Systematic improvement of the momentum average approximation for the Green's function of a Holstein polaron

Mona Berciu and Glen L. Goodvin

Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1
(Received 27 May 2007; revised manuscript received 20 August 2007; published 9 October 2007)

We show how to systematically improve the momentum average (MA) approximation for the Green's function of a Holstein polaron by systematically improving the accuracy of the self-energy diagrams in such a way that they can still all be summed efficiently. This allows us to fix some of the problems of the MA approximation, e.g., we now find the expected polaron+phonon continuum at the correct location, and a momentum-dependent self-energy. The quantitative agreement with numerical data is further improved as expected since the number of exactly satisfied spectral weight sum rules is increased. The corrections are found to be larger in lower dimensional systems.

I. INTRODUCTION

One of the main challenges in condensed matter physics is to find simple yet accurate analytical approximations for systems in nonperturbative (strongly coupled) regimes. This is needed for interpretation of experiments and to help supplement the results available from numerical simulations of model Hamiltonians.

In this general context, we have recently proposed the so-called momentum average (MA) approximation for the Green's function of a single Holstein polaron.[1,2] The essence of this approximation is to sum all diagrams contributing to the polaron self-energy; however, each diagram is approximated to such a degree as to allow the analytical summation of the entire series. Specifically, each free propagator appearing in a self-energy diagram is replaced by its momentum average. The resulting MA self-energy is a trivial-to-evaluate continued fraction which gives remarkably accurate results over most of the parameter space, including intermediate electron-phonon coupling strengths, where perturbational methods completely fail to capture the correct physics.

In Refs. 1 and 2, we identified some of the reasons for this good agreement: first, the MA approximation becomes asymptotically exact for both very weak and very strong couplings. More importantly, the resulting MA spectral weight satisfies the first six spectral weight sum rules exactly, and remains highly accurate for all higher order sum rules. However, we also pointed out some shortcomings of the MA approximation: (i) it fails to predict the continuum that must appear at $E_{GS}+\Omega$, where E_{GS} is the polaron ground-state (GS) energy and Ω is the frequency of the Einstein phonons (we set $\hbar=1$ throughout this work). This continuum arises from states that have a phonon excited very far from where the polaron is. As a result, their interactions are negligible and the energy of the system is simply the sum of the two. MA either predicts a wrong location for this continuum (at weak electron-phonon couplings) or no continuum at all in that range of energies (at intermediary and strong electron-phonon couplings). We noted in Ref. 2 that numerical simulations show that there is very little spectral weight in this continuum, hence its absence or wrong positioning does not significantly upset the agreement with the sum rules. Never-

theless, it would be reassuring to have an approximation that correctly predicts its existence; (ii) the accuracy of the MA approximation worsens as $\Omega \rightarrow 0$; (iii) the MA self-energy $\Sigma_{MA}(\omega)$ is independent of the momentum **k** of the electron. Given how featureless the Holstein model is (electron-phonon coupling and phonon frequency are both constants), one may expect a rather weak momentum dependence of the self-energy; however, it is certainly not entirely absent.

In this paper, we show how to systematically improve the MA approximation, generating a hierarchy of approximations that we call $MA^{(n)}$ (the original MA is $MA^{(0)}$ in this notation). As explained below, the idea is to systematically improve the accuracy of the "simplified" self-energy diagrams. The results become more and more accurate as n increases, for example, while the MA spectral weight satisfies only the first six sum rules exactly, this improves to eight and ten exact sum rules, respectively, for $MA^{(1)}$ and $MA^{(2)}$ spectral weights. While the numerical effort also increases, it is still trivial for the $n=1$ and $n=2$ levels that we discuss explicitly here. Level $MA^{(1)}$ already solves the continuum problem, while all levels with $n \geq 2$ produce momentum-dependent self-energies. The accuracy in the limit $\Omega \rightarrow 0$ is also shown to improve significantly with increasing n. In effect, for a slightly increased numerical effort, $MA^{(2)}$ solves all the known problems of the $MA^{(0)}$ approximation.

The work is organized as follows: in Sec. II, we briefly review the $MA^{(0)}$ approximation, presenting a different argument to explain its accuracy. In Sec. III, we describe the systematic approach to obtain the improved versions $MA^{(n)}$, $n \geq 1$, and give explicit formulas for the self-energies corresponding to the $MA^{(1)}$ and $MA^{(2)}$ approximations. In Sec. IV, we compare the predictions of these approximations against numerical simulations to gauge the improved accuracy as n increases. Spectral sum rules, as well as variational arguments, will also be used to explain the systematic improvement of accuracy with increasing n. Finally, Sec. V contains our conclusions.

II. BRIEF REVIEW OF THE MA(0) APPROXIMATION

The Holstein model is the simplest lattice model that includes electron-phonon coupling. Its Hamiltonian is[3]

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

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 electron (polaron) problem of interest to us, the spin of the electron is irrelevant and we suppress its index. $\varepsilon_{\mathbf{k}}$ is the free-particle dispersion. In all the results shown here, we assume nearest-neighbor hopping on a d-dimensional simple cubic lattice of constant a (we set $a = 1$) with a total of N sites and with periodic boundary conditions. In this case,

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

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 $V_{\text{el-ph}} = g \sum_i c_i^{\dagger} c_i (b_i^{\dagger} + b_i)$, written in $\mathbf{k}$ space. All sums over momenta are over the first Brillouin zone, namely, $-\pi < k_i \leq \pi$, $i = 1, \dots, d$.

The quantity of interest to us is the (retarded) single polaron Green's function, defined as[1,2]

$$G(\mathbf{k}, \omega) = \langle 0 | c_{\mathbf{k}} \hat{G}(\omega) c_{\mathbf{k}}^{\dagger} | 0 \rangle = \left\langle 0 \left| c_{\mathbf{k}} \frac{1}{\omega - \mathcal{H} + i\eta} c_{\mathbf{k}}^{\dagger} \right| 0 \right\rangle,$$

(1)

where $|0\rangle$ is the vacuum $c_{\mathbf{k}}|0\rangle = b_{\mathbf{q}}|0\rangle = 0$, and $\eta > 0$ is infinitesimally small.

As described in detail in Ref. 2, using repeatedly Dyson's identity $\hat{G}(\omega) = \hat{G}_0(\omega) + \hat{G}(\omega) V_{\text{el-ph}} \hat{G}_0(\omega)$, where $V_{\text{el-ph}} = \mathcal{H} - \mathcal{H}_0$ is the electron-phonon interaction potential and $\hat{G}_0(\omega) = [\omega - \mathcal{H}_0 + i\eta]^{-1}$, we generate the infinite hierarchy of equations of motion whose exact solution is the desired Green's function:

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

(2)

and for $n \geq 1$,

$$F_n(\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n, \omega) = \frac{g}{\sqrt{N}} G_0(\mathbf{k} - \mathbf{q}_T, \omega - n\Omega)$$

$$\times \left[\sum_{i=1}^{n} F_{n-1}(\mathbf{k}, \dots, \mathbf{q}_{i-1}, \mathbf{q}_{i+1}, \dots, \omega) \right.$$

$$\left. + \sum_{\mathbf{q}_{n+1}} F_{n+1}(\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_{n+1}, \omega) \right].$$

(3)

Here, $\mathbf{q}_T = \sum_{i=1}^{n} \mathbf{q}_i$ is the total momentum carried by phonons, $G_0(\mathbf{k}, \omega) = (\omega - \epsilon_{\mathbf{k}} + i\eta)^{-1}$ is the free electron Green's function, and we introduced the generalized Green's functions

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

These equations can be recast in a more convenient form after observing that if we treat Eq. (3) as an inhomogeneous system of linear equations in unknowns $F_1, F_2, \dots$, then the only inhomogeneous term appears in the first equation and is proportional to $F_0(\mathbf{k}, \omega) = G(\mathbf{k}, \omega)$. It follows that all generalized Green's functions $F_1, F_2, \dots$, must be proportional to $G(\mathbf{k}, \omega)$. As a result, we introduce the more convenient variables:

$$f_n(\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n, \omega) = \frac{N^{n/2} g^n F_n(\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n, \omega)}{G(\mathbf{k}, \omega)}.$$

(4)

In terms of these, Eq. (2) becomes

$$G(\mathbf{k}, \omega) = G_0(\mathbf{k}, \omega) \left[1 + \frac{1}{N} \sum_{\mathbf{q}_1} f_1(\mathbf{k}, \mathbf{q}_1, \omega) G(\mathbf{k}, \omega) \right]$$

so that the exact self-energy is

$$\Sigma(\mathbf{k}, \omega) = \frac{1}{N} \sum_{\mathbf{q}_1} f_1(\mathbf{k}, \mathbf{q}_1, \omega),$$

(5)

giving the standard solution

$$G(\mathbf{k}, \omega) = \frac{1}{\omega - \epsilon_{\mathbf{k}} - \Sigma(\mathbf{k}, \omega) + i\eta}.$$

(6)

To find $f_1(\mathbf{k}, \mathbf{q}_1, \omega)$, we must solve the infinite system of coupled equations that result from Eq. (3). For later convenience, we write the first few equations explicitly here, using the shorthand notation $f_n(\mathbf{k}, \mathbf{q}_1, \dots, \mathbf{q}_n, \omega) \equiv f_n(\mathbf{q}_1, \dots, \mathbf{q}_n)$ (i.e., the dependence of $\mathbf{k}$ and ω of these functions is implicitly assumed from now on). Then

$$f_1(\mathbf{q}_1) = G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega) \left[g^2 + \frac{1}{N} \sum_{\mathbf{q}_2} f_2(\mathbf{q}_1, \mathbf{q}_2) \right], \quad (7)$$

$$f_2(\mathbf{q}_1, \mathbf{q}_2) = G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \omega - 2\Omega) \left[g^2 [f_1(\mathbf{q}_1) + f_1(\mathbf{q}_2)] \right.$$

$$\left. + \frac{1}{N} \sum_{\mathbf{q}_3} f_3(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \right],$$

(8)

and for all $n \geq 3$,

$$f_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = G_0 \left(\mathbf{k} - \sum_{i=1}^{n} \mathbf{q}_i, \omega - n\Omega \right)$$

$$\times \left[g^2 \sum_{i=1}^{n} f_{n-1}(\mathbf{q}_1, \dots, \mathbf{q}_{i-1}, \mathbf{q}_{i+1}, \dots, \mathbf{q}_n) \right.$$

$$\left. + \frac{1}{N} \sum_{\mathbf{q}_{n+1}} f_{n+1}(\mathbf{q}_1, \dots, \mathbf{q}_{n+1}) \right].$$

(9)

Clearly, all the dependence on free propagators of the self-energy comes from the free propagator prefactors on the right-hand side of these equations.

