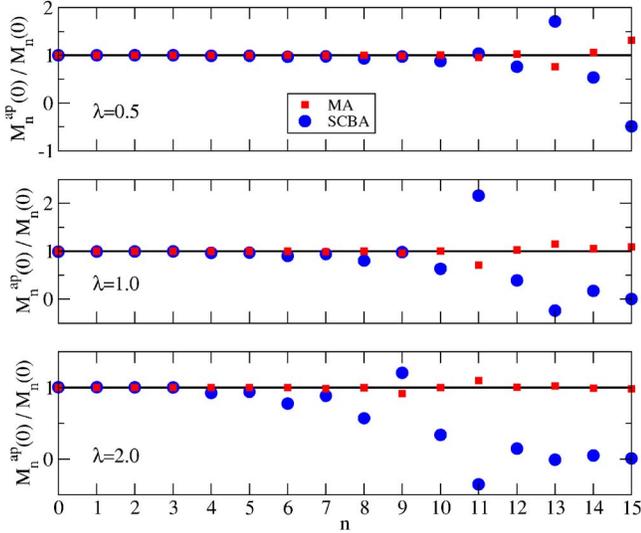
FIG. 2. (Color online) Ratio of MA (squares), respectively SCBA (circles) sum rules and the exact sum rules vs order n. Results are for 1D and $k=0$, $\Omega=0.5t$, and $\lambda=0.5$, 1, and 2.

FIG. 3. (Color online) Ratio of MA (squares), respectively SCBA (circles) sum rules and the exact sum rules vs order n. Results are for 1D and $k=\pi$, $\Omega=0.5t$, and $\lambda=0.5$, 1, and 2.

$3g^4$ in $M_4(\mathbf{k})$, $5g^6$ instead of $15g^6$ in $M_6(\mathbf{k})$, etc. This shows that SCBA fails badly all sum rules with $n>3$ in the strong coupling regime where the term proportional to g^{2n} gives the most significant contribution to $M_{2n}(\mathbf{k})$ (similar conclusions hold for odd sum rules). So even though SCBA always satisfies exactly the first four sum rules, it is a bad approximation for large g, where big discrepancies appear for $n>3$.

MA satisfies $M_4(\mathbf{k})$ and $M_5(\mathbf{k})$ exactly as well, because these only depend on having the correct number and topology for the second order diagrams. MA fails from $M_6(\mathbf{k})$ onward, however, in a very different manner than SCBA. The leading term in g^6 has the correct prefactor, because MA has the correct number of third order diagrams. The error comes from the second order diagram containing the noncrossed self-energy diagram, as discussed. Indeed, instead of the exact sum rule

$$M_6(\mathbf{k}) = \varepsilon_{\mathbf{k}}^6 + g^2[5\varepsilon_{\mathbf{k}}^4 + 6t^2(2d^2-d) + 4\varepsilon_{\mathbf{k}}^3\Omega + 3\varepsilon_{\mathbf{k}}^2\Omega^2 \\ + 6dt^2(\varepsilon_{\mathbf{k}}^2 + \varepsilon_{\mathbf{k}}\Omega + 2\Omega^2) + 2\varepsilon_{\mathbf{k}}\Omega^3 + \Omega^4] \\ + g^4(18dt^2 + 12\varepsilon_{\mathbf{k}}^2 + 22\varepsilon_{\mathbf{k}}\Omega + 25\Omega^2) + 15g^6,$$

$$(25)$$

MA finds a sum rule $M_6^{\text{MA}}(\mathbf{k})=M_6(\mathbf{k})-2dt^2g^4$. The leading terms in the $g\ll t$ and $g\gg t$ limits are always exact (as expected, since MA becomes exact in these limits), and this is true for all orders n. For $n\geq6$ some of the cross terms are missing, but these are a minority related to noncrossed diagrams, as explained. We therefore expect the MA sum rules to remain highly accurate for larger n values. That this is indeed true for higher sum rules is shown numerically in Figs. 2 and 3, where we plot the ratio of the MA, respectively, SCBA sum rules, and the exact sum rules of the same order n. The results shown are for one dimension (1D) and $k=0$ and π, but similar trends are found in the other cases. For $k=0$, all the spectral weight is at negative frequencies,

therefore $M_n(0)$ alternate signs for even/odd n, and this is reflected in the nonmonotonic behavior with n. For $k=\pi$, most of the weight is at positive frequencies and sum rules are always positive. The magnitude of the exact sum rules increases roughly exponentially with n; for instance, for $\lambda=2$ and $\Omega=0.5t$, $M_{14}(0)=119\,516\,000$. For $k=0$ and $\lambda=0.5$, both MA and SCBA are reasonably accurate, with a slight edge for MA at higher n. However, MA is clearly much more accurate for all the other cases shown, and its accuracy is expected to improve even more as one moves further into the asymptotic regions of weak or strong coupling.

Before ending this section, one more issue needs to be addressed. It is obvious that the MA sum rules must capture correctly the contributions to $M_n(\mathbf{k})$ proportional to t^n and to g^n, since MA is exact for both $g=0$ and $t=0$. One may assume that this alone is sufficient for a good interpolation at finite t and g. That this is not so is shown by the generalized LF approximation, which is also exact for $t=0$ or $g=0$. However, in this approximation one finds $M_0^{\text{LF}}(\mathbf{k})=1, M_1^{\text{LF}}(\mathbf{k})=\varepsilon_{\mathbf{k}}\exp(-g^2/\Omega^2)$, etc. M_1 and all higher sum rules have unacceptable dependence on the energy scales g and Ω (see second observation above), even though they become exact asymptotically. As shown in the next section, this approximation indeed performs rather poorly for finite t and g.

We conclude that while MA satisfies exactly the first six sum rules, it remains accurate for all higher order sum rules, and is asymptotically exact. This is another argument in favor of the accuracy of this approximation over the whole parameter space. In the next section, we compare the MA predictions to those of existing numerical simulations to further support this claim.

IV. RESULTS

We first list the explicit expressions of the momentum averaged Green's function $\bar{g}_0(\omega)$ of Eq. (17). For nearest-

neighbor hopping, it is straightforward to derive

$$\bar{g}_0^{1D}(\omega) = \frac{\text{sgn}(\omega)}{\sqrt{(\omega + i\eta)^2 - 4t^2}},$$

$$\bar{g}_0^{2D}(\omega) = \frac{2}{\pi(\omega + i\eta)} \mathcal{K}\left(\frac{4t}{\omega + i\eta}\right),$$

and

$$\bar{g}_0^{3D}(\omega) = \frac{1}{2\pi^2 t} \int_0^\pi dk_z \, \text{sgn} \, \nu |\nu| \mathcal{K}(\nu),$$

respectively, where

$$\nu = \frac{4t}{\omega + 2t\cos(k_z a) + i\eta}$$

and

$$\mathcal{K}(\nu) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - \nu^2\cos^2\phi}}$$

is the complete elliptical function of the first kind.[46] These integrals can be performed numerically very efficiently. More generally, for any free electron dispersion $\epsilon_\mathbf{k}$ to which corresponds the free electron density of states (DOS) $\rho_0(\epsilon)$ $= (1/N)\Sigma_\mathbf{k}\delta(\epsilon - \epsilon_\mathbf{k})$, we have $\bar{g}_0(\omega) = \int_{-\infty}^{\infty} d\epsilon \, \rho_0(\epsilon)(\omega - \epsilon + i\eta)^{-1}$ (also see the Appendix).

In our calculations we employ a small but finite value for η. This moves the poles of the Green's function off of the real axis and changes the δ peaks of the spectral weight $A(\mathbf{k},\omega) = \Sigma_\alpha |\langle\alpha|c_\mathbf{k}^\dagger|0\rangle|^2\delta(\omega - E_\alpha)$ into Lorentzians. In practice, it is necessary to choose η small enough to allow detection of the Lorentzian peaks in regimes where the qp weight Z is extremely small. This is relevant when ground-state and low-energy properties are investigated. In such cases, we typically use $\eta/t \simeq 10^{-5}$. For cases where $A(\mathbf{k},\omega)$ is plotted for a large frequency range, we use η large enough to allow detection of the Lorentzian peaks with modest resolution in the step $\delta\omega$. The values used in these cases are listed explicitly.

The self-energy $\Sigma_{MA}(\omega)$ is then calculated easily from Eq. (19) by truncating the continued fraction to a high-enough level. For an error of order ϵ, it is necessary to go to a level with n such that $ng^2\bar{g}_0(\omega - n\Omega)\bar{g}_0(\omega - (n+1)\Omega) < \epsilon$. Using the fact that for large enough n we can approximate $\bar{g}_0(\omega - n\Omega) \approx \bar{g}_0(\omega - (n+1)\Omega) \approx -1/(n\Omega)$, it follows that we must have

$$n > \frac{1}{\epsilon}\frac{g^2}{\Omega^2}.$$

This result is expected, since g^2/Ω^2 is roughly the average number of phonons in the polaron cloud (see below). This condition shows that all diagrams with at least that many phonons have to be included. In practice, we always use n large enough so that the change in $\Sigma_{MA}(\omega)$ after doubling n is below a threshold much smaller than η. All MA error bars in the figures we show are less than the thickness of lines/symbols used for the plots.

With an explicit form for $\Sigma_{MA}(\omega)$ we are now in a position to calculate the Green's function $G_{MA}(\mathbf{k},\omega)$ and extract various polaron properties.

A. Polaron ground state properties

We begin by discussing polaron GS properties. Most of these are already known from numerical studies, but they give us an opportunity to further test the accuracy of the MA approximation. Given the simplicity and efficiency of the MA approximation, we can also investigate higher dimensionality and larger parameter ranges than typical numerically intensive approaches. In this section, we use for comparison 1D and 2D numerical results obtained from diagrammatic quantum Monte Carlo (QMC) simulations from Ref. 32 unless otherwise noted.

For $k=0$, we track the energy and weight of the lowest pole in the spectral weight, which give the ground-state energy E_0 and the ground-state quasiparticle weight Z_0 $= |\langle GS|c_{\mathbf{k}=0}^\dagger|0\rangle|^2$. Using the Hellmann-Feynman theorem,[47] we then find the average number of phonons in the ground state to be

$$N_{ph} \equiv \left\langle GS \left| \sum_\mathbf{q} b_\mathbf{q}^\dagger b_\mathbf{q} \right| GS \right\rangle = \frac{\partial E_0}{\partial\Omega}. \tag{26}$$

Note that one can also calculate the correlation function

$$\left\langle GS \left| \sum_i c_i^\dagger c_i (b_i^\dagger + b_i) \right| GS \right\rangle = \frac{\partial E_0}{\partial g} \tag{27}$$

just as easily. We do not show it here because we do not have the corresponding numerical data for the comparison; however, some typical results for this quantity are shown at the end of this section. Also note that all these quantities can be calculated similarly for other eigenstates. We will show such results in other sections.

We also show the effective mass, m^*. Because the MA self-energy is momentum independent for this simple Holstein model, one has[43]

$$\frac{m^*}{m} = \frac{1}{Z_0} = 1 - \left.\frac{d\Sigma_{MA}(\omega)}{d\omega}\right|_{\omega=E_0}. \tag{28}$$

This result also gives us a consistency check on our calculations. For MA, we generally show effective mass results obtained from the first equality.