As already stated, MA$^{(0)}$ consists in replacing all free propagators in all self-energy diagrams by their momentum average:
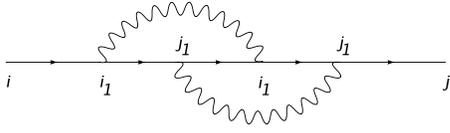
FIG. 1. A second-order diagram contribution to $G(i,j,\omega)$.

$$\bar{g}_0(\omega) = \frac{1}{N}\sum_{\mathbf{k}} G_0(\mathbf{k},\omega). \tag{10}$$

Obviously, this corresponds to replacing the free propagator prefactors on the right-hand side of Eqs. (7)–(9) by the corresponding $\bar{g}_0(\omega-n\Omega)$. In this case, it is straightforward to see that the resulting solutions, denoted $f_n^{(0)}$, are functions of ω only (all dependence on phonon momenta disappears at this level of approximation). The resulting equations $f_1^{(0)}(\omega)$ $=\bar{g}_0(\omega-\Omega)[g^2+f_2^{(0)}(\omega)]$ and for $n\geq 2$, $f_n^{(0)}(\omega)=\bar{g}_0(\omega-n\Omega)$ $\times[ng^2 f_{n-1}^{(0)}(\omega)+f_{n+1}^{(0)}(\omega)]$ are solved in terms of continued fractions[4] (also see Appendix A) to find[2]

$$\Sigma_{MA^{(0)}}(\omega)=f_1^{(0)}(\omega)=g^2 A_1(\omega), \tag{11}$$

where we define the infinite continued fractions:

$$A_n(\omega) = \frac{n\bar{g}_0(\omega-n\Omega)}{1-g^2\bar{g}_0(\omega-n\Omega)A_{n+1}(\omega)}$$

$$= \cfrac{n\bar{g}_0(\omega-n\Omega)}{1-\cfrac{(n+1)g^2\bar{g}_0(\omega-n\Omega)\bar{g}_0(\omega-(n+1)\Omega)}{1-\cdots}}. \tag{12}$$

Results based on this $MA^{(0)}$ approximation have been analyzed in detail in Ref. 2.

Before discussing how to systematically improve this approximation, it is worth pointing out an alternative explanation for its good accuracy.[5] This comes from realizing that in real space, the meaning of the $MA^{(0)}$ approximation is that it replaces all free propagators $G_0(i,j,\omega-n\Omega)$ appearing in all self-energy diagrams by $\delta_{i,j}\bar{g}_0(\omega-n\Omega)$. For example, consider the real-space, second-order Green's function diagram depicted in Fig. 1. It has the exact value $\Sigma_{i_1,j_1}G_0(i,i_1,\omega)\Sigma_{2,c}(i_1,j_1,\omega)G_0(j_1,j,\omega)$, where the contribution to self-energy from the second-order crossed self-energy diagram is $\Sigma_{2,c}(i_1,j_1,\omega)=g^4 G_0(i_1,j_1,\omega-\Omega)G_0(j_1,i_1,\omega-2\Omega)G_0(i_1,j_1,\omega-\Omega)$. Within $MA^{(0)}$, $\Sigma_{2,c}(i_1,j_1,\omega)$ is approximated as $\Sigma_{2,c}^{(0)}(i_1,j_1,\omega)=\delta_{i_1,j_1}\Sigma_{2,c}^{(0)}(\omega)=\delta_{i_1,j_1}g^4\bar{g}_0(\omega-\Omega)\bar{g}_0(\omega-2\Omega)\bar{g}_0(\omega-\Omega)$. Inserting this into the Green's function diagram removes one of the sums, and after Fourier transforming, we find that the contribution of this diagram to $G(\mathbf{k},\omega)$ is $G_0(\mathbf{k},\omega)\Sigma_{2,c}^{(0)}(\omega)G_0(\mathbf{k},\omega)$. The same holds for all higher order diagrams. Summing all of them, we see that the free propagators in the *proper self-energy parts* have, indeed, been replaced by their momentum averages.

The reason why it is a good zero-order approximation to keep only the diagonal (in real space) contributions $G_0(i,j,\omega-n\Omega)\rightarrow\delta_{i,j}\bar{g}_0(\omega-n\Omega)$ is straightforward to understand, at least for low energies $\omega\sim E_{GS}$. Because of interac-

tions, $E_{GS}<-2dt$ (the polaron ground state is below the free electron continuum). It follows that for $\omega\sim E_{GS}$, the free propagators $G_0(i,j,\omega-n\Omega)$ are needed at energies below the free electron continuum, and the larger n is, the further below the band edge these energies are. However, it is well known that for $\omega<-2dt$, the free propagator decreases exponentially with increasing distance $|i-j|$. The reason is that there are no free-electron eigenstates outside the free-electron band, and the electron has to tunnel from one site to another. For example, in one dimension,[6]

$$G_0(i,j,\omega) = e^{ik_0|i-j|}G_0(i,i,\omega) = e^{ik_0|i-j|}\bar{g}_0(\omega),$$

where k_0 is the first quadrant solution of the equation $\omega=-2t\cos k_0$. For $\omega<-2t$, $k_0=i\kappa$, where $\kappa>0$ increases as ω decreases. We also used the fact that in terms of its Fourier transform,

$$G_0(i,i,\omega) = \frac{1}{N}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)}G_0(\mathbf{k},\omega) = \bar{g}_0(\omega).$$

It follows that at low energies, $MA^{(0)}$ keeps the largest contributions to the self-energy (from real-space diagonal terms), while ignoring exponentially small contributions coming from off-diagonal terms. This is expected to become more and more accurate for higher order diagrams with many phonons: the larger n is, the faster the exponential decay with distance of $G_0(i,j,\omega-n\Omega)$ becomes.

Based on these arguments, one expects at least low-energy properties to be well described by MA, at least if Ω is not too small. Together with the good agreement[1,2] with the spectral weight sum rules, this leads to the conclusion that the self-energy and the Green's function at all energies should, indeed, be quite accurate within MA. Comparison with numerics validates this.[2]

III. HIGHER LEVEL $MA^{(n)}$: $n\geq 1$

The above arguments also suggest a systematic way toward improving the $MA^{(0)}$ approximation. The biggest error at low energies is due to the momentum average of the propagators of energy $\omega-\Omega$. This is the energy closest to the free-electron continuum and, therefore, these propagators have the slowest decay in real space. The next slowest decay is for the propagators of energy $\omega-2\Omega$, etc. If one could selectively keep these propagators exactly while momentum averaging the ones with more phonons (lower energy, faster decay), this should improve the accuracy of the approximation at low energies. In fact, all sum rules would also be improved (individual diagrams are more accurate), therefore one would expect an improvement at all energies. This is precisely what the higher level $MA^{(n)}$, $n\geq 1$, achieve.

We define $MA^{(n)}$ as being the approximation where in all self-energy diagrams, all free propagators with energy $\omega-m\Omega$, where $m\leq n$, are kept exactly, while the ones with more than n phonons are momentum averaged.

In terms of the equations of motion that need to be solved [Eqs. (7)–(9)], achieving this is straightforward, since a propagator of energy $\omega-m\Omega$ appears only once, in the right-hand side prefactor of the equation for $f_m(\mathbf{q}_1,\ldots,\mathbf{q}_m)$. It fol-

lows that if we keep the first n of Eqs. (7)–(9) as they are, and approximate the equations for $f_{n+1}, f_{n+2}, \ldots$, by momentum averaging the free propagator appearing in the right-hand side prefactor $G_0(\mathbf{k} - \mathbf{q}_T, \omega - m\Omega) \to \bar{g}_0(\omega - m\Omega)$ if $m \geq n+1$, we achieve our goal—provided that we can find the solution of the resulting infinite system of coupled equations.

We now derive explicitly the solutions for $n=1$ and $n=2$ levels.

A. MA$^{(1)}$ level

In this case, the equations to be solved are

$$\Sigma_{MA^{(1)}}(\mathbf{k}, \omega) = \frac{1}{N} \sum_{\mathbf{q}_1} f_1^{(1)}(\mathbf{q}_1), \qquad (13)$$

where

$$f_1^{(1)}(\mathbf{q}_1) = G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega) \left[g^2 + \frac{1}{N} \sum_{\mathbf{q}_2} f_2^{(1)}(\mathbf{q}_1, \mathbf{q}_2) \right] \qquad (14)$$

and for all $n \geq 2$,

$$f_n^{(1)}(\mathbf{q}_1, \ldots, \mathbf{q}_n) = \bar{g}_0(\omega - n\Omega) \left[g^2 \sum_{i=1}^{n} f_{n-1}^{(1)}(\ldots, \mathbf{q}_{i-1}, \mathbf{q}_{i+1}, \ldots) \right.$$

$$\left. + \frac{1}{N} \sum_{\mathbf{q}_{n+1}} f_{n+1}^{(1)}(\mathbf{q}_1, \ldots, \mathbf{q}_{n+1}) \right]. \qquad (15)$$

As before, the dependence on $\mathbf{k}, \omega$ is implicitly assumed everywhere. We use the upper labels (1) because these are the approximative solutions corresponding to MA$^{(1)}$.

The solution of this infinite set of recurrent equations is discussed in Appendix B. The end result is

$$\Sigma_{MA^{(1)}}(\omega) = \frac{g^2 \bar{g}_0(\tilde{\omega})}{1 - g^2 \bar{g}_0(\tilde{\omega})[A_2(\omega) - A_1(\omega - \Omega)]}, \qquad (16)$$

where [see Eq. (11)]

$$\tilde{\omega} = \omega - \Omega - g^2 A_1(\omega - \Omega) = \omega - \Omega - \Sigma_{MA}(\omega - \Omega). \qquad (17)$$

The continued fractions $A_1(\omega - \Omega)$, $A_2(\omega)$ are defined in Eq. (12). This expression is slightly more complicated than $\Sigma_{MA^{(0)}}(\omega)$, since it involves two different continued fractions, however, it is still very trivial to compute.

Note that based on this and other results derived in Appendix B, we can now calculate the MA$^{(1)}$ expressions for the generalized Green's functions $f_n^{(1)}(\mathbf{k}, \mathbf{q}_1, \ldots, \mathbf{q}_n, \omega) \sim F_n(\mathbf{k}, \mathbf{q}_1, \ldots, \mathbf{q}_n, \omega)$. These will be more accurate than the values obtained within the MA$^{(0)}$ approximation, where none of the $f_n^{(0)}$ expression had any momentum dependence. These generalized Green's functions contain further information about the polaron, for example, regarding the phonon statistics.

As can be seen from Eq. (16), the self-energy is still momentum independent at this level. However, it is clear that this is because the Holstein model is so featureless. If, e.g., the coupling was dependent on the phonon momentum, then

the first self-energy diagram $\frac{1}{N} \Sigma_{\mathbf{q}} |g(\mathbf{q})|^2 G_0(\mathbf{k} - \mathbf{q}, \omega)$ would be $\mathbf{k}$ dependent, and so would $\Sigma_{MA^{(1)}}(\mathbf{k}, \omega)$ (this diagram is exact at MA$^{(1)}$ level). Indeed, work in progress on generalizing this approach to models with $g(\mathbf{q})$ coupling verifies this.

It is also clear that even for the Holstein model, all expressions $\Sigma_{MA^{(n)}}(\mathbf{k}, \omega)$ with $n \geq 2$ will have momentum dependence, since Holstein self-energy diagrams of second order are momentum dependent. We demonstrate below that this is indeed the case.

Finally, the MA approximation becomes exact in the limit $g \to 0$ and $t \to 0$. The first limit is trivial, since $\Sigma \to 0$. The second is due to the fact that if $t=0$, then free propagators $G_0(\mathbf{k}, \omega)$ are, in fact, independent of $\mathbf{k}$, and thus, the momentum averages become irrelevant. Clearly, the same must hold true for all higher level MA$^{(n)}$ approximations. Indeed, one can verify directly that if $\bar{g}_0(\omega) = (\omega + i\eta)^{-1}$ (corresponding to $t=0$), then the expressions in Eqs. (16) and (11) are equal. The same is true for the MA$^{(2)}$ results we present below.

B. MA$^{(2)}$ level

In this case, the equations to be solved are [compare to the exact Eqs. (7)–(9)]

$$\Sigma_{MA^{(2)}}(\mathbf{k}, \omega) = \frac{1}{N} \sum_{\mathbf{q}_1} f_1^{(2)}(\mathbf{q}_1), \qquad (18)$$

where

$$f_1^{(2)}(\mathbf{q}_1) = G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega) \left[g^2 + \frac{1}{N} \sum_{\mathbf{q}_2} f_2^{(2)}(\mathbf{q}_1, \mathbf{q}_2) \right], \qquad (19)$$

$$f_2^{(2)}(\mathbf{q}_1, \mathbf{q}_2) = G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \omega - 2\Omega) \left\{ g^2 [f_1^{(2)}(\mathbf{q}_1) + f_1^{(2)}(\mathbf{q}_2)] \right.$$

$$\left. + \frac{1}{N} \sum_{\mathbf{q}_3} f_3^{(2)}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) \right\}, \qquad (20)$$

and for all $n \geq 3$,

$$f_n^{(2)}(\mathbf{q}_1, \ldots, \mathbf{q}_n) = \bar{g}_0(\omega - n\Omega) \left[g^2 \sum_{i=1}^{n} f_{n-1}^{(2)}(\ldots, \mathbf{q}_{i-1}, \mathbf{q}_{i+1}, \ldots) \right.$$

$$\left. + \frac{1}{N} \sum_{\mathbf{q}_{n+1}} f_{n+1}^{(2)}(\mathbf{q}_1, \ldots, \mathbf{q}_{n+1}) \right]. \qquad (21)$$

Dependence on $\mathbf{k}, \omega$ is again implicitly assumed.

This can be reduced to a closed system of equations in terms of only $f_1^{(2)}(\mathbf{q}_1)$ and $f_2^{(2)}(\mathbf{q}_1, \mathbf{q}_2)$ after solving for $\frac{1}{N} \Sigma_{\mathbf{q}_3} f_3^{(2)}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$ from Eq. (19). The details are provided in Appendix C.

We use the shorthand notation

$$A_1 \equiv A_1(\omega - 2\Omega), \quad A_2 \equiv A_2(\omega - \Omega), \quad A_3 \equiv A_3(\omega),$$

where the continued fractions are defined in Eq. (12). We also define various momentum averages (dependence on $\mathbf{k}$, ω is implicit):

$$\mathcal{F}_1 = \Sigma_{MA^{(2)}} = \frac{1}{N}\sum_{\mathbf{q}_1} f_1^{(2)}(\mathbf{q}_1),$$

$$\mathcal{F}_2 = \frac{1}{N^2}\sum_{\mathbf{q}_1,\mathbf{q}_2} f_2^{(2)}(\mathbf{q}_1,\mathbf{q}_2) = \frac{2g^2\bar{g}_0(\tilde{\omega})F_1}{1 - g^2\bar{g}_0(\tilde{\omega})(A_3 - A_1)},$$

$$\delta\bar{f}_2(\mathbf{q}_1) = \frac{1}{N}\sum_{\mathbf{q}_2} f_2^{(2)}(\mathbf{q}_1,\mathbf{q}_2) - \mathcal{F}_2.$$

The link between $\mathcal{F}_1$ and $\mathcal{F}_2$ is proven in Eq. (C4).

In terms of these, the closed system of equations to be solved becomes (see Appendix C for more details)

$$f_1^{(2)}(\mathbf{q}_1) = G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega)[g^2 + \delta\bar{f}_2(\mathbf{q}_1) + \mathcal{F}_2],$$

$$\delta\bar{f}_2(\mathbf{q}_1) = g^2\bar{g}_0(\tilde{\omega})[f_1^{(2)}(\mathbf{q}_1) + (A_2 - A_1)\delta\bar{f}_2(\mathbf{q}_1) - 2\mathcal{F}_1]$$
$$+ \frac{g^2}{N}\sum_{\mathbf{q}_2} G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \tilde{\omega})[f_1^{(2)}(\mathbf{q}_2)$$
$$+ (A_2 - A_1)\delta\bar{f}_2(\mathbf{q}_2)].$$

These equations can be solved in a variety of ways. We present here the most efficient solution that we have found, and then comment briefly on other possible solutions. First, given the form of these equations, it is advantageous to introduce the unknown

$$x_{\mathbf{q}} = f_1^{(2)}(\mathbf{q}) + (A_2 - A_1)\delta\bar{f}_2(\mathbf{q}). \tag{22}$$

Consider its Fourier transform at various lattice sites $\mathbf{R}_i$, namely, $x(i) = \frac{1}{N}\sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{R}_i} x_{\mathbf{q}}$. First, observe that $x(0) = \frac{1}{N}\sum_{\mathbf{q}} x_{\mathbf{q}} = \mathcal{F}_1 = \Sigma_{MA^{(2)}}$ by definition.

As shown in Appendix C, the set of two closed equations above can be rewritten as an inhomogeneous equation involving $x(i)$ (at all lattice sites):

$$\sum_j M_{ij}(\mathbf{k},\omega)x(i) = e^{i\mathbf{k}\cdot\mathbf{R}_i}g^2 G_0(-i,\tilde{\tilde{\omega}}), \tag{23}$$

where

$$\tilde{\tilde{\omega}} = \omega - \Omega - \frac{g^2\bar{g}_0(\tilde{\omega})}{1 - g^2\bar{g}_0(\tilde{\omega})(A_2 - A_1)},$$

$G_0(i,\omega) = \frac{1}{N}\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}_i} G_0(\mathbf{k},\omega)$, and the expression of the matrix elements $M_{ij}(\mathbf{k},\omega)$ is given in Eqs. (C5)–(C7). They have simple expressions, involving only (the same) three continued fractions A_1, A_2, and A_3 as well as various $G_0(i,\omega)$ values, therefore they can be calculated easily.