A comparison of the 1D results for these four quantities as obtained with QMC (black circles) and the different approximations is shown in Fig. 4. As expected, SCBA (blue line) fares well at small couplings λ but very poorly at strong couplings. The generalized LF (green line) is asymptotically exact, but quite wrong at finite λ. Because these are GS properties, we can also use perturbation theory in the two asymptotic limits to estimate them. At weak coupling we use the Rayleigh-Schrödinger (RS) perturbation theory (violet line), which gives the lowest energy for a state with total momentum $\mathbf{k}$ as[48]
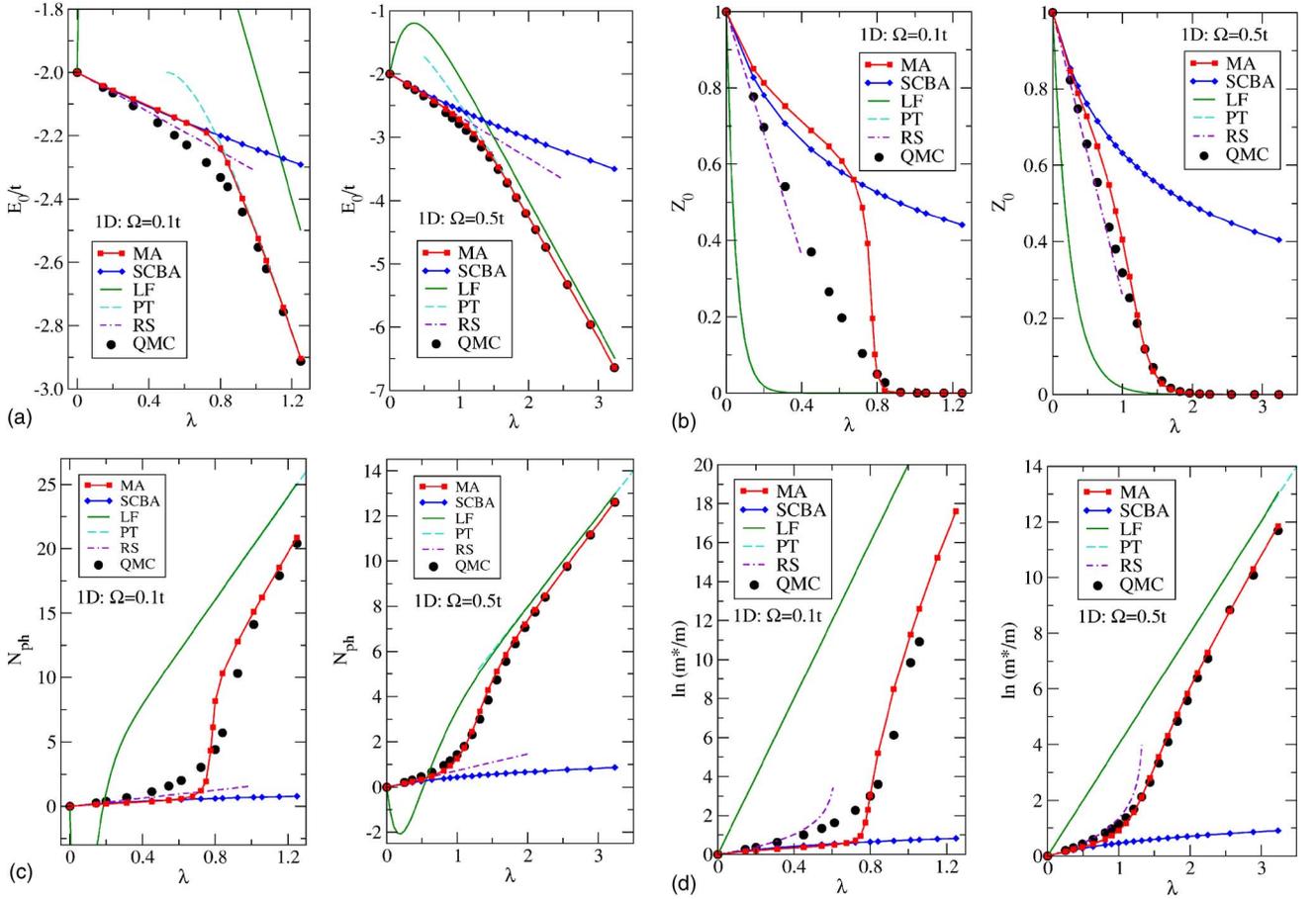
FIG. 4. (Color online) Ground state results in 1D. Shown as a function of the coupling λ are the ground state energy E_0, the qp weight Z_0, the average number of phonons N_{ph}, and the effective mass m^* on a logarithmic scale. The left panels correspond to $\Omega/t=0.1$ and the right ones to $\Omega/t=0.5$.

$$E_{\mathbf{k}} = \varepsilon_{\mathbf{k}} - \frac{1}{N}\sum_{\mathbf{q}} \frac{g^2}{\Omega + \varepsilon_{\mathbf{k-q}} - \varepsilon_{\mathbf{k}}}. \tag{29}$$

At strong couplings, we use the second order perturbation theory (PT) result (cyan line):[16,37]

$$E_{\mathbf{k}} = -\frac{g^2}{\Omega} + \epsilon_{\mathbf{k}} e^{-g^2/\Omega^2} - \frac{d\Omega t^2}{g^2}. \tag{30}$$

The GS energy is simply $E_0 = E_{\mathbf{k}=0}$. N_{ph} is obtained from E_0 as before; m^* can be evaluated from the second derivative of $E_{\mathbf{k}}$ with respect to $\mathbf{k}$, and Z_0 is extracted from the effective mass [Eq. (28)].

Figure 4 shows that one or the other of these perturbational values describe the GS energy E_0 quite well, especially for the larger Ω value. However, the agreement for the other quantities is somewhat poorer, especially at strong couplings (PT and LF give identical results for Z_0). Part of the reason is that the t^2 term in Eq. (30) has in fact a more complicated dependence of g and Ω, which is only asymptotically equal to the one used here.[16] More importantly, neither perturbational theory describes well the crossover regime, or can be easily applied to high energy states.

Clearly, MA (red line with red symbols, for easier comparison with QMC data) has the best agreement with the

QMC data. As expected from the sum rule analysis and the discussion on the convergence of Σ_{MA}, MA improves for larger Ω. The worst disagreements we ever found are the ones shown in the intermediary λ regime for $\Omega/t=0.1$. Even there, the error in the GS energy is always below 5%. The qp weight has a more significant disagreement; however, note that it indeed becomes asymptotically correct for $\lambda < 0.2$ and $\lambda > 0.8$. The second claim is supported by the m^* data, which indeed shows convergence towards the expected PT values. Most importantly, even though it is quantitatively somewhat wrong in this intermediary regime for small Ω, the MA approximation is the only one that clearly captures the crossover from the large to the small polaron, which is accompanied by the collapse of the qp weight and the increase in the number of phonons trapped in the cloud, and thus of the effective polaron mass.

The comparison with the QMC data for 2D polarons is shown in Fig. 5. Here MA gives excellent agreement at all couplings λ. [Some of this is because the results correspond to larger $\Omega/(4dt)$ values. We do not show $\Omega/t=0.2$ results because we lack QMC data for comparison. Given the sum rule arguments, we expect the agreement to be better than for the 1D, $\Omega/t=0.1$ case.] The agreement is all the more remarkable when considering what a numerically trivial task it is to evaluate the MA results compared to the QMC simula-
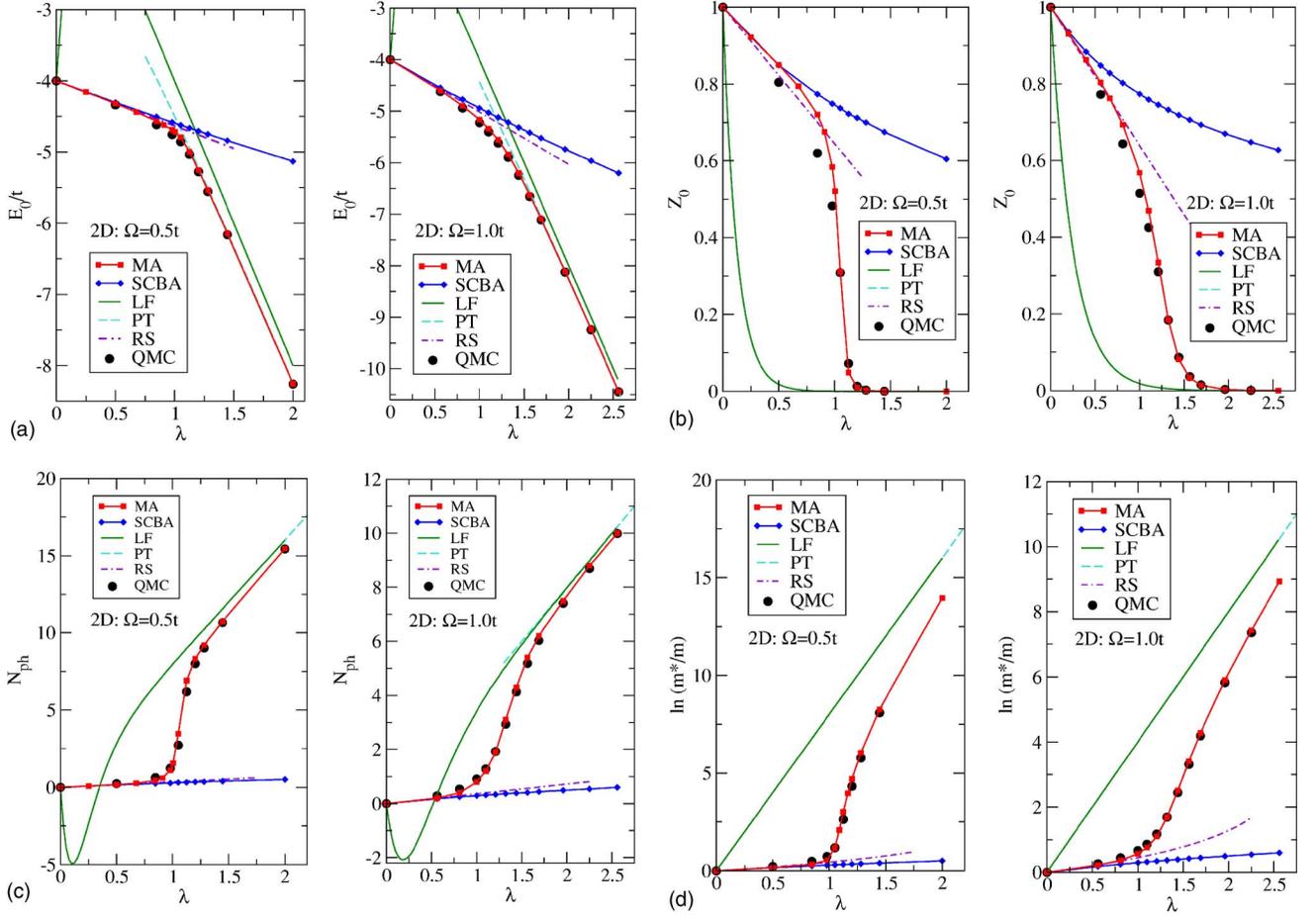
FIG. 5. (Color online) Ground state results in 2D. Shown as a function of the coupling λ are the ground state energy E_0, the qp weight Z_0, the average number of phonons N_{ph}, and the effective mass m^* on a logarithmic scale. The left panels correspond to $\Omega/t=0.5$ and the right ones to $\Omega/t=1.0$.

tions. The physics is similar to that seen in 1D; however, the crossover from the large (light) polaron at weak couplings to the small (heavy) polaron at strong couplings becomes somewhat sharper, especially for smaller Ω/t values.

Given the simplicity of MA, we can also generate contour plots of these quantities as a function of both λ and Ω/t, and investigate the entire parameter space. Such results are shown in Figs. 6 and 7 for $d=1$ and 2, respectively. The crossover from the large to the small polaron can now be tracked (for instance, from the collapse of Z_0) and one can quantitatively trust the results to a high degree. The crossover occurs for $\lambda \approx 1$, slowly shifting to higher values of λ as Ω increases, in agreement with numerical results.[27,28]

We also show MA results in 3D; see Fig. 8. There are very few numerical three-dimensional results available for the Holstein polaron, due to the computational effort required to investigate such cases. Ground state properties of the three-dimensional polaron have been calculated using variational methods[21] and QMC (Ref. 26) for a single phonon frequency of $\Omega/t=1.0$. Comparison of the MA ground state energy and effective mass to those found in Ref. 26 shows excellent agreement. The good agreement with one or the other perturbational theories for most coupling strengths further supports the accuracy of MA in 3D. The crossover from large to

small polaron becomes even sharper, especially for lower Ω. It is still located in the neighborhood of $\lambda \approx 1$. This influence of dimensionality on the crossover regime of the Holstein polaron has also been reported by Ku *et al.* in Ref. 21.