Because at low energies the free propagators in real space decay exponentially, one expects that $x(i)$ decreases fast with increasing distance $\mathbf{R}_i$. Alternatively, consider, for instance, the $f_1^{(2)}(\mathbf{q})$ contribution to $x_{\mathbf{q}}$. When Fourier transformed, the initial state $c_{\mathbf{k}-\mathbf{q}}^\dagger b_{\mathbf{q}}^\dagger|0\rangle$ goes into $c_j^\dagger b_{j+i}^\dagger|0\rangle$, i.e., the phonon is further and further apart from the electron. A similar interpretation can be given for the second contribution to $x(i)$. One expects the amplitudes for such processes to decay with i.

As a result, we can truncate the system of equations (23), assuming that $x(i) = 0$ for all $\mathbf{R}_i$ larger than a cutoff. We vary this cutoff to ensure that convergence has, indeed, been achieved. In this formulation, we find that convergence is reached extremely fast, typically for a cutoff distance of order $5a$ (see Sec. IV). In other words, to obtain $\Sigma_{MA^{(2)}}(\mathbf{k},\omega) = x(0)$, in one dimension we typically have to solve a system of 11 or so inhomogeneous equations, which can be done very efficiently. Higher dimensions imply larger systems, but overall, the numerical task is still trivial and results can be obtained very fast and with little computational resources.

It is important to emphasize that such low cutoffs are *not* inherent in the problem. In fact, one could also solve these equations, for instance, by eliminating $f_1^{(2)}(\mathbf{q})$ to get an equation only in terms of $\delta\bar{f}_2(\mathbf{q}_1)$ and $\mathcal{F}_2 \sim \Sigma$. If one Fourier transforms this, it turns out that cutoffs as large as $100a$ are needed before convergence is achieved, and this is especially so for the polaron+one-phonon continuum (the bound polaron states converge quickly). This is not surprising, since states in the polaron+one-phonon continuum do have a free phonon, i.e., one that could be infinitely far from the polaron. In reality, we expect that if we have a finite but large enough system, all quantities will eventually converge to their bulk values. In particular, here this suggests that one needs to allow the free phonon to be hundreds of sites away from the polaron before convergence for the continuum is achieved.

The much faster convergence for the formulation of Eq. (23) is due in part to the particular choice of variable $x_{\mathbf{q}}$. Even more important is the infinite summation of diagrams performed when the new frequency $\tilde{\tilde{\omega}}$ appears (see Appendix C). Without this, the convergence for continuum energies remains slow and large cutoffs are needed. Examples are discussed in Sec. IV.

Of course, one could also attempt to solve these equations directly in $\mathbf{k}$ space. Without having tried it, we believe this to be an inefficient approach. The goal is to find $\mathcal{F}_1 = \Sigma$, i.e., an average over the Brillouin zone. Within MA$^{(1)}$, $f_1^{(1)}(\mathbf{q}) \sim G_0(\mathbf{k}-\mathbf{q},\tilde{\omega})$ (it is a constant in MA). Presuming that $f_1^{(2)}(\mathbf{q})$ is not too different, it is clear that these functions are of comparable size everywhere in the Brillouin zone, and therefore, one should sample many points in the Brillouin zone to obtain an accurate average.

Finally, going back to Eq. (23), we would like to point out that if we set the cutoff at zero distance, i.e., use $M_{00}x(0) = g^2\bar{g}_0(\tilde{\tilde{\omega}})$, we obtain an analytical, momentum-independent approximation to the true $\Sigma_{MA^{(2)}}(\mathbf{k},\omega)$:

$$\tilde{\Sigma}_{MA^{(2)}}(\omega) = \frac{g^2\bar{g}_0(\tilde{\tilde{\omega}})}{1 - g^2\bar{g}_0(\tilde{\omega})\bar{g}_0(\tilde{\tilde{\omega}})\left[\dfrac{2}{a_{31}(\omega)} - \dfrac{1}{a_{21}(\omega)}\right]}, \tag{24}$$

where

$$a_{ij}(\omega) = 1 - g^2\bar{g}_0(\tilde{\omega})(A_i - A_j).$$

One can think of this as the variant of MA$^{(2)}$ that keeps some of the free propagators of energy $\omega - 2\Omega$ exactly (typically those appearing in noncrossed diagrams), but averages over
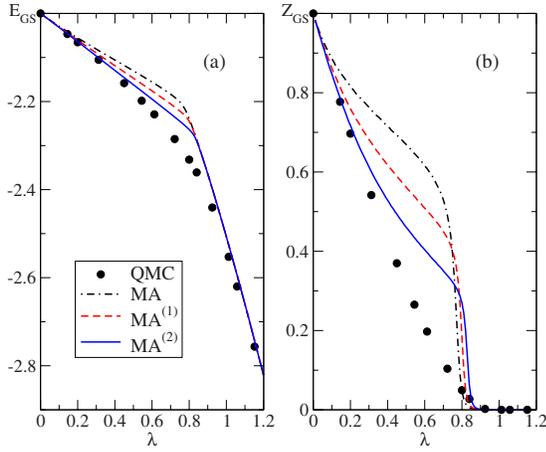
FIG. 2. (Color online) (a) Ground-state energy and (b) ground-state quasiparticle weight, as a function of the effective coupling λ, for $d=1$, $t=1$, and $\Omega/t=0.1$. The QMC results are from Ref. 13.

those that give momentum dependence to the self-energy. The self-energy $\widetilde{\Sigma}_{MA^{(2)}}(\omega)$ is more accurate than $\Sigma_{MA^{(1)}}(\omega)$ but less accurate than $\Sigma_{MA^{(2)}}(\mathbf{k}, \omega)$. Such "zero-cutoff" analytical approximations can be obtained for higher levels of $MA^{(n)}$ quite easily. The full $MA^{(n)}$ for $n > 2$ can also be done. The reduction to a closed system of n coupled equations is always straightforward. Its solution, however, becomes more computationally involved as n increases, and leads to gradually smaller improvements in accuracy.

IV. RESULTS

A detailed comparison of the predictions of the MA approximation vs numerical simulations is available in Ref. 2. Instead of another comprehensive investigation, here we will focus on several properties where the higher level $MA^{(n)}$ approximations show a significant improvement over MA results. The way of extracting quantities of interest from the Green's function, e.g., GS energies, quasiparticle (qp) weights, effective masses, average number of phonons in the polaron cloud, etc., is described in detail in Ref. 2. Throughout we use

$$\lambda = \frac{g^2}{2dt\Omega} \qquad (25)$$

as the effective coupling strength, d being the dimension of the lattice. All energies are measured in units of t.

A. Ground-state properties

The ground-state energies predicted by MA are quite accurate for a large range of parameters. The accuracy is known to worsen as $\Omega/t \rightarrow 0$; however, even for $\Omega/t=0.1$, the MA energies have less than 5% relative error. On the other hand, the GS qp weight for this low Ω/t ratio is quite wrong for intermediary couplings $\lambda \sim 1$, although it does become asymptotically exact, as expected. The comparison with quantum Monte Carlo (QMC) results is shown in Fig. 2,

where we also show the $MA^{(1)}$ and $MA^{(2)}$ predictions.

The accuracy is improved significantly for the higher MA levels. Improvements are observed for all other sets of parameters (not shown) analyzed in Ref. 2, however, for higher Ω/t ratios, MA is much more accurate to begin with, so the supplementary improvements due to $MA^{(1)}$ and $MA^{(2)}$ are comparatively smaller. As discussed, this systematic improvement is expected since all self-energy diagrams become more and more accurate. This is reflected in the sum rules for spectral weight, which are also systematically improved (the link between diagrams and sum rules is discussed at length in Ref. 2). While MA satisfies the first six sum rules exactly, $MA^{(1)}$ satisfies the first eight sum rules exactly, $MA^{(2)}$ satisfies the first ten sum rules exactly, etc. Of course, the accuracy of all higher order sum rules is also systematically improved by the use of more accurate expressions for the diagrams.

This systematic improvement can also be understood in variational terms. As pointed out in Ref. 5, MA is equivalent to a variational approach where the eigenfunctions are built within a restricted Hilbert space that allows phonons only at one site. Indeed, it can be verified straightforwardly that wave functions of the general type $|\Phi\rangle = \sum_{n=0}^{\infty} \sum_{i,j} \alpha_n(i,j) c_i^{\dagger} (b_j^{\dagger})^n |0\rangle$ lead to the same spectrum as MA. Of course, the highest probability is to find the electron and phonon cloud at the same site, however, hopping of the electron away from the phonon cloud is needed to allow for polaron motion.

This observation immediately explains the absence of the polaron+one-phonon continuum, since within this restricted Hilbert space it is not allowed to have a single phonon far from where the polaron is. To compensate for the missing continuum's weight and to fulfill sum rules, the GS qp weight is increased within MA, as seen in Fig. 2(b) and many other examples discussed in Ref. 2.

From this variational point of view, $MA^{(1)}$ almost corresponds to using an enlarged variational Hilbert space; after Fourier transforming to real space, it is straightforward to verify that (see Appendix B) $\delta \bar{f}_n(\mathbf{q}_1) \sim \sum_{i \neq j} e^{i(\mathbf{k}-\mathbf{q}_1) \cdot \mathbf{R}_i + i\mathbf{q}_1 \cdot \mathbf{R}_j} c_i (b_i^{\dagger})^{n-1} b_j^{\dagger} |0\rangle$, i.e., one phonon is now allowed to be away from the main polaron cloud. The equivalence is not exact because the resulting $MA^{(2)}$ equations do not contain a term that would appear in a full variational treatment, so more correctly one could view this as also being accompanied by a change in the effective Hamiltonian. Similarly, $MA^{(2)}$ allows for up to two phonons to be located anywhere away from the main phonon cloud, again with a slight change of the effective Hamiltonian when acting on these isolated phonons.

This systematic increase in the size of the variational Hilbert space is another way to explain the gradual improvement of the GS energy. Also, it is now clear that the polaron+one-phonon continuum should appear at level $MA^{(1)}$ (see below). As a result, the GS qp weight no longer needs to account for it and it decreases, improving the agreement with QMC results as shown in Fig. 2(b). We will return to this issue when we investigate spectral weights. For the time being, we note that in the adiabatic limit $\Omega/t \rightarrow 0$, many phonons can be created in the system at low energetic cost.

In the intermediary region $\lambda \sim 1$ where the polaron cloud is still relatively large, one expects many of these phonons to be relatively far from the polaron and, therefore, a high order n would be required in order to accurately describe them within this approach. As a result, it is expected that the strongly adiabatic regime will not be quantitatively well described for $\lambda \sim 1$ by the low-level MA approximations, even though the qualitative behavior is correctly captured. Of course, this limit can be investigated by other means, such as in Ref. 7 and references therein. However, for most of the parameter space, i.e., any $\Omega/(dt) > 0.1$ or so and any coupling λ, the MA set of approximations gives very easy to evaluate yet remarkably accurate results.

B. Polaron band

We can also track how the lowest eigenstate of momentum **k**, and its various properties, changes with various parameters. Results are shown in Fig. 3 for the energy $E_{\mathbf{k}}$, qp weight $Z_{\mathbf{k}}$, and average number of phonons $N_{\mathrm{ph}}(\mathbf{k})$ for one dimension and two couplings, $\lambda = 0.25$ and $\lambda = 1.00$. We found similar improvements in two-dimensional (2D) cases. Clearly, $MA^{(2)}$ leads to an obvious improvement, even though for this value of $\Omega/t = 0.5$, MA itself is quite accurate already. It should be noted that $\lambda \sim 1$ is where the MA accuracy is generally expected to be at its worst. In particular, for the weak-coupling value $\lambda = 0.25$, we see that MA overestimates the distance to the continuum, i.e., $E_{\pi} - E_0 > \Omega$. This should be Ω, but it is larger for MA because the polaron +one-phonon continuum is not predicted at the correct energy.[2] (For $\lambda = 1$, there is a second bound state between the one shown and the continuum, therefore the bandwidth is much less than Ω.) For $MA^{(1)}$ and $MA^{(2)}$, this problem is indeed fixed, and the polaron dispersion width (at weak couplings) is Ω. All other quantities are also clearly more accurate.

C. Polaron+one-phonon continuum and higher-energy states

In order to understand the effects on higher-energy states, we study the spectral weight $A(\mathbf{k}, \omega) = -\frac{1}{\pi} \mathrm{Im}\, G(\mathbf{k}, \omega)$. As is well known, this is finite only at energies ω, where eigenstates of momentum **k** exist. For discrete (bound) states, the spectral weight is a Lorentzian of width η [see Eq. (1)] and height proportional to the qp weight. In a continuum, the lifetime is determined by $\mathrm{Im}\, \Sigma(\mathbf{k}, \omega)$ and is independent of η, provided that η is chosen small enough.

In Fig. 4(a), we show results for the one-dimensional (1D) spectral weight $A(k=0, \omega)$ vs ω for a relatively weak coupling $\lambda = 0.6$. The MA spectral weight shows two discrete states at low energies, and a continuum starting for $\omega > -1.5t$. Within $MA^{(1)}$, the second peak spreads into a continuum whose lower edge is at roughly Ω above the energy of the GS peak. In fact, since $\bar{g}_0(\omega)$ acquires an imaginary part when $-2dt \leq \omega \leq 2dt$, from Eq. (16) it follows that the $MA^{(1)}$ continuum appears when $\tilde{\omega} \geq -2t \rightarrow \omega > E_{GS}^{(0)} + \Omega$, where $E_{GS}^{(0)}$ is the MA prediction for the GS energy. Since the $MA^{(1)}$ GS energy $E_{GS}^{(1)} < E_{GS}^{(0)}$, it follows that the gap is, in fact, slightly larger than Ω. In $MA^{(2)}$, the weight is redistrib-
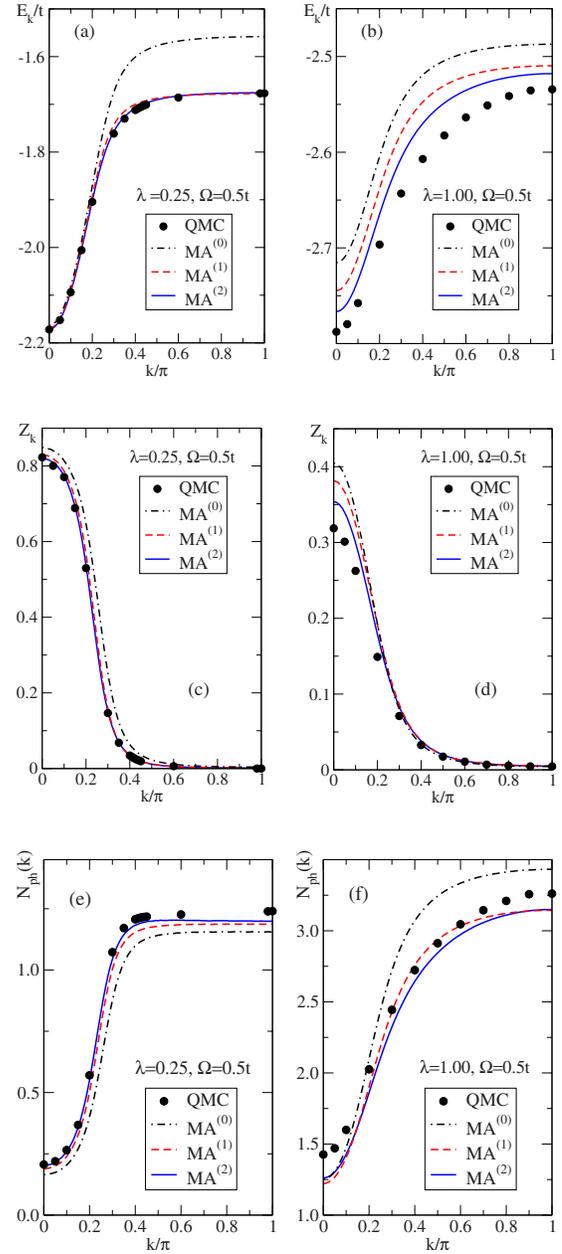
FIG. 3. (Color online) [(a) and (b)] Polaron dispersion E_k, [(c) and (d)] qp weight Z_k, and [(e) and (f)] average number of phonons $N_{\mathrm{ph}}(k)$ vs k, in $d=1$, for $\Omega = 0.5t$ and $\lambda = 0.25$ in (a) (c), and (e), respectively, and $\lambda = 1.00$ in (b), (d), and (f). The QMC results are from Ref. 13.

uted within this continuum, and its lower band edge is at Ω above the ground-state energy, within numerical precision.

The convergence of $MA^{(2)}$ with the cutoff value used to truncate Eq. (23) is shown in Fig. 4(b). For a cutoff value of 0, we obtain the momentum-independent self-energy of Eq. (24), which gives a continuum with a shape rather similar to that predicted by $MA^{(1)}$. As the cutoff value is increased, weight shifts toward the lower band edge. Convergence is reached very quickly, with little difference visible between results corresponding to a cutoff of 5 or 10 [these imply solving an inhomogeneous system of 11 or 21 equations,
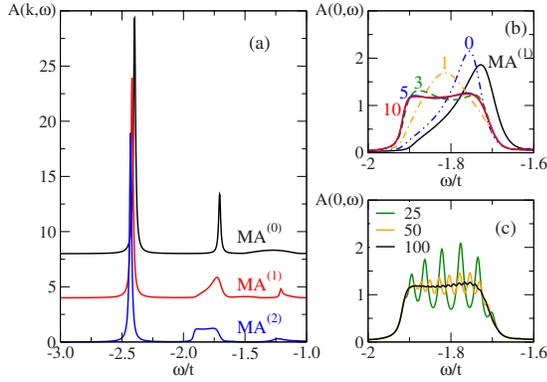
FIG. 4. (Color online) (a) Spectral weight $A(k=0,\omega)$ vs ω in one dimension for $t=1$, $\Omega=0.5$, $\lambda=0.6$, and $\eta=0.01$ in MA, $MA^{(1)}$, and $MA^{(2)}$ (curves are shifted for clarity). (b) Polaron+one-phonon continuum convergence with cutoff within $MA^{(2)}$. Results for cutoffs of 0, 1, 3, 5, and 10 are shown (the last two are almost identical). For comparison, the $MA^{(1)}$ continuum is also shown (black full line); (c) same as in (b) but for an inefficient computation scheme.

respectively, in Eqs. (23)]. Other solutions of the coupled equations (briefly discussed in the previous section) converge much more slowly. In Fig. 4(c), we show results for three cutoffs for such an alternative scheme. Even for a cutoff as large as 100, one can still see small oscillations, very reminiscent of finite-size effects. The finding of an efficient solution for the $MA^{(2)}$ self-energy is, thus, quite important.