Using the Hellmann-Feynman theorem like in Eqs. (26) and (27) also allows us to separate the individual contributions of

$$T_{elec} = \left\langle GS \left| \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} c_{\mathbf{k}}^{\dagger} c_{\mathbf{k}} \right| GS \right\rangle,$$

$$E_{ph} = \left\langle GS \left| \Omega \sum_{\mathbf{q}} b_{\mathbf{q}}^{\dagger} b_{\mathbf{q}} \right| GS \right\rangle = \Omega N_{ph},$$

and

$$V_{corr} = \left\langle GS \left| g \sum_{i} c_i^{\dagger} c_i (b_i^{\dagger} + b_i) \right| GS \right\rangle$$

to the total GS energy. Plots of these individual contributions as a function of coupling strength λ are shown in Fig. 9 for $d=1$, 2, and 3. The results for the kinetic energy of the electron are in good agreement with those found in Ref. 27 for $d=1$, 2, and 3, using QMC cluster sizes of 32, 12, and 6, respectively. We expect that the agreement for $d=2,3$ would
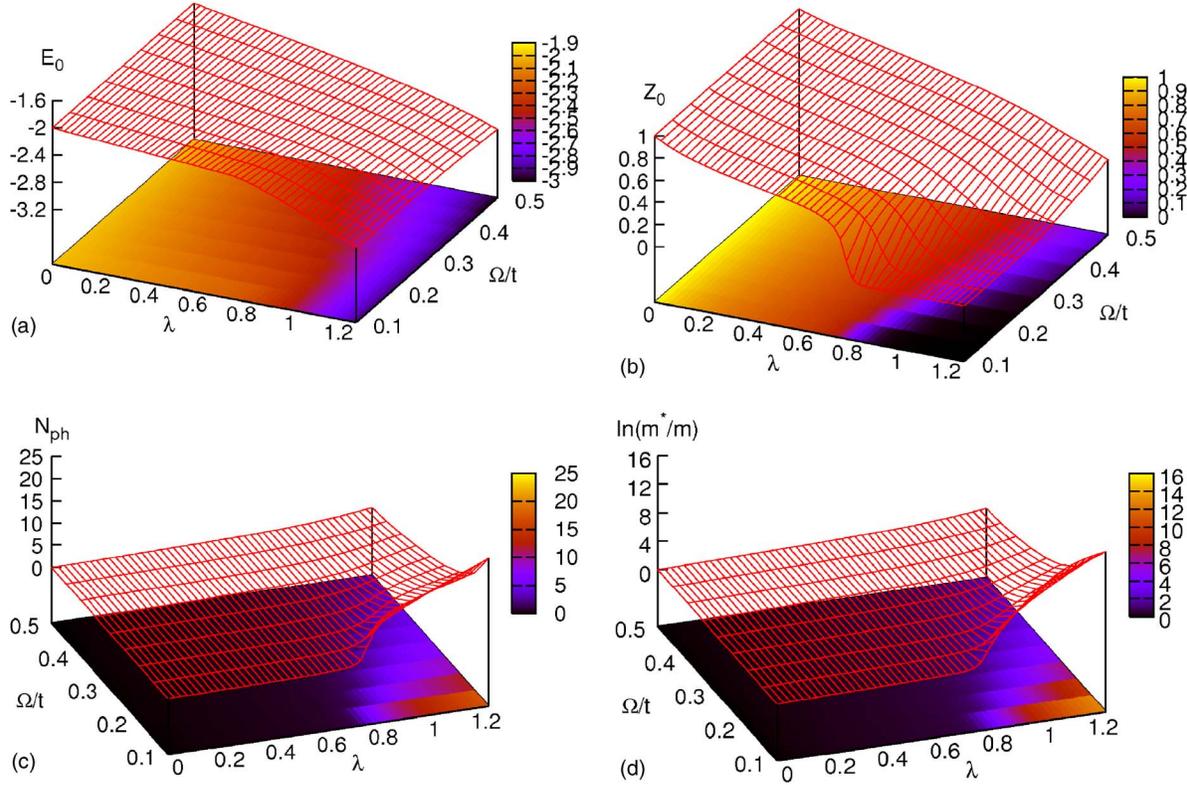
FIG. 6. (Color online) Ground state energy E_0, the qp weight Z_0, the average number of phonons N_{ph}, and the effective mass m^* as a function of λ and of Ω/t, for $d=1$.

be further improved with larger cluster sizes; however, this is computationally prohibitive for QMC simulations.[27] As expected, the kinetic energy is close to $-2dt$ at weak couplings, but it becomes vanishingly small in the strong coupling limit, where the polaron becomes very heavy. The phonon energy E_{ph} increases roughly like g^2/Ω in the strong coupling limit, where $N_{\mathrm{ph}} \approx g^2/\Omega^2$. It follows that the decrease in the total GS energy is due to the interaction term, as expected. Note that this energy is proportional to the correlation of Eq. (27). Since $E_0 \approx -g^2/\Omega$ in the strong coupling limit (see agreement with PT results), it follows that this correlation becomes asymptotically equal to $-2g/\Omega$ in the strong coupling limit.

B. Low energy states: Momentum dependence

We can also calculate the same properties for the lowest energy state corresponding to each given momentum $\mathbf{k} \neq 0$ to find the low-energy behavior of the polarons. In this section, we present comparisons with available QMC results[32] for 1D and 2D systems, and with variational results for the 3D systems.[21]

In Fig. 10 we show 1D results for the polaron dispersion E_k, the associated qp weight Z_k, and average phonon number $N_{\mathrm{ph}}(k)$ for two couplings. For the very weak $\lambda=0.25$, we see that MA and SCBA are equally good at small k, however, for large k, MA overestimates the energy, such that the continuum which is expected to appear at a distance Ω above the GS energy E_0 is in this case pushed somewhat higher. We

will return to a discussion of this discrepancy later on. As expected, the qp weight is large for small k, where the main contribution to the eigenstate comes from the free electron state $c_k^\dagger|0\rangle$. Z_k goes to zero for larger k, since these are primarily linear combinations of states of type $c_{k-q}^\dagger b_q^\dagger|0\rangle$, as confirmed also by the average phonon number of about unity. Note that the RS perturbation works well for small k. However, it breaks down at a finite k where $\epsilon_k \approx -2t+\Omega$, i.e., the free electron dispersion crosses into the continuum of electron plus one phonon states. Here, RS predicts an unphysical peak in the polaron dispersion [denominator of Eq. (29)] and it fails for larger k.

The second coupling $\lambda=1$ is roughly in the crossover regime; see also Fig. 4. Here MA gives a much better agreement with QMC than SCBA or the perturbational theories. The polaron bandwidth is already renormalized and well below the weak coupling value of Ω. In fact, as we show later, there is another bound state between these states and the continuum. This is not captured in SCBA, which always predicts a polaron bandwidth of Ω with a continuum above, and roughly between zero and one average number of phonons as k increases from 0 to π. For the strong coupling $\lambda=1.96$, low-energy properties become almost k independent, as expected since the Lang-Firsov impurity limit is being approached.

A second such comparison is possible for the 2D case with $\Omega=0.5t$ and $\lambda=0.845$, where QMC data is available; see Fig. 11. This coupling is on the weak side of the crossover, with the GS qp weight still large, $Z_0 \approx 0.6$. Here MA is already doing better than SCBA. As in the 1D weak-coupling
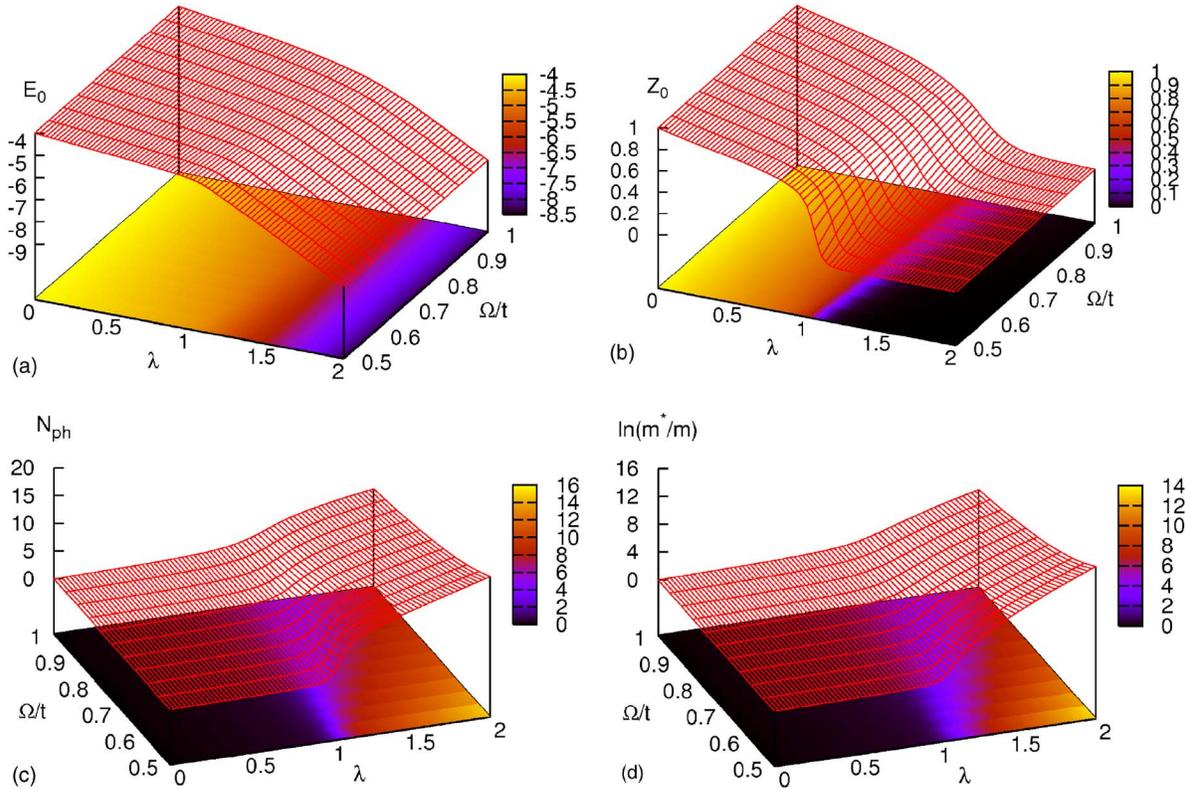
FIG. 7. (Color online) Ground state energy E_0, the qp weight Z_0, the average number of phonons N_{ph}, and the effective mass m^* as a function of λ and of Ω/t, for $d=2$.

case, one can again see that the MA polaron bandwidth is slightly larger than Ω. MA also somewhat overestimates the average number of phonons for large **k** values. Overall, given that all this data is in the crossover regime where the MA is at its worst, one can conclude that MA is also reasonably accurate in capturing low-energy polaron behavior.

In the 3D case, momentum-dependent data is available for energies and qp weights, when $\Omega=2.0t$ and $\lambda=1.0833$.[21] Although this coupling is in the center of the crossover regime, MA is in excellent agreement with the variational results for all momenta **k** along the high-symmetry cuts of the Brillouin zone, as shown in Fig. 12. The agreement is particularly good in this case because MA is highly accurate for large phonon energies.

C. High-energy states

The main motivation in trying to find an approximation for the Green's function is that this quantity gives not only low-energy information, but the whole spectrum. We now compare MA predictions with various high-energy results available in the literature. Unfortunately, there are much fewer of these, because the computational effort to obtain the whole spectrum through the usual numerical approaches is generally forbidding.

We begin with a comparison against exact diagonalization (ED) 1D data, from Ref. 18. The results are shown in Fig. 13. As already discussed, for weak coupling there is a continuum starting at $E_0+\Omega$, but for stronger coupling a so-

called second bound state appears below the continuum. In Fig. 13(a) we track the energy E_1 of the second $k=0$ state. For small couplings, the data actually shows the maximum DOS in the continuum, not its edge (the maximum is generally located close to the lower edge. This data shows again that MA somewhat overestimates this energy, which should be $\approx\Omega$). When $E_1<E_0+\Omega$, there is a true discrete state. Note that panel (a) is in very good quantitative agreement with similar data shown in Fig. 8 of Ref. 18. The only difference is for strong coupling, where the ED data shows $E_1>E_0+\Omega$ again, however, with significant finite-size dependence on the chosen Hilbert space cutoff.

We can thus find the coupling g/t where the second bound state appears, for different values of Ω/t. This line is shown in panel (b), together with the ED results. The agreement between the two data sets is excellent, even at small Ω/t values where we expect MA to be less accurate. We also show in panel (c) the qp weight of this second bound state, where stable. This data is not given in Ref. 18; however, one QMC point is available in Ref. 32, in good agreement with the MA prediction.