Similar behavior is observed for higher couplings, as seen in Figs. 5(a) and 6(a) for intermediate and strong couplings, respectively, $\lambda=1.2$ and 1.8. In both cases, there are now two bound, discrete states below the continuum starting at $E_{GS}+\Omega$ as expected.[8] The weight of the polaron+one-phonon continuum decreases very fast, so that for $\lambda=1.8$, it is barely visible just above the second peak. Figure 6(b) shows it clearly on a logarithmic scale. Interestingly, it is not only the
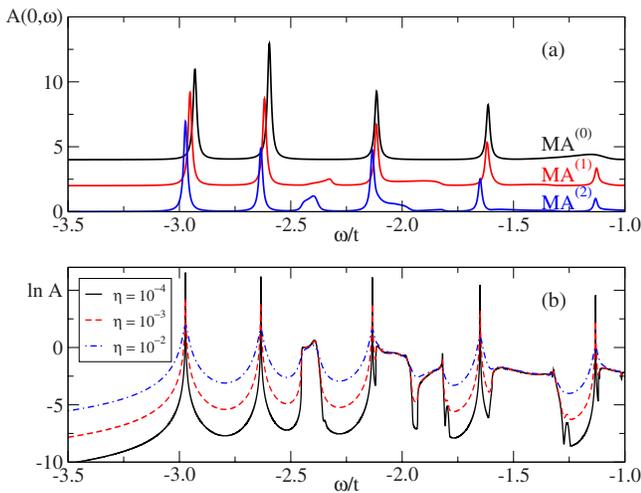
FIG. 5. (Color online) (a) Spectral weight $A(0,\omega)$ vs ω in one dimension for $t=1$, $\Omega=0.5$, $\lambda=1.2$, and $\eta=0.01$ in MA, $MA^{(1)}$, and $MA^{(2)}$ (curves shifted for clarity). (b) $\ln A(0,\omega)$ vs ω for $MA^{(2)}$ and $\eta=10^{-2}$, 10^{-3}, and 10^{-4}. Other parameters are as in (a).
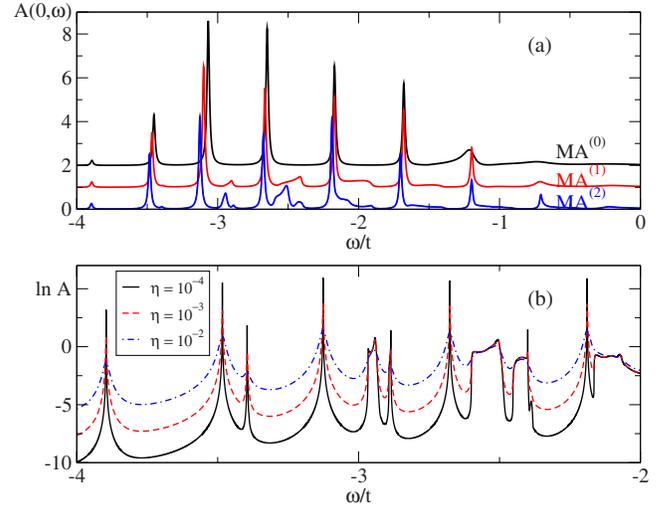
FIG. 6. (Color online) (a) Spectral weight $A(0,\omega)$ vs ω in one dimension for $t=1$, $\Omega=0.5$, $\lambda=1.8$, and $\eta=0.01$ in MA, $MA^{(1)}$, and $MA^{(2)}$ (curves shifted for clarity). (b) $\ln A(0,\omega)$ vs ω for $MA^{(2)}$ and $\eta=10^{-2}$, 10^{-3}, and 10^{-4}. Other parameters are as in (a).

height of this feature that is much smaller as λ increases, but its width as well. In fact, scaling vs η in Fig. 6(b) shows that this is more like a Lorentzian, i.e., a single bound state, and not a finite-width continuum as was the case for lower couplings [see Fig. 5(b)]. This is, in fact, reasonable, since at such large couplings the lowest energy polaron state is basically dispersionless (the effective mass is already considerable and the polaron is well into the small-polaron regime). Since the width of the continuum is due to the polaron dispersion (the phonon being dispersionless), it is reasonable that as the polaron bandwidth decreases exponentially with increasing coupling, so does the width of the polaron+one-phonon continuum.

The higher-energy features are also quite interesting. For the intermediate coupling $\lambda=1.2$, at some distance above the polaron+one-phonon continuum, one can see the feature evolved from what was the third discrete state in the MA approximation. For a large η value, this looks like a continuum, however, scaling with η reveals a discrete state just below another continuum. In fact, the spectrum is broken up into discrete states and continua separated by gaps where no states are present. Of course, the detailed shape of the spectral weight above the third bound state is likely to change as one goes to $MA^{(3)}$ and higher orders; however, we doubt that these gaps should all close and a single continuum should form above $E_{GS}+\Omega$, as is the case at very weak couplings (see below). Instead, to us these results suggest that most weight is inside bound states which are reminiscent of the Lang-Firsov "comb" of discrete states separated by a frequency Ω. Here, the distance between discrete states is generally less than Ω and there are narrow continua in between them, which evolve from lower-energy bound states+one or more phonons. This is nicely illustrated by the results in Fig. 6(b), which show two different features between the third and fourth bound states. The lower-energy one is a continuum that starts roughly at Ω above the second bound state and has a finite width. Since the second bound state still has

some finite bandwidth at this coupling (see below), it seems reasonable to interpret this feature as being the polaron in the second bound state plus one phonon somewhere far from it. The higher-energy feature is much narrower, but close inspection reveals that it also is a continuum, with the lower edge starting at $E_{GS}+2\Omega$, so its origin is obvious. The continua between higher bound states become wider, in agreement with expectations if one assumes that, indeed, they result from adding a distant phonon to a polaron in a higher bound state, which has a larger bandwidth. There is also an overlap between different types of states, which also leads to increased bandwidth.

This structure of the spectral weight explains why the MA (which predicts only bound states at low energies for medium to strong coupling) still obeys sum rules with such good accuracy.[2] Most of the weight is, indeed, in the bound states, not in the narrow continua that appear in between them. These results also show clearly how the convergence toward the Lang-Firsov limit $g \gg t$ is achieved: the width and weight of the continua shrink to zero, and one is left only with the equally spaced discrete states. This is illustrated in Fig. 7, where $k=0$ MA spectral weights are contrasted with MA$^{(2)}$ spectral weights for different λ values. The largest difference is observed at small couplings, where MA predicts the (wrong) continuum pinned at $-2dt+\Omega=-1.5$ for these values, whereas MA$^{(2)}$ clearly shows a continuum starting at Ω above the ground state for as long as it is still visible. For $\lambda > 0.7$ or so, a second bound state splits from this continuum and comes quite close to the GS peak before moving away so that it asymptotically goes to $E_{GS}+\Omega$. The existence of multiple bound states below the first continuum, for moderate and strong couplings, is well known for the Holstein polaron and has also been demonstrated for other polaron models.[8,14] It further serves to validate these approximations. There are clear similarities between the two plots, with most of the weight concentrated in the bound states that are roughly Ω apart, but MA overestimates their weights in order to compensate for the missing narrow continua that appear in between these discrete states. As stated, the precise shape and weight of the higher continua are very likely to change as one goes to a higher level MA approximation; however, we expect the general picture to remain essentially the same.

We are unable to check quantitatively the accuracy of the higher-energy spectral weight against detailed numerical predictions, beyond the sort of comparisons shown for the polaron dispersion in Fig. 3. The reason is that most of the numerical work is focused on computing low-energy properties (a detailed overview of such work is provided in Ref. 2 or, for example, in Ref. 9). The much fewer high-energy results, such as those based on a variational treatment in Ref. 10 and a QMC and/or exact diagonalization approach in Ref. 11, use a rather large η and are already in reasonable agreement with MA, as discussed in Ref. 2. Cluster perturbation theory results such as shown in Ref. 12, while also an approximation, are in good qualitative and quantitative agreement with ours. It would be very interesting to be able to compare our results against detailed high-accuracy, high-energy numerical predictions.

Finally, we contrast the difference between MA and MA$^{(2)}$ spectral weights for different momenta and energies.
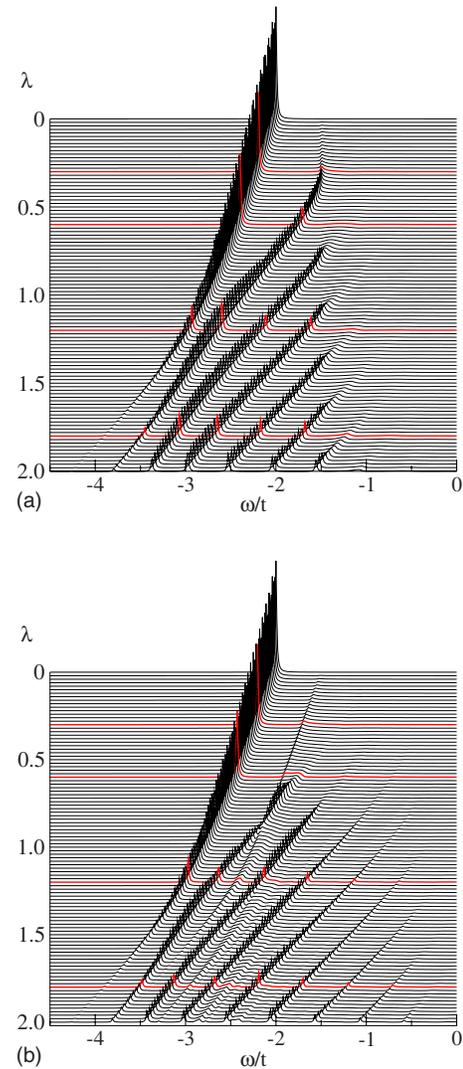
FIG. 7. (Color online) (a) MA spectral weight $A(0,\omega)$ vs ω in one dimension for $t=1$, $\Omega=0.5$, $\eta=0.01$, and λ varying from 0 to 2. (b) Same for MA$^{(2)}$. Curves corresponding to $\lambda=0.3$, 0.6, 1.2, and 1.8 are highlighted.

Typical results are shown in Figs. 8 and 9 for weak, intermediate, and strong couplings $\lambda=0.3$, 0.6, 1.2, and 1.8, respectively. For the weak coupling, as already discussed, the most obvious difference is that the polaron bandwidth is decreased to its correct value of Ω in MA$^{(2)}$. The qp weights and all other features are very similar. Based on the MA$^{(2)}$ results, it is now clear that the strong resonance seen in the electron +phonon continuum occurs at an energy of 2Ω above the ground-state energy. One expects that here is where the second bound state will arise from. For $\lambda=0.6$, Figs. 8(c) and 8(d) show a bigger contrast. Here, MA already predicts a second bound state that has evolved in between the polaron band and the higher-energy continuum. In contrast, MA$^{(2)}$ shows that there is no second bound state yet; however, the electron+one-phonon continuum is split off the higher-energy continuum, which also starts to fractionalize at higher energies (roughly multiples of Ω). Within MA$^{(2)}$, a true second bound state is observed for the higher couplings shown
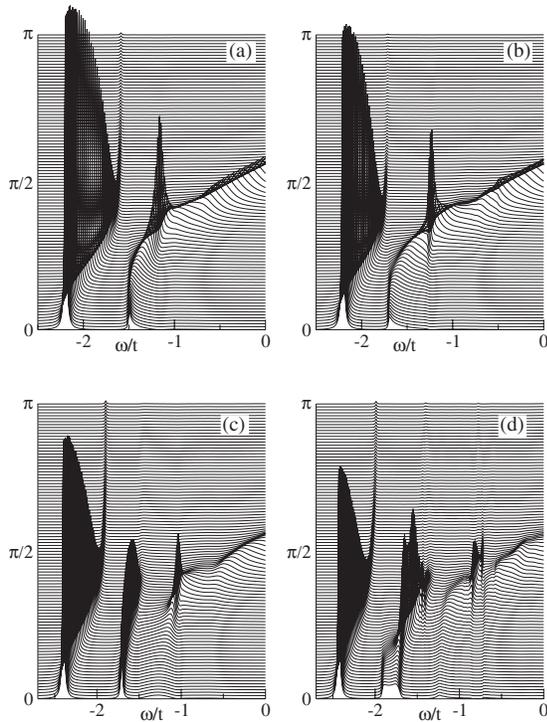
FIG. 8. $A(k,\omega)$ vs k and ω in one dimension for $t=1$, $\Omega=0.5$, $\eta=0.01$, and $\lambda=0.3$ in (a) and (b), and $\lambda=0.6$ in (c) and (d). Results for MA are shown in (a) and (c), MA$^{(2)}$ is shown in (b) and (d).

FIG. 9. $A(k,\omega)$ vs k and ω in one dimension for $t=1$, $\Omega=0.5$, $\eta=0.01$, and $\lambda=1.2$ in (a) and (b), and $\lambda=1.8$ in (c) and (d). Results for MA are shown in (a) and (c), while MA$^{(2)}$ is shown in (b) and (d).

in Fig. 9. In contrast to MA, which shows several bound states which disperse as k increases, MA$^{(2)}$ also clearly shows continua in between these discrete states. These account for some of the spectral weight that was in the MA peaks. These results again suggest a very fractionalized spectrum at intermediate and strong couplings. Instead of the polaron band, a second bound state, and a rather featureless continuum at higher energies, we instead find many sets of discrete states interspersed with continua. As $\lambda \to \infty$, the relative weight in these continua decreases and the spectral weights evolve continuously toward the Lang-Firsov set of discrete states with energies $-g^2/\Omega+n\Omega$.

Qualitatively similar results are observed in higher dimensions. Rather ironically, the most time-consuming part in the MA$^{(2)}$ calculation for higher dimensions is finding the spatial dependence of the noninteracting Green's functions $G_0(i,\Omega)$ which is needed to generate Eq. (23), and not the solving of the system of equations. The reason is that for nearest-neighbor hopping in higher dimensions, the evaluation of these propagators must be done numerically. Of course, one could choose simpler forms of the dispersion $\epsilon_\mathbf{k}$ for which analytical results are possible. However, as we show now, in higher dimensions, the improvements in going to MA$^{(1)}$ and MA$^{(2)}$ are quantitatively smaller, because the relative weight in the continua is reduced compared to the 1D case. This is in agreement with our general observation that MA itself becomes more accurate with increased dimensionality.[2]

In Figs. 10–13, we show the 2D spectral weight $A(\mathbf{k}=0,\omega)$ vs ω for effective couplings $\lambda=0.3$, 0.6, 0.95, and 1.2, both on linear and logarithmic scales. Qualitatively, every-

thing is similar to the behavior seen in the 1D case. Quantitatively, we find that the continuum that appears at Ω above the ground state is broader, but of lower height. In fact, its height is so small that it is invisible on curves like those in Fig. 7, which is the reason why we do not show such curves here. The overall weight in this continuum also decreases much faster with increasing λ. In one dimensions, for $\lambda=1.2$, the first continuum is still clearly visible (see Fig. 5).
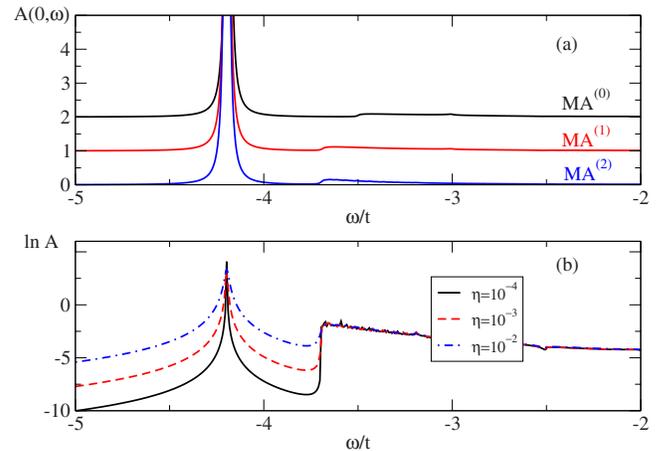
FIG. 10. (Color online) (a) $A(0,\omega)$ vs ω in two dimensions for $t=1$, $\Omega=0.5$, $\lambda=0.3$, and $\eta=0.01$ in MA, MA$^{(1)}$, and MA$^{(2)}$ (curves shifted for clarity). (b) $\ln A(0,\omega)$ vs ω for MA$^{(2)}$ and $\eta=10^{-2}$, 10^{-3}, and 10^{-4}. Other parameters are as in (a).
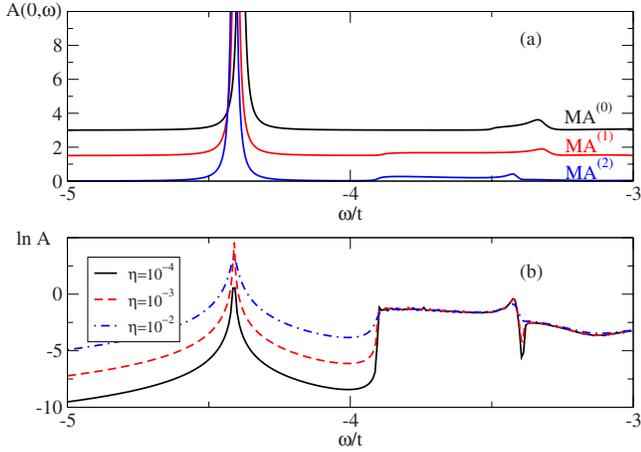
FIG. 11. (Color online) (a) $A(0,\omega)$ vs ω in two dimensions for $t=1$, $\Omega=0.5$, $\lambda=0.6$, and $\eta=0.01$ in MA, MA$^{(1)}$, and MA$^{(2)}$ (curves shifted for clarity). (b) $\ln A(0,\omega)$ vs ω for MA$^{(2)}$ and $\eta=10^{-2}$, 10^{-3}, and 10^{-4}. Other parameters are as in (a).

FIG. 13. (Color online) (a) $A(0,\omega)$ vs ω in two dimensions for $t=1$, $\Omega=0.5$, $\lambda=1.2$, and $\eta=0.01$ in MA, MA$^{(1)}$, and MA$^{(2)}$ (curves shifted for clarity). (b) $\ln A(0,\omega)$ vs ω for MA$^{(2)}$ and $\eta=10^{-2}$, 10^{-3}, and 10^{-4}. Other parameters are as in (a).