We now move to comparisons for the entire spectral weight $A(k,\omega)$. In Figs. 14–16, we show comparisons for a 1D system with $\Omega=0.4t$ and three different coupling strengths $\lambda=0.5$, 1, and 2, respectively. In each case, data for five values of k, namely 0, $\pi/4$, $\pi/2$, $3\pi/4$, and π, are shown. The numerical data (black line) is obtained using a variational method by De Filippis et al.[23] Numerical data obtained through exact diagonalization of a finite system and by QMC, for the same parameters but somewhat different k
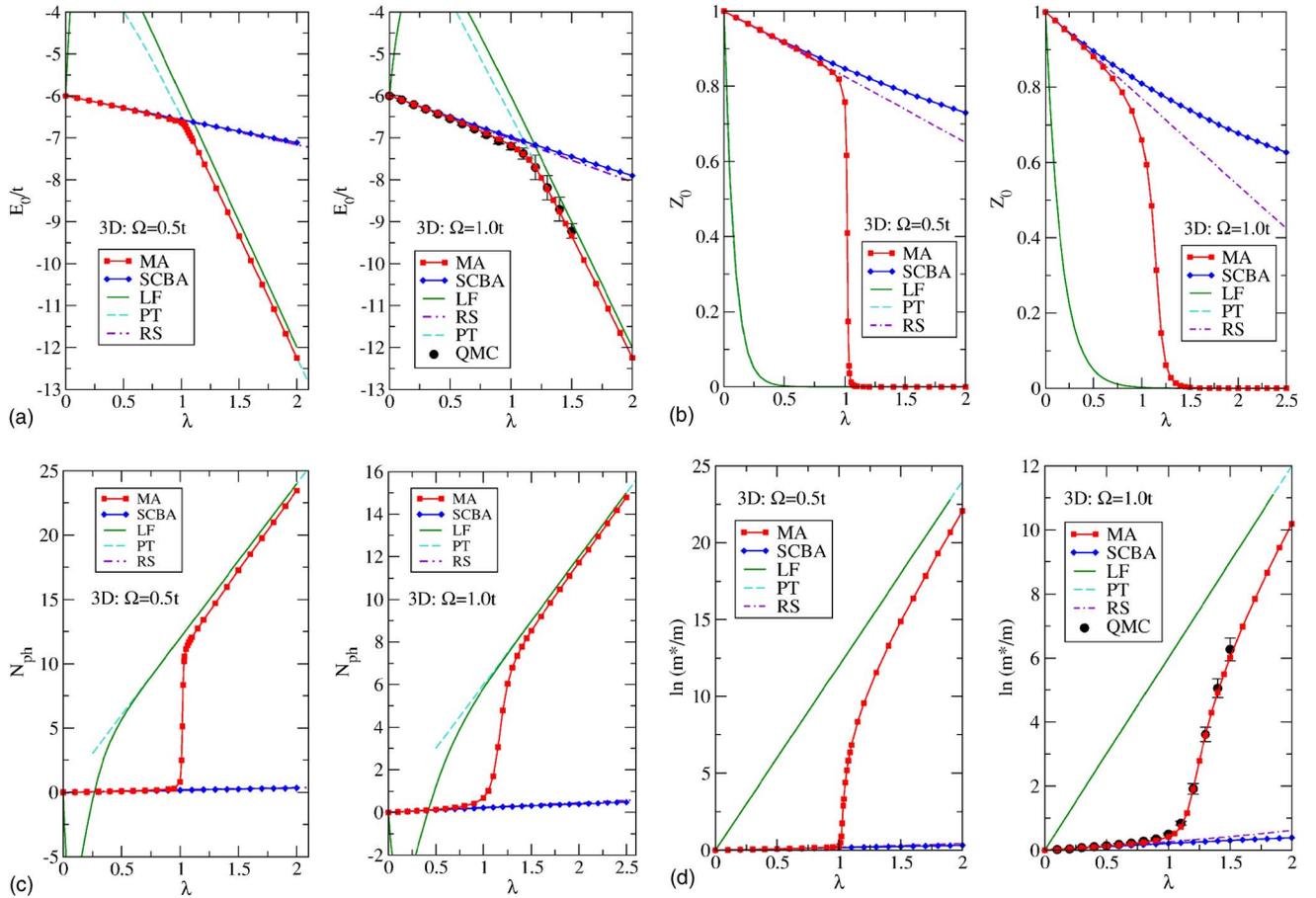
FIG. 8. (Color online) Ground state results in 3D. Shown as a function of the coupling λ are the ground state energy E_0, the qp weight Z_0, the average number of phonons N_{ph}, and the effective mass m^* on a logarithmic scale. The left/right panels are for $\Omega/t=0.5$ and 1. QMC data is from Ref. 26.

values, is also presented by Hohenadler *et al.* in Refs. 39 and 40. These sets of numerical data are in good agreement with one another.

In all three cases, the agreement between MA results and the numerical data is very good. As expected, it is best for the largest λ, but even for the smaller λ values, which are just below and within the crossover region, the agreement is very satisfactory. For $\lambda=0.5$ and $k=0$ (upper panel of Fig. 14) we see the polaron state as a Lorentzian peak (a broadening $\eta=0.1\ \Omega$ was used), which accounts for most of the weight, and a small continuum at a higher energy. MA overestimates the gap between the two, which should be Ω. As k increases, the polaron peak disperses but also loses significant weight, as discussed in the previous section. Most of the weight is now in the high energy continuum, located roughly near the corresponding ϵ_k value. This simply shows that these higher energy states are not significantly affected by this rather weak coupling. The VM data shows somewhat more structure in these continua than the MA data, but most of the weight occupies similar frequency ranges.

For $\lambda=1$ and $k=0$ (upper panel of Fig. 15), the MA data shows three Lorentzian peaks plus a continuum starting at $\omega/t=-1.6$. For the rather large η used, it is hard to distinguish which peaks come from individual poles, and which are true continua. This can be easily done by studying their

behavior as a function of the broadening η, as shown in Fig. 17. The height of peaks corresponding to discrete states scales precisely like $1/\eta$, as expected for Lorentzians. The continuum is affected very little by changes in η, except the peak near its lower edge where the finite η smoothes out a singularity in the DOS. Since this singularity is not of the $1/\omega$ type, its scaling with η is different from that of the Lorentzians. The two lower states are closer to one another than Ω, however, the MA data shows no sign of the continuum that is expected to start at $E_0+\Omega$. Note that the numerical data in Fig. 15 shows more structure that could be consistent with this continuum. We will address the issue of this continuum below. As k is increased (see Fig. 15), the low-energy peaks show some dispersion, but with a strongly renormalized bandwidth. At higher k most weight shifts again at high energies, in a rather broad continuum. Finally, for $\lambda=2$ and $k=0$, Fig. 16 shows even more discrete peaks spaced by Ω. The GS is at $E_0\approx-4.25t$, but its weight is so small that it cannot be seen on this scale, unless η is decreased significantly. A continuum is seen above $\omega=-1.6t$. As k is now increased, there is almost no dispersion of the discrete peaks; however, the weight shifts again to an even broader high-energy continuum.

The issue of the continuum at $E_0+\Omega$ in the exact case and of its absence in the MA approximation for moderate and
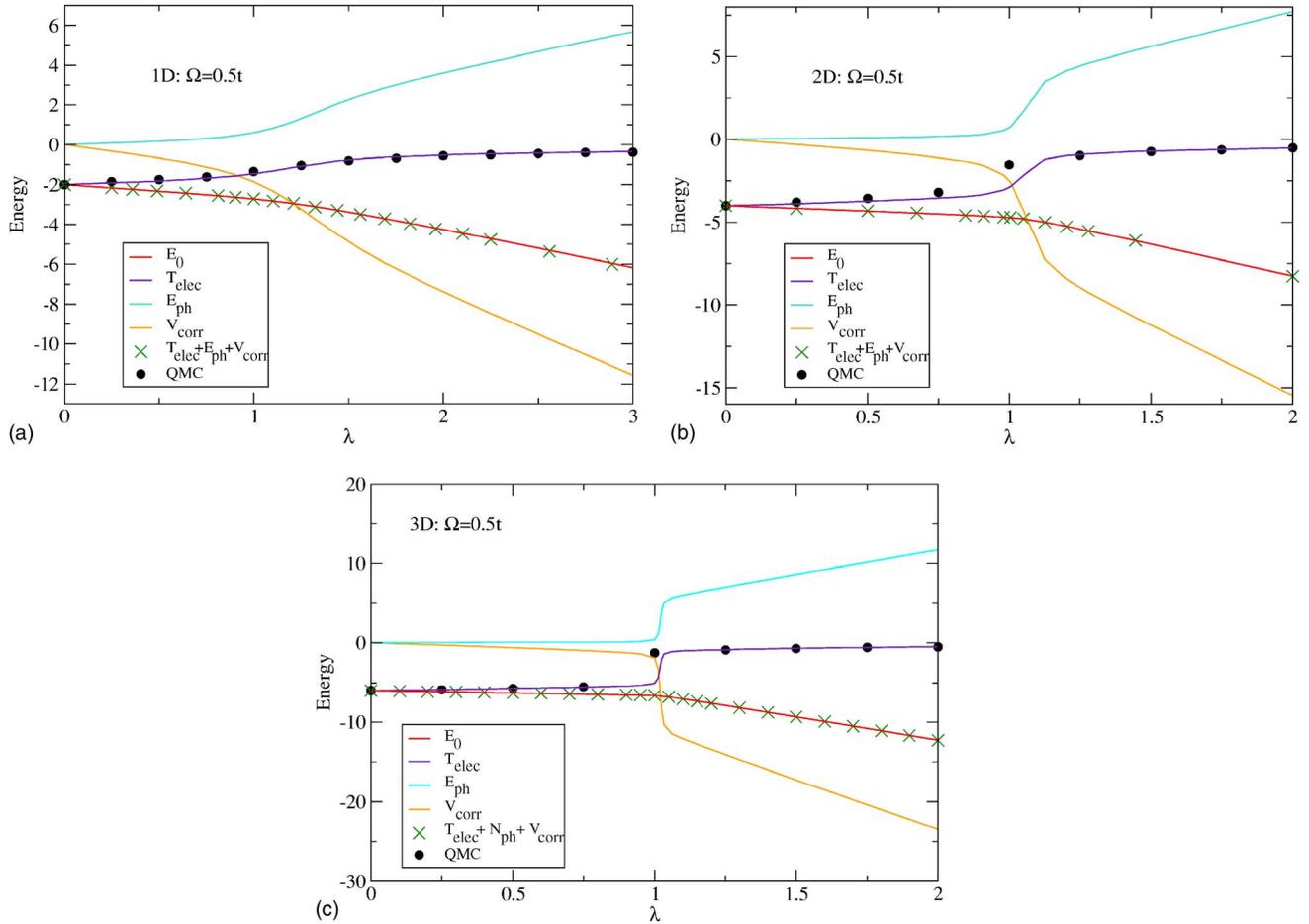
FIG. 9. (Color online) GS expectation values for the electron kinetic energy (violet), the phonon energy (cyan), and the electron-phonon interaction (orange) as a function of λ, for $\Omega=0.5t$ and $d=1$, 2, and 3. Circles represent numerical results for the kinetic energy, from Ref. 27.

large couplings can be understood from Fig. 18. Here we show a comparison of $A(k=0,\omega)$ in 2D obtained from exact diagonalization,[41] vs MA results, for various couplings λ. The first remark is that MA captures quite well all the large-weight features, both as far as their energy and their weight are concerned. This is expected, given the good sum rules agreement demonstrated previously. However, the ED results clearly show more states than MA predicts. There is always a low-weight peak at precisely $E_0+\Omega$ (the GS and this state are marked by arrows in the insets). For couplings of up to $\lambda \approx 1$, this peak is followed by several nearby peaks with comparably low weight, which can be argued to be part of the expected continuum. For larger λ, however, only the state at $E_0+\Omega$ can still be seen, although more states suggesting more continua are seen between the large higher-energy peaks. The gradual disappearance of the first continuum is not surprising, since one expects its width to narrow exponentially as the coupling increases. Moreover, one expects that the largest contribution to this continuum is from states with one or more phonons, explaining their low qp weight.

As far as MA is concerned, Fig. 18 suggests that for couplings where there is more than one discrete state, the very little weight in the $E_0+\Omega$ and similar higher-energy continua is absorbed in the discrete states predicted by MA. This is

consistent with the systematic upshift of the MA peaks compared to the ED data.

In fact, it is straightfoward to see that the MA approximation can only predict a continuum starting at $-2dt+\Omega$. A continuum is signaled by a finite imaginary part of $\Sigma_{MA}(\omega)$, and the lowest frequency where this can occur is that for which $\bar{g}_0(\omega-\Omega)$ acquires a finite imaginary part [see Eq. (19)]. However, the imaginary part of $\bar{g}_0(\omega)$ is proportional to the total density of states of the free model, i.e., it is finite for $\omega \in [-2dt,2dt]$ for nearest-neighbor hopping. It follows that the MA continuum always starts precisely at $-2dt+\Omega$. This explains why for small λ, where there is only one peak below this continuum, the gap between the two is somewhat larger that the expected Ω value: The GS energy decreases below $-2dt$ with increasing λ, whereas the continuum edge is pinned at $-2dt+\Omega$, in the MA approximation. As the coupling increases, bound states start to split from this continuum, and spectral weight is shifted to lower energies, in good agreement with the sum-rule predictions of the exact solution. These new bound states have to account for the (small) weight that is present in lower energy continua, in the exact solution, and this is precisely what Fig. 18 shows. Clearly, a self-energy that would account for these continua
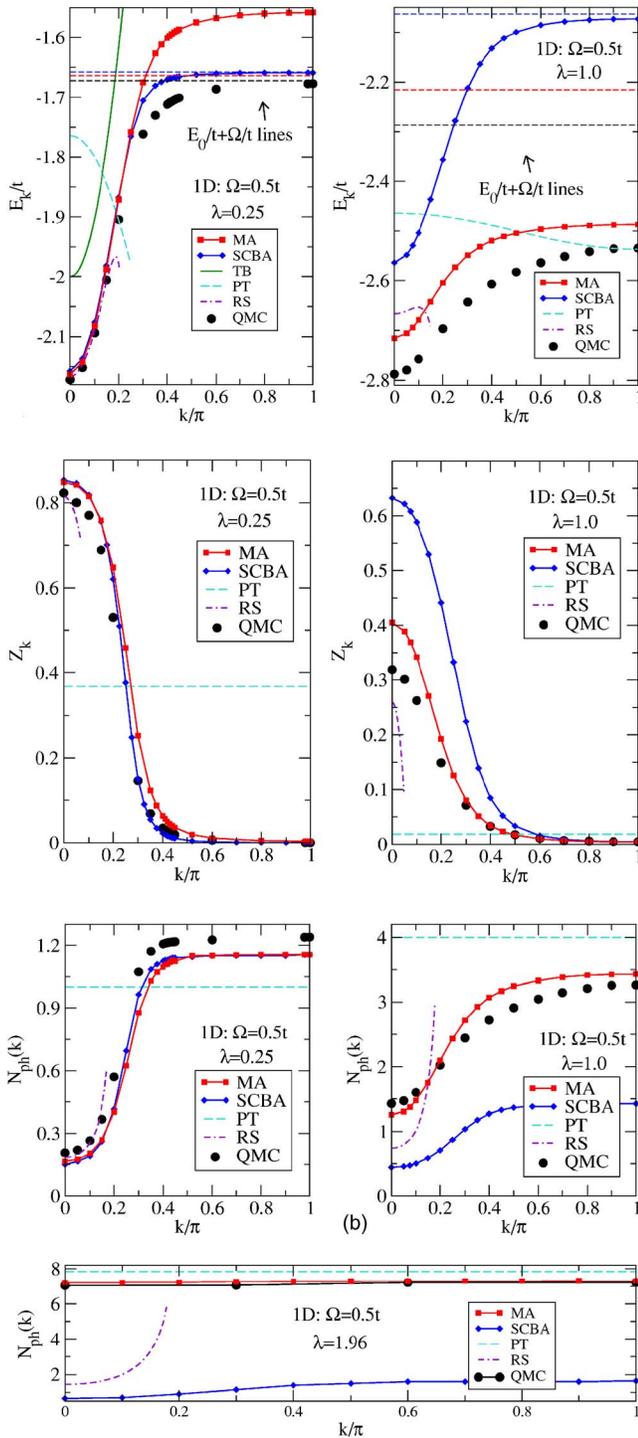
FIG. 10. (Color online) Lowest polaron eigenenergy for a given k, E_k, and the corresponding qp weight Z_k and average phonon number $N_{ph}(k)$. Results are for 1D, $\Omega/t=0.5$, and $\lambda=0.25$ and 1 (for E_k and Z_k) as well as 1.96, for $N_{ph}(k)$. Only half of the BZ is shown.

as well would have to be a lot more complicated than that of Eq. (19).

This shows again that MA is remarkably successful in predicting the main features of the Green's function, given its simplicity and trivial numerical cost. This should make it (and generalizations of it to other models) of large interest for comparison against experiments.
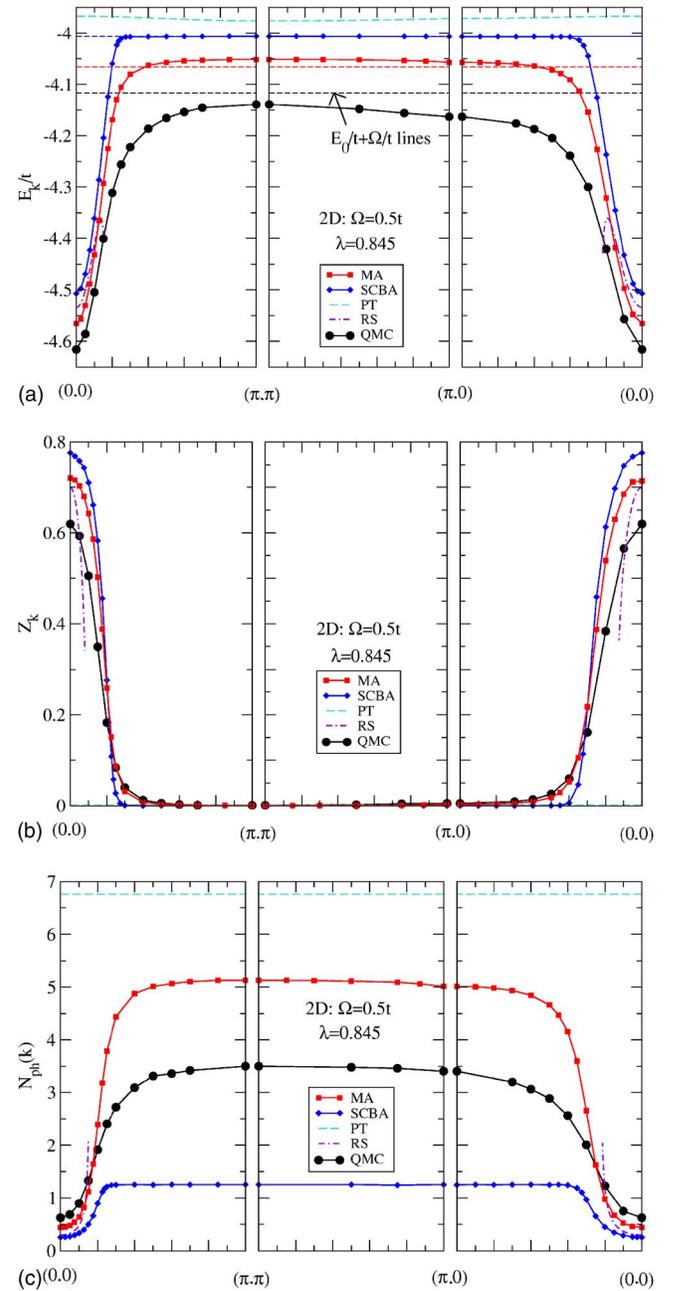
FIG. 11. (Color online) Lowest polaron energy for a given **k**, $E_{\mathbf{k}}$, and the corresponding qp weight $Z_{\mathbf{k}}$ and average phonon number $N_{ph}(\mathbf{k})$. Results are for 2D, $\Omega/t=0.5$, and $\lambda=0.845$. Three high-symmetry cuts in the Brillouin zone are shown.

In the following, we use MA to investigate more properties of the Green's function. We begin with Fig. 19, where we plot the qp weight and average phonon numbers for a few of the higher-energy peaks, once they appear below the continuum. For comparison, we also show the already discussed GS results (black line). Results in higher dimension are qualitatively similar and we do not show them here. Unlike for the GS, both these quantities are nonmonotonic functions of λ for all higher-energy bound states. Each of these states disperses with k, as in Figs. 15 and 16 (more data for this is shown below), so an effective mass can be associated with
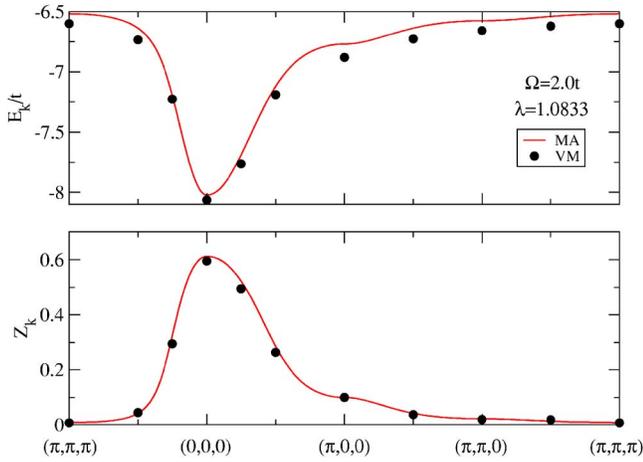
FIG. 12. (Color online) Lowest polaron energy for a given **k**, $E_{\mathbf{k}}$, and the corresponding qp weight $Z_{\mathbf{k}}$. Results are for 3D, Ω/t =2.0, and λ=1.0833. Three high-symmetry cuts in the Brillouin zone are shown. Variational data is from Ref. 21.

each such band. This effective mass satisfies $m^*/m=1/Z$, and therefore also shows nonmonotonic behavior, first decreasing and then increasing as λ is increased. The average phonon number in the nth state must approach $n+g^2/\Omega^2$ asymptotically, as can be verified in the Lang-Firsov limit. This is indeed observed in Fig. 19, however, the plateaus seen at moderate λ suggest some crossover from one to another type of wave function associated with these higher levels. As $\lambda \rightarrow \infty$, an infinite sequence of such bound states appear, as expected in the Lang-Firsov limit.

For a better illustration of the appearance of these bound states and of their evolution, we show contour plots of the
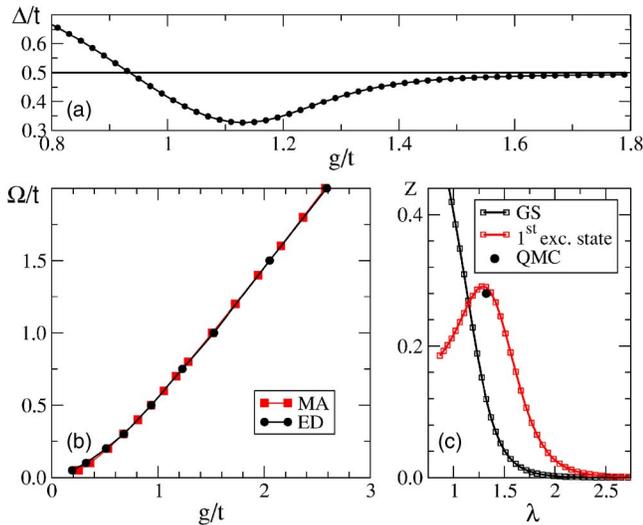
FIG. 13. (Color online) (a) Energy gap $\Delta=E_1-E_0$ between GS and first excited k=0 state. A second bound state is stable if $\Delta < \Omega$ (here Ω=0.5t); (b) line below which a second bound state appears: ED (Ref. 18) (circles) and MA (squares); (c) MA qp weight of the second bound state when stable (red squares), and that of the GS (black squares), for Ω/t=0.5. The circle is the one QMC result available for the qp weight of the second bound state (Ref. 32). These results are for 1D.
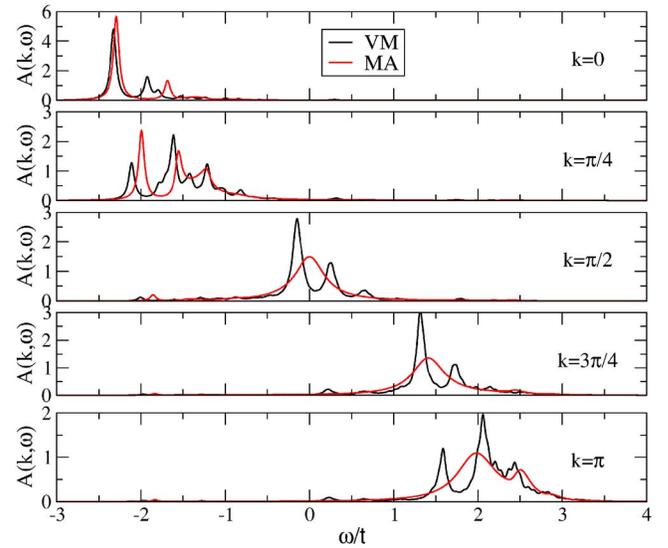
FIG. 14. (Color online) 1D spectral weight $A(k,\omega)$ vs ω, for k =0, $\pi/4$, $\pi/2$, $3\pi/4$, and π. MA results (red line) vs data from Ref. 23 (black line). Here Ω=0.4t, λ=0.5, and η=0.1 Ω.

spectral weight $A(\mathbf{k},\omega)$. We begin by plotting $A(\mathbf{k},\omega)$ as a function of **k** and ω, for fixed parameters g, t, and Ω. In Fig. 20 we show 1D results corresponding to Ω=t and λ=0.4, 1, and 2, respectively. Only half of the BZ is shown, since time-invariance guarantees that $G(\mathbf{k},\omega)=G(-\mathbf{k},\omega)$. Each of these MA contour plots takes below ten seconds to generate. Note that similar plots for the same parameters were provided by Hohenadler *et al.* in Ref. 39, based on a cluster perturbation theory approach. The agreement between the main features of our and their plots is excellent. As expected, their data does show a few more low-weight features at lower energies, below $-2t+\Omega=-t$, in this case, where our continuum starts. Such contour plots are richer versions of plots like those shown in Figs. 14–16. They illustrate basi-
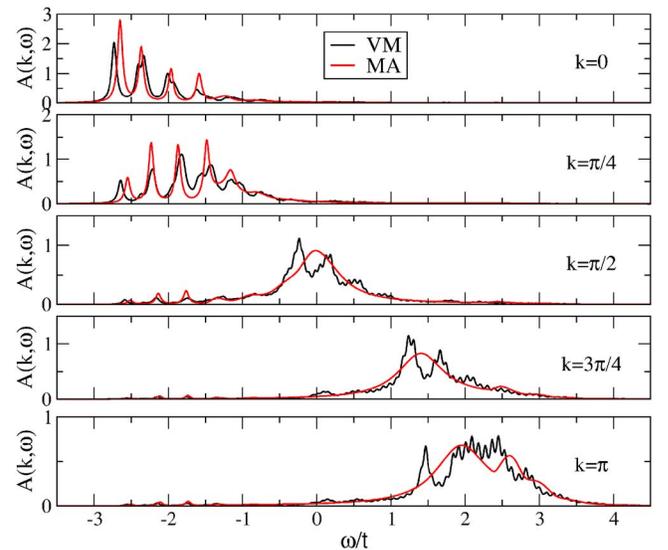
FIG. 15. (Color online) 1D spectral weight $A(k,\omega)$ vs ω, for k =0, $\pi/4$, $\pi/2$, $3\pi/4$, and π. MA results (red line) vs data from Ref. 23 (black line). Here Ω=0.4t, λ=1, and η=0.1 Ω.
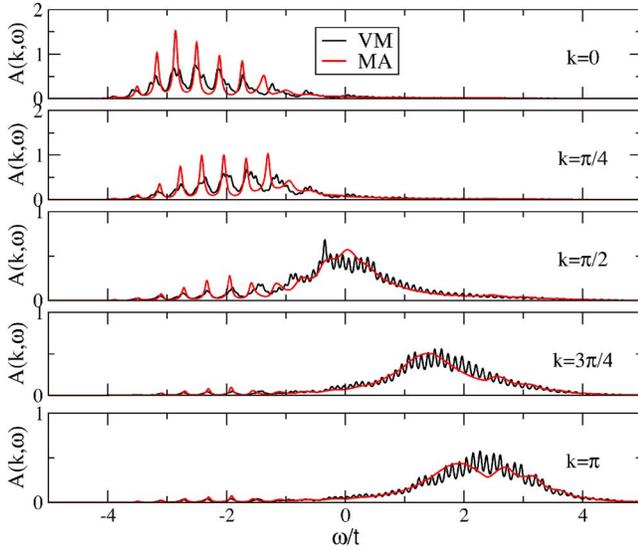
FIG. 16. (Color online) 1D spectral weight $A(k,\omega)$ vs ω, for $k=0$, $\pi/4$, $\pi/2$, $3\pi/4$, and π. MA results (red line) vs data from Ref. 23 (black line). Here $\Omega=0.4t$, $\lambda=2$, and $\eta=0.1\,\Omega$.

cally the same points, although one can now also see clearly the dispersion of various features. For the lowest λ, the free electron dispersion $\epsilon_k=-2t\cos(ka)$ is still almost visible, except that electron-phonon interactions split it into the lower polaron band and the higher continuum. This continuum is not featureless, instead one can already see weight accumulating near its lower edge. As λ increases, a new bound state will split off from it. This is seen for $\lambda=1$, where there are two bound states below the continuum starting at -1 (for these parameters). The bandwidth of each of these states is now narrowed below Ω. The weight in the continuum at higher energies is redistributed suggesting the impending formation of yet more bound states. Indeed, the $\lambda=2$ data shows four even narrower bound states below the continuum, which is showing yet more resonances at multiples of the phonon frequency.
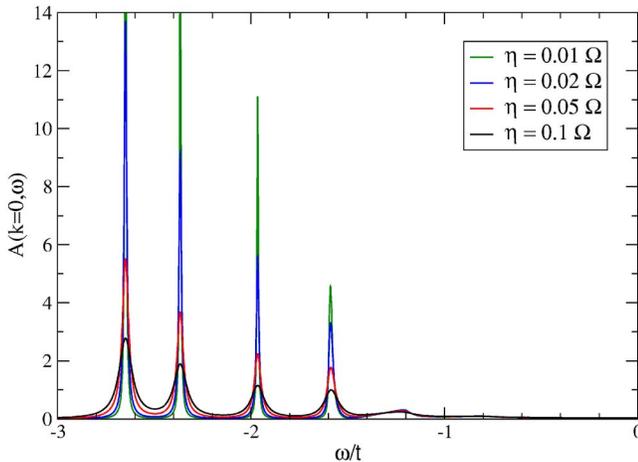
FIG. 17. (Color online) MA 1D spectral weight $A(k=0,\omega)$ vs ω, for $\Omega=0.4t$, $\lambda=1$, and $\eta/\Omega=0.1$, 0.05, 0.02, and 0.01. The first three peaks are discrete states (Lorentzians), whereas the fourth marks the band-edge singularity of the continuum.
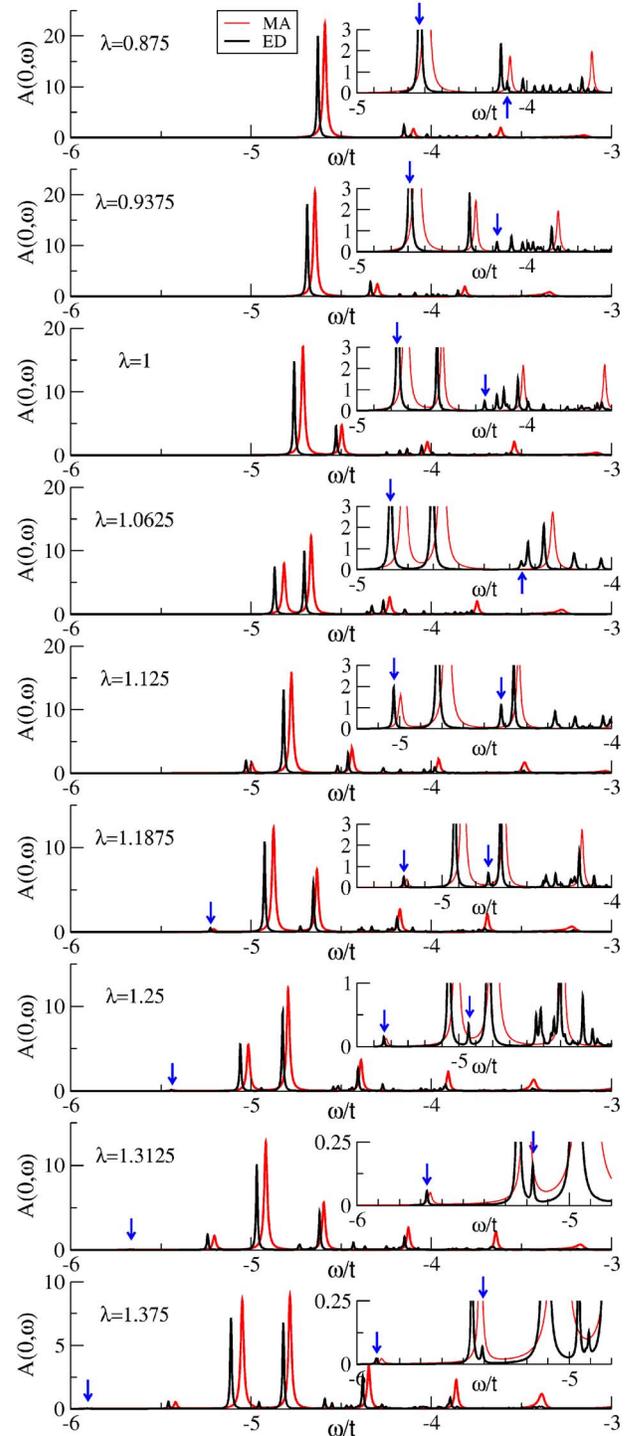
FIG. 18. (Color online) 2D spectral weight $A(k=0,\omega)$ for $\Omega=0.5t$, $\eta=0.01t$, and various λ values, from exact diagonalization (Ref. 41) (black) and MA (red). For $\lambda>1.125$, the arrows point to the GS location. The insets show the same data on a smaller scale, so that low-weight states are more visible. The arrows show the GS and the state appearing at precisely Ω above GS, in the ED results.

Similar behavior is expected, and indeed seen, in higher dimensions. Here we only show similar 2D contour plots, for $\Omega=2t$ and $\lambda=0.5$, 0.945, and 2, in Fig. 21. The middle panel again agrees very well with data shown in Ref. 39. In this
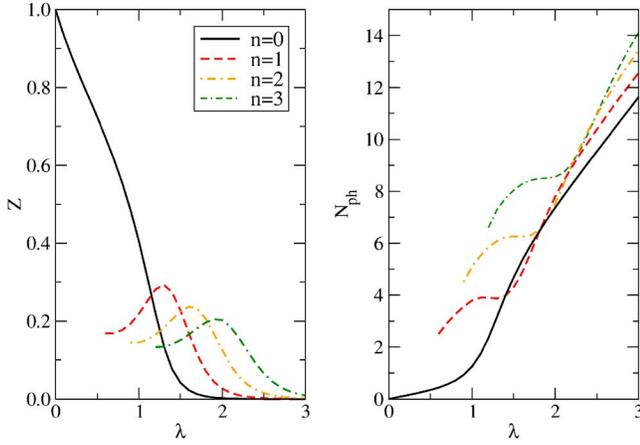
FIG. 19. (Color online) qp weight and average number of phonons in the GS (black line) and the next three higher $k=0$ bound states, when they become stable according to MA. Results are for 1D, $\Omega=0.5t$.

case, the MA continuum starts at $-4t+\Omega=-2$. As λ increases, we see again first 1, then 2, and then 4 bound states below the continuum. Their bandwidths decrease with increasing λ, so that the lowest band for $\lambda=2$ is already almost dispersionless, even though its weight still varies with **k**. As in the 1D case, as the interaction becomes stronger, the weight in the continuum also redistributes itself, with strong resonances seen around multiples of the phonon frequency.

Another way to understand the dependence on the coupling λ (or any other parameter) is to plot a contour of the spectral weight $A(\mathbf{k},\omega)$ vs λ and ω for a fixed value of **k**. Such a task is equally trivial at the MA level. In fact, one can also just as easily calculate and plot the total density of states or spectral weight,

$$A(\omega)=\frac{1}{N}\sum_{\mathbf{k}}A(\mathbf{k},\omega),$$

since within the MA approximation this is given by

$$A(\omega)=-\frac{1}{\pi}\,\mathrm{Im}\left[\frac{1}{N}\sum_{\mathbf{k}}\frac{1}{\omega-\varepsilon_{\mathbf{k}}-\Sigma_{\mathrm{MA}}(\omega)+i\eta}\right]$$

$$=-\frac{1}{\pi}\,\mathrm{Im}\{\bar{g}_0[\omega-\Sigma_{\mathrm{MA}}(\omega)]\}.$$

In Fig. 22, we show four such contour plots for the 1D case. Panel (a) shows $A(k=0,\omega)$ vs ω. For $\lambda=0$ (noninteracting case), only one state exists at $-2t$, as expected. As the coupling turns on, the energy of this state (the ground state) decreases, but $k=0$ weight is also transferred to higher energies, due to hybridization with the states in the electron-plus-one-phonon continuum. The MA continuum here starts at $-1.5t$. For moderate and larger λ one can clearly see how weight is rearranged inside the continuum as λ increases, and new bound states split from it and move toward lower energies. The apparent "break" in the slope of the GS energy, as λ increases, is now seen to occur when the first bound state approaches the GS, and is suggestive of an avoided crossing. From this point on, the GS lowers its energy much faster, but
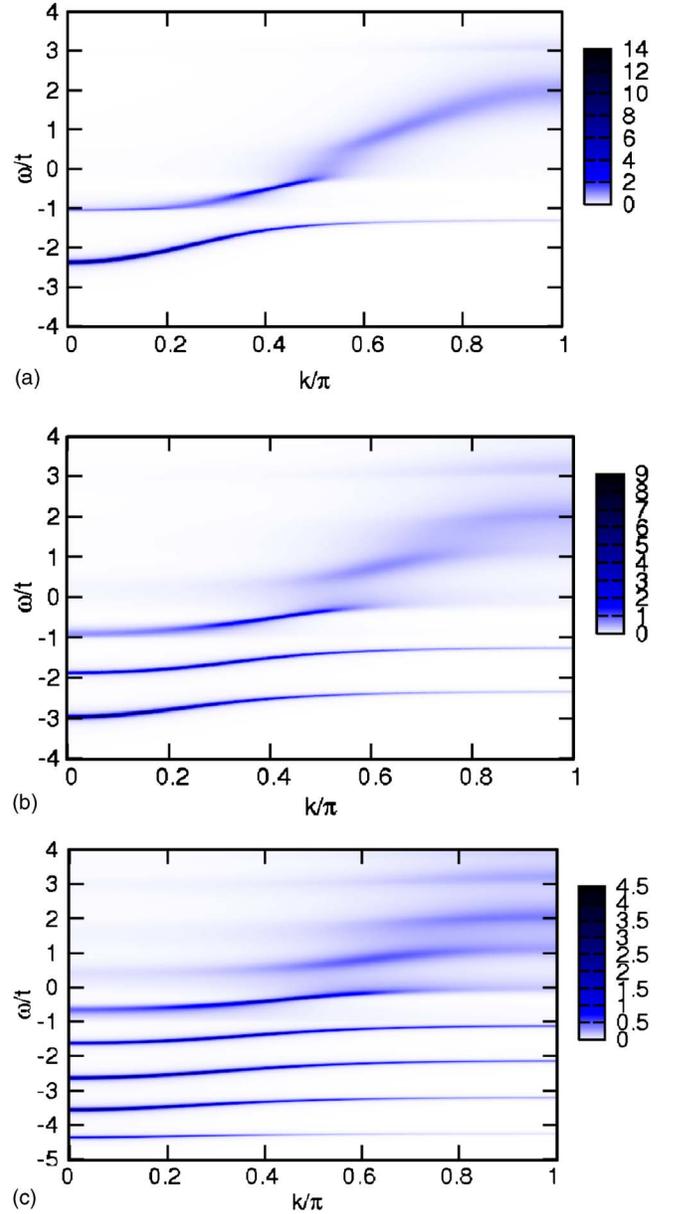
FIG. 20. (Color online) Contour plots of the 1D spectral weight $A(k,\omega)$ as a function of k and ω. The intensity scales are shown to the right of each plot. Parameters are $\Omega=t$, $\lambda=0.4$, 1, and 2, respectively, and $\eta=0.02$. These results are in excellent agreement with the results of Hohenadler *et al.* (Ref. 39) and the GS results of Bonca *et al.* (Ref. 18).

its weight also decreases dramatically and it becomes difficult to see. Panel (b) of Fig. 22 shows $A(k=\pi,\omega)$ vs ω, on a linear scale. At $\lambda=0$, there is only one peak at $+2t$, as expected. As the coupling is turned on, this weight seems to spread around in a rather featureless, broad continuum. In fact, on a logarithmic scale [panel (c) of Fig. 22] one can see that $k=\pi$ weight is pulled down into all the bound states, in agreement with the previous data we showed. This weight, however, is so small that it is not visible on the linear scale. Finally, panel (d) of Fig. 22 shows the total spectral weight or DOS. At $\lambda=0$, we see the usual 1D DOS, with the singularities near the band edges rounded off because of the finite
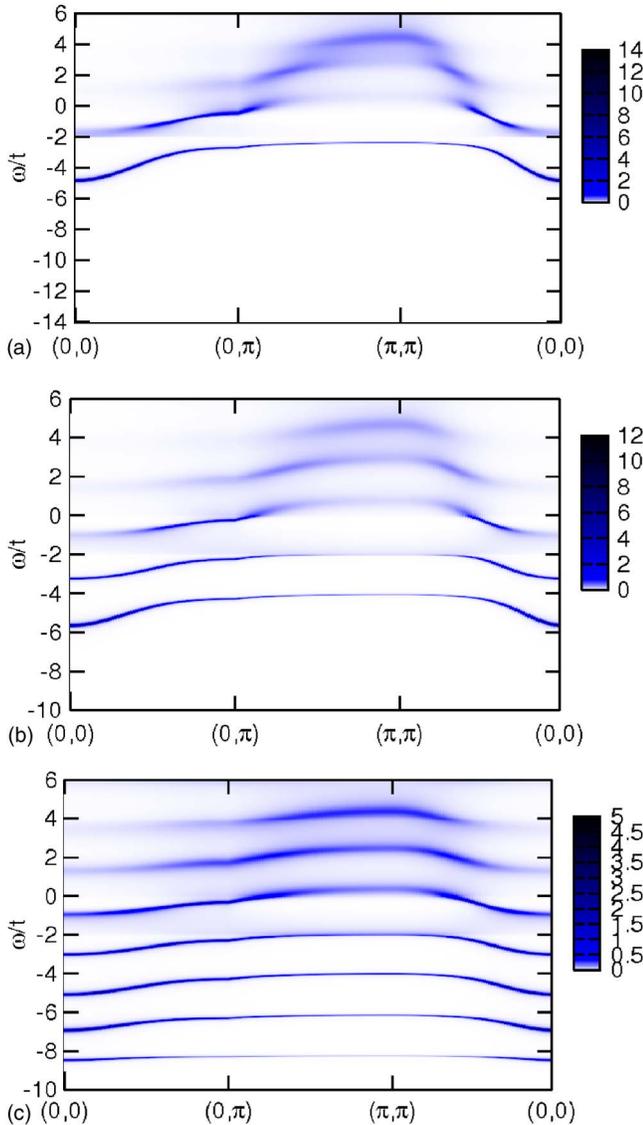
FIG. 21. (Color online) Contour plots of the 2D spectral weight $A(\mathbf{k},\omega)$ as a function of $\mathbf{k}$ and ω, along several cuts in the BZ. The intensity scales are shown to the right of each plot. Parameters are $\Omega=2t$, $\lambda=0.5$, 0.945, and 2, respectively, and $\eta=0.02$. The middle panel is in excellent agreement with the results of Hohenadler *et al.* given in Ref. 39.

η used. As λ is turned on, one can recognize both the contribution from the $k=0$ and $k=\pi$ states to the total DOS: Each bound state has a finite bandwidth due to its dispersion (this is to be contrasted to the upper panels, where the bound states are true delta functions, with a width defined by the broadening η). As the coupling strength increases, the number of bound states increases; they are spaced by roughly Ω, their bandwidths narrow down, and their weights also decrease. In other words, they approach the expected Lang-Firsov behavior.

Thus this figure actually answers the question posed in the Introduction, regarding the evolution of the spectral weight from that of a free electron toward that of the impurity limit. While the MA results are certainly not exact, we can claim with a high degree of certainty that the main features are accurately captured, especially in the weak and in the strong coupling limits. This suffices to understand the physics of this problem and, given the simplicity of the approximation, to investigate in detail a number of other quantities we have not shown here, such as the self-energy. Of course, if one is interested in exact results for some particular set of parameters, numerical methods have to be used.

Qualitatively similar plots are obtained in 2D and 3D, as shown in Figs. 23 and 24. Of course, the $\lambda=0$ DOS is very different, with a van Hove singularity at $\omega=0$ for the 2D case, and the characteristic nearest-neighbor hopping DOS in the 3D case. However, the appearance of multiple bands below the continuum as λ increases and all the remaining phenomenology is very similar. Thus we see no evidence of any qualitative differences in the polaron physics due to different dimensionality.

V. SUMMARY AND CONCLUSIONS

In this paper, we analyzed the Green's function of the Holstein polaron, using the *momentum average* approximation, which consists in summing all the self-energy diagrams, but with each individually averaged over all its free propagator momenta. The resulting self-energy can be written as an infinite continuous fraction that is numerically trivial to evaluate. This procedure becomes exact in the limits $g=0$ and $t=0$.

We gauged the accuracy of this approximation by computing its corresponding spectral weight sum rules and comparing them against the exact sum rules, which are known for this type of Hamiltonian. We showed that the MA spectral weight satisfies exactly the first six sum rules. Even though this is quite impressive at first sight, it is actually no guarantee of overall accuracy, as the case of SCBA demonstrates. The SCBA spectral weight always satisfies exactly the first four sum rules, even at large couplings λ where it predicts very wrong results. The meaningful criterion of accuracy for the sum rules is to show that the vast majority of terms in *all sum rules*, and in particular the dominant terms in the various asymptotic limits, are captured by the approximation. MA indeed satisfies this very restrictive criterion.

The accuracy of the approximation was also tested by direct comparison with data obtained through numerically intensive methods. In all cases, we obtain remarkable agreement, especially considering the ease of evaluation of the MA results. The MA approximation is not exact and some features are not correctly captured (for example, the continuum starting at $E_0+\Omega$); however, in all cases, all the higher-weight features in the spectral weight are quantitatively and qualitatively well described by the MA approximation. Trading some of the accuracy of numerically exact but often time consuming methods in exchange for very fast results which capture the main features accurately is a useful approach when trying to understand the main aspects of the physics of a problem, as well as when one is concerned about comparison with experiments. It is very unlikely that ARPES could capture very low-weight features; these would be lost in the noise background. Thus an approximation like MA, that quickly but accurately estimates results, is very useful,
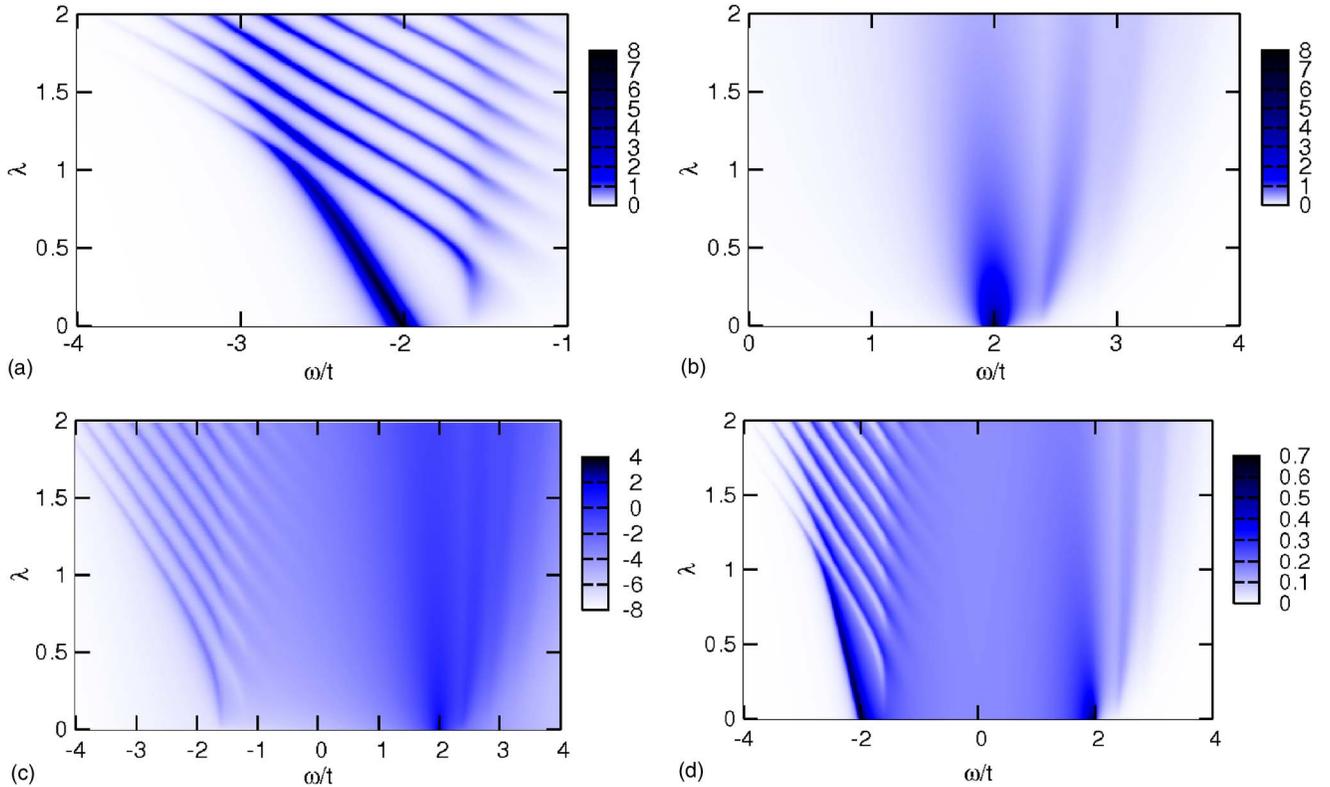
FIG. 22. (Color online) 1D results for $\Omega=0.4t$ and $\eta=0.04t$. (a) $A(k=0,\omega)$ vs ω and λ; (b) $A(k=\pi,\omega)$ vs ω and λ, on a linear scale; (c) $A(k=\pi,\omega)$ vs ω and λ, on a logarithmic scale; (d) total spectral weight $A(\omega)$ vs ω and λ. The intensity scales are shown to the right of each plot.

to be followed, of course, by detailed numerics for the parameter sets of interest.

It is important to note that the existence and accuracy of approximations like MA is not guaranteed; in fact, it can be regarded as a surprise for the case of the Holstein polaron, that has been under investigation for almost five decades. However, this demonstration of its existence and efficiency in the Holstein polaron problem gives some hope for making progress for the general class of strongly correlated systems problems. One can always use some flavor of perturbation theory to understand behavior in asymptotic limits, but the really challenging problems are set in regimes where perturbation does not apply. The MA approximation suggests that one way to make nontrivial progress is to sum all diagrams, with each simplified enough so as to make the calculation feasible, but not so much as to really alter the physics. This is a very different approach from the usually employed summation of a subclass of diagrams. Note that there are many classes of problems with diagrams similar to the ones arising in the single polaron problem, although of course with different propagators and/or vertices.

A first possible generalization of this work is to models with several phonon branches, and/or momentum-dependent coupling $g_\mathbf{q}$, and/or dispersive phonons $\Omega_\mathbf{q}$. The way to achieve this for a single polaron is briefly discussed in Ref. 42. Given the length of this article, we postpone this discussion for future publications where results can also be shown. Other directions of generalization are to finite particle densities and/or finite temperatures, and indeed to Hamiltonians

which also include electron-electron interactions.

Such work is currently in progress. It is still far from clear which cases will admit useful generalizations; however, the proof of existence of this method for the Holstein polaron problem is in itself encouraging.

APPENDIX: COMPARISON WITH DMFT

The MA self-energy of Eq. (19) looks similar to the DMFT self-energy, discussed in Ref. 35. This is not so surprising, since both become equal to the exact Lang-Firsov limit if the bandwidth goes to zero, and this can be rewritten as an infinite continued fraction, as shown in Eq. (12). However, this similarity may raise questions about the relationship between the two approximations. In this Appendix, we briefly show that the two approximations are very different at all finite t and g.

The meaning of the MA $\bar{g}_0(\omega)$ and of corresponding DMFT $G_0(\omega)$ (notation of Ref. 35) is very different. The DMFT $G_0(\omega)$ is obtained by solving exactly the problem of
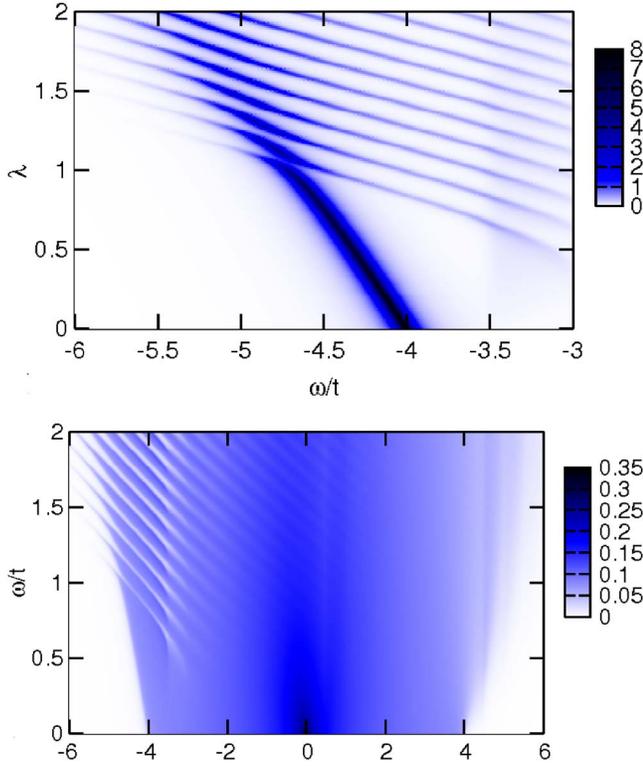
FIG. 23. (Color online) 2D results for $\Omega=0.5t$ and $\eta=0.04t$. Top: $A(\mathbf{k}=(0,0),\omega)$ vs ω and λ. Bottom: Total 2D spectral weight $A(\omega)$ vs ω and λ. The intensity scales are shown to the right of each plot.

FIG. 24. (Color online) 3D results for $\Omega=0.5t$. Top: $A(\mathbf{k}=(0,0,0),\omega)$ vs ω and λ, for $\eta=0.15t$. Bottom: Total 3D spectral weight $A(\omega)$ vs ω and λ, for $\eta=0.04t$. The intensity scales are shown to the right of each plot.

an impurity coupled to an environment in the $d\rightarrow\infty$ limit, and then imposing the self-consistency condition that the impurity site behaves similar to all other sites in the environment. The DMFT $G_0(\omega)$ is calculated self-consistently using the following steps:[35] (i) With some initial guess for $G_0(\omega)$, one calculates the DMFT self-energy, given by a formula similar to Eq. (19), with $\bar{g}_0(\omega)$ replaced by $G_0(\omega)$; this self-energy is then used to calculate the total Green's function

$$G(\omega) = \int_{-\infty}^{\infty} d\epsilon \, \rho_0(\epsilon) \frac{1}{\omega - \epsilon - \Sigma(\omega) + i\eta},$$

where $\rho_0(\epsilon)$ is the density of states of the noninteracting electrons. The usual procedure is to take $d\rightarrow\infty$ and use as DOS the semielliptical formula corresponding to an infinitely branched Bethe tree,

$$\rho_0(\omega) = \frac{2}{\pi\left(\frac{W}{2}\right)^2} \sqrt{\left(\frac{W}{2}\right)^2 - \epsilon^2},$$

where W is the bandwidth for the noninteracting system. Then (iii) the new $G_0(\omega)$ is extracted from the condition that $G(\omega)=[G_0^{-1}(\omega)-\Sigma(\omega)]^{-1}$ and the procedure is repeated until self-consistency is reached. Reaching self-consistency is a nontrivial numerical task, especially compared to obtaining the MA $\bar{g}_0(\omega)$ (see below). More importantly, the DMFT $G_0(\omega)$ is an explicit function of g and Ω.

In contrast, the MA $\bar{g}_0(\omega)$ is the momentum average of the free propagator: Thus it is a known function, independent of the parameters g and Ω. In particular, for the semielliptical DOS, we have simply

$$\bar{g}_0(\omega) = \int_{-\infty}^{\infty} d\epsilon \frac{\rho_0(\epsilon)}{\omega - \epsilon + i\eta}$$

$$= \frac{8}{W^2}(\omega+i\eta)\left[1 - \sqrt{1 - \frac{W^2}{4(\omega+i\eta)^2}}\right].$$
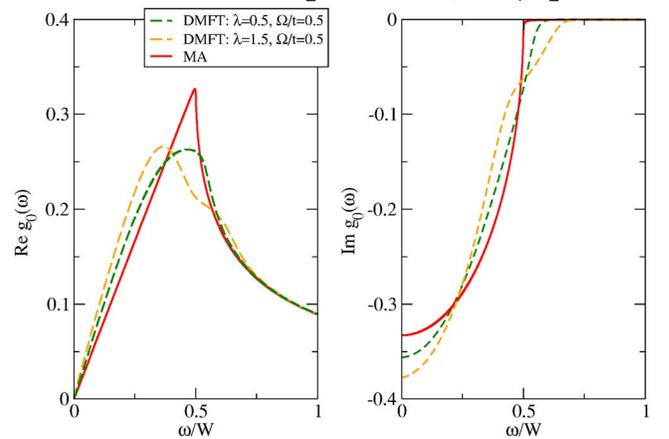
FIG. 25. (Color online) Comparison between the function $G_0(\omega)$ entering the DMFT self-energy, for $\omega/t=0.5$ and $\lambda=0.5$ (green) and $\lambda=1.5$ (yellow), and $\bar{g}_0(\omega)$ entering the MA self-energy, for the $d\rightarrow\infty$ semielliptical DOS.

A comparison of these functions is provided in Fig. 25. They are clearly different. Moreover, note that in the MA approximation, $G(\mathbf{k}, \omega)$ is an explicit function of the momentum. The MA self-energy for the Holstein polaron happens to be independent of the momentum, but this

a consequence of the simplicity of the model, not an "in-built" feature as in DMFT. Generalizations to models with a momentum-dependent coupling and/or dispersive phonons lead to momentum-dependent self-energies.[42]

[1] A. Lanzara, P. V. Bogdanov, X. J. Zhou, S. A. Kellar, D. L. Feng, E. D. Lu, T. Yoshida, H. Eisaki, A. Fujimori, K. Kishio, J.-I. Shimoyama, T. Noda, S. Uchida, Z. Hussain and Z.-X. Shen, Nature (London) **412**, 510 (2001).

[2] K. M. Shen, F. Ronning, D. H. Lu, W. S. Lee, N. J. C. Ingle, W. Meevasana, F. Baumberger, A. Damascelli, N. P. Armitage, L. L. Miller, Y. Kohsaka, M. Azuma, M. Takano, H. Takagi, and Z.-X. Shen, Phys. Rev. Lett. **93**, 267002 (2004).

[3] A. S. Mishchenko and N. Nagaosa, Phys. Rev. Lett. **93**, 036402 (2004).

[4] X. J. Zhou, T. Cuk, T. Devereaux, N. Nagaosa, and Z.-X. Shen, cond-mat/0604284 (unpublished).

[5] M. Hengsberger, D. Purdie, P. Segovia, M. Garnier, and Y. Baer, Phys. Rev. Lett. **83**, 592 (1999).

[6] O. Gunnarsson, Rev. Mod. Phys. **69**, 575 (1997), and references therein.

[7] W. P. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).

[8] A. S. Mishchenko, N. V. Prokof'ev, and B. V. Svistunov, Phys. Rev. B **64**, 033101 (2001).

[9] S. Isihara, Y. Murakami, T. Inami, K. Ishii, J. Mizuki, K. Hirota, S. Maekawa, and Y. Endoh, New J. Phys. **7**, 119 (2005), and references therein.

[10] T. Hotta, Rep. Prog. Phys. **69**, 2061 (2006).

[11] K. P. Schmidt, M. Grüninger, and G. S. Uhrig, cond-mat/0603766 (unpublished).

[12] T. Holstein, Ann. Phys. (N.Y.) **8**, 325 (1959); **8**, 343 (1959).

[13] G. D. Mahan, *Many Particle Physics* (Plenum, New York, 1981).

[14] A. Damascelli, Z. Hussain, and Z.-X. Shen, Rev. Mod. Phys. **75**, 473 (2003).

[15] A. S. Alexandrov, V. V. Kabanov, and D. K. Ray, Phys. Rev. B **49**, 9915 (1994); G. Wellein, H. Röder, and H. Fehske, Phys. Rev. B **53**, 9666 (1996); G. Wellein and H. Fehske, Phys. Rev. B **56**, 4513 (1997); E. V. L. de Mello and J. Ranninger, Phys. Rev. B **55**, 14872 (1997); M. Capone, W. Stephan, and M. Grilli, Phys. Rev. B **56**, 4484 (1997); G. Wellein and H. Fehske, Phys. Rev. B **58**, 6208 (1998); O. S. Barisic, Phys. Rev. B **65**, 144301 (2002).

[16] F. Marsiglio, Physica C **244**, 21 (1995).

[17] O. S. Barisic, Phys. Rev. B **73**, 214304 (2006).

[18] J. Bonca, S. A. Trugman, and I. Batistic, Phys. Rev. B **60**, 1633 (1999).

[19] A. H. Romero, D. W. Brown, and K. Lindenberg, Phys. Rev. B **59**, 13728 (1999); **60**, 4618 (1999); **60**, 14080 (1999).

[20] V. Cataudella, G. De Filippis, and G. Iadonisi, Phys. Rev. B **60**, 15163 (1999); **62**, 1496 (2000).

[21] L.-C. Ku, S. A. Trugman, and J. Bonca, Phys. Rev. B **65**, 174306 (2002).

[22] V. Cataudella, G. De Filippis, F. Martone, and C. A. Perroni,

Phys. Rev. B **70**, 193105 (2004).

[23] G. De Filippis, V. Cataudella, V. Marigliano Ramaglia, and C. A. Perroni, Phys. Rev. B **72**, 014307 (2005).

[24] O. S. Barisic, Phys. Rev. B **69**, 064302 (2004).

[25] F. Marsiglio, Phys. Rev. B **42**, 2416 (1990); H. De Raedt and A. Lagendijk, Phys. Rev. Lett. **49**, 1522 (1982); Phys. Rev. B **27**, 6097 (1983); **30**, 1671 (1984); P. Kornilovitch, J. Phys.: Condens. Matter **9**, 10675 (1997); P. E. Kornilovitch and E. R. Pike, Phys. Rev. B **55**, R8634 (1997); P. E. Kornilovitch, Phys. Rev. B **60**, 3237 (1999).

[26] P. E. Kornilovitch, Phys. Rev. Lett. **81**, 5382 (1998).

[27] M. Hohenadler, H. G. Evertz, and W. von der Linden, Phys. Status Solidi B **242**, 1406 (2005).

[28] M. Hohenadler, H. G. Evertz, and W. von der Linden, Phys. Rev. B **69**, 024301 (2004) and references therein.

[29] H. Fehske and S. A. Trugman, in *Polarons in Advanced Materials*, edited by A. S. Alexandrov (Canopus Publishing and Springer-Verlag GmbH, Bath, UK, 2007).

[30] N. V. Prokof'ev and B. V. Svistunov, Phys. Rev. Lett. **81**, 2514 (1998).

[31] A. S. Mishchenko, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000).

[32] A. Macridin, Ph. D. thesis, Rijkuniversiteit Groningen, 2003.

[33] E. Jeckelmann and S. R. White, Phys. Rev. B **57**, 6376 (1998).

[34] H. Sumi, J. Phys. Soc. Jpn. **36**, 770 (1974).

[35] S. Ciuchi, F. de Pasquale, S. Fratini, and D. Feinberg, Phys. Rev. B **56**, 4494 (1997).

[36] A. S. Alexandrov and N. F. Mott, *Polarons and Bipolarons* (World Scientific, Singapore, 1995).

[37] I. G. Lang and Y. A. Firsov, Sov. Phys. JETP **16**, 1301 (1963).

[38] H. Löwen, Phys. Rev. B **37**, 8661 (1988).

[39] M. Hohenadler, M. Aichhorn, and W. von der Linden, Phys. Rev. B **68**, 184304 (2003).

[40] M. Hohenadler, D. Neuber, W. von der Linden, G. Wellein, J. Loos, and H. Fehske, Phys. Rev. B **71**, 245111 (2005).

[41] Bayo Lau and George Sawatzky (unpublished).

[42] M. Berciu, Phys. Rev. Lett. **97**, 036402 (2006).

[43] C. Slezak, A. Macridin, G. A. Sawatzky, M. Jarrell, and T. A. Maier, Phys. Rev. B **73**, 205122 (2006).

[44] A. S. Alexandrov and J. Ranninger, Phys. Rev. B **45**, 13109 (1992).

[45] P. E. Kornilovitch, Europhys. Lett. **59**, 735 (2002).

[46] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. (Dover, New York, 1964).

[47] R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[48] J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, Redwood City, CA, 1988).