For $\lambda=1.8$, it becomes harder to see on the linear scale, but it is clearly seen on the logarithmic scale. By contrast, in two dimensions, for $\lambda=1.2$, the continuum is no longer visible on the linear scale, and even in the logarithmic scale it only barely starts to be visible for $\eta=10^{-4}$ (small shoulder marked by arrow). The spectral weight for a Lorentzian contribution $Z/(\omega+i\eta)$ is $Z\eta/[\pi(\omega^2+\eta^2)]$, so the maximum height of the peak, at resonance, is $Z/\pi\eta$. This is visible only if it is larger than the background, which, of course, depends on how close is the next spectral feature. However, for a very small Z, η has to be really small before the peak is seen.

As in the 1D case, we also observe the fractionalization of the spectrum for moderate and large effective couplings, with a succession of discrete peaks and continua at higher energies. We expect similar behavior to be observed in three dimensions as well. Clearly, the changes in going from MA to MA$^{(1)}$ and MA$^{(2)}$ are quantitatively much smaller in two

dimensions than in one dimension, because the continua have so little weight. We expect the trend to continue in going to three dimensions, meaning that in three dimensions, the difference between MA$^{(2)}$ and MA should be quantitatively even less. Indeed, all the comparisons of MA results with available three-dimensional (3D) numerics, shown in Ref. 2, are already in excellent agreement, even for intermediary couplings $\lambda \sim 1$. As a result, we find it unnecessary to present 3D results, although they can be obtained very straightforwardly.

V. CONCLUSIONS

In summary, we presented a way to systematically improve the MA approximation by systematically improving the accuracy of self-energy diagrams, but in such a way that they can still all be efficiently summed.

This allows us to rather easily fix various known failings of the MA approximation, such as the absence of the polaron+one-phonon continuum at the correct energy, and its momentum-independent self-energy. It also allows us to understand in more detail the effects of the Holstein-type electron-phonon coupling on the polaron spectrum, both at low and high energies. While agreement with exact numerical results is improved, unfortunately there are not many such results for higher-energy states, so detailed comparisons are not possible there. However, the hierarchy of MA$^{(n)}$ approximations clearly provides a simple way toward obtaining quantitatively more and more accurate results for the Green's function of the Holstein polaron, in any dimension and for any free-electron dispersion.

The next direction of obvious interest is to study to what extent this work can be extended to other models, for example, models where the electron-phonon coupling depends on the phonon momentum (e.g., the breathing mode[15] or Fröhlich[16] Hamiltonians). For such models, there are very few reliable high-energy results available. A simple approximation such as MA could easily investigate the whole pa-

FIG. 12. (Color online) (a) $A(0,\omega)$ vs ω in two dimensions for $t=1$, $\Omega=0.5$, $\lambda=0.95$, and $\eta=0.01$ in MA, MA$^{(1)}$, and MA$^{(2)}$ (curves shifted for clarity). (b) $\ln A(0,\omega)$ vs ω for MA$^{(2)}$ and $\eta=10^{-2}$, 10^{-3}, and 10^{-4}. Other parameters are as in (a).

rameter space and identify interesting regimes. Such work is currently in progress.

APPENDIX A: CONTINUED FRACTIONS

Consider the set of recurrence relations $f_{-1} = 1$ and for $n \geqslant 0$,

$$f_n = \alpha_n f_{n-1} + \beta_n f_{n+1},$$

where $f_n \to 0$ as $n \to \infty$. (This is justified for all cases of interest to us since f_n is related to a generalized Green's function with n phonons in the initial state and none in the final state. If n is large enough, the amplitude of probability to evolve between such states must vanish.)

Truncating the system at a large enough N so that $f_{N+1} \approx 0$, we solve backwards to find $f_N = \alpha_N f_{N-1}$, $f_{N-1} = \alpha_{N-1} f_{N-2}/(1 - \beta_{N-1}\alpha_N)\dots$. In general,

$$f_n = \cfrac{\alpha_n f_{n-1}}{1 - \cfrac{\alpha_{n+1}\beta_n}{1 - \cfrac{\alpha_{n+2}\beta_{n+1}}{1 - \cdots}}} = A_n f_{n-1}.$$

If we allow $N \to \infty$, then the solution becomes exact and $A_n = \alpha_n/[1 - \beta_n A_{n+1}]$ are infinite continued fractions.

APPENDIX B: MA$^{(1)}$ DETAILS

First, we use Eq. (15) to find how $\frac{1}{N}\Sigma_{\mathbf{q}_2} f_2^{(1)}(\mathbf{q}_1,\mathbf{q}_2)$ depends on $f_1^{(1)}(\mathbf{q}_1)$. This is then used in Eq. (14) to solve for $f_1^{(1)}(\mathbf{q}_1)$, and Σ follows from Eq. (13).

We introduce the partial momentum averages:

$$\bar{f}_n(\mathbf{q}_1) = \frac{1}{N^{n-1}} \sum_{\mathbf{q}_2,\dots,\mathbf{q}_n} f_n^{(1)}(\mathbf{q}_1,\dots,\mathbf{q}_n), \qquad \text{(B1)}$$

[we use the convention that $\bar{f}_1(\mathbf{q}_1) = f_1^{(1)}(\mathbf{q}_1)$] and the full momentum averages:

$$\mathcal{F}_n = \frac{1}{N^n} \sum_{\mathbf{q}_1,\dots,\mathbf{q}_n} f_n^{(1)}(\mathbf{q}_1,\dots,\mathbf{q}_n) = \frac{1}{N}\sum_{\mathbf{q}_1} \bar{f}_n(\mathbf{q}_1) \qquad \text{(B2)}$$

so that $\Sigma_{MA^{(1)}}(\mathbf{k},\omega) = \mathcal{F}_1$.

First, we average Eq. (15) over all the corresponding momenta to find that for $n \geqslant 2$, $\mathcal{F}_n = \bar{g}_0(\omega - n\Omega)[ng^2\mathcal{F}_{n-1} + \mathcal{F}_{n+1}]$. This has the continued-fraction solution (see Appendix A)

$$\mathcal{F}_2 = g^2 A_2(\omega)\mathcal{F}_1, \qquad \text{(B3)}$$

where $A_2(\omega)$ is a continued fraction defined in Eq. (12) and $\mathcal{F}_1$ is still unknown.

We now average Eq. (15) over all the momenta except $\mathbf{q}_1$ to find that for all $n \geqslant 2$, $\bar{f}_n(\mathbf{q}_1) = \bar{g}_0(\omega - n\Omega)[g^2\mathcal{F}_{n-1} + (n-1)g^2\bar{f}_{n-1}(\mathbf{q}_1) + \bar{f}_{n+1}(\mathbf{q}_1)]$. This can also be solved by rewriting it in terms of $\delta\bar{f}_n(\mathbf{q}_1) = \bar{f}_n(\mathbf{q}_1) - \mathcal{F}_n$. Using the recurrence relation for $\mathcal{F}_n$, we find that

$$\delta\bar{f}_n(\mathbf{q}_1) = \bar{g}_0(\omega - n\Omega)[(n-1)g^2\delta\bar{f}_{n-1}(\mathbf{q}_1) + \delta\bar{f}_{n+1}(\mathbf{q}_1)].$$

This also has solutions in terms of continued fractions, in particular, $\delta\bar{f}_2(\mathbf{q}_1) = g^2 A_1(\omega - \Omega)\delta\bar{f}_1(\mathbf{q}_1)$. Using Eq. (B3), we then find $\bar{f}_2(\mathbf{q}_1) = g^2 A_1(\omega - \Omega)f_1(\mathbf{q}_1) + g^2[A_2(\omega) - A_1(\omega - \Omega)]\mathcal{F}_1$. This expression is now inserted in Eq. (14) to give an equation with only $f_1^{(1)}(\mathbf{q}_1)$ and $\mathcal{F}_1 = \frac{1}{N}\Sigma_{\mathbf{q}_1} f_1^{(1)}(\mathbf{q}_1)$ as unknowns, which can be rearranged as

$$[(G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega))^{-1} - g^2 A_1(\omega - \Omega)]f_1^{(1)}(\mathbf{q}_1)$$
$$= g^2 + g^2[A_2(\omega) - A_1(\omega - \Omega)]\mathcal{F}_1.$$

However, $[G_0(\mathbf{k},\omega)]^{-1} - a(\omega) = \omega - \epsilon_{\mathbf{k}} + i\eta - a(\omega) = [G_0(\mathbf{k}, \omega - a(\omega))]^{-1}$ and, therefore, we find

$$f_1^{(1)}(\mathbf{q}_1) = G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega - g^2 A_1(\omega - \Omega))$$
$$\times [g^2 + g^2[A_2(\omega) - A_1(\omega - \Omega)]\mathcal{F}_1].$$

After momentum averaging over $\mathbf{q}_1$ on both sides, we find $\mathcal{F}_1 = \Sigma$ [see Eq. (16)]. Based on the knowledge of $f_1^{(1)}(\mathbf{q}_1)$ and the various partial and total averages, one can now compute $f_2^{(1)}(\mathbf{q}_1, \mathbf{q}_2)$, etc.

APPENDIX C: MA$^{(2)}$ DETAILS

As for MA$^{(1)}$, the strategy is to solve Eq. (21) to find how $\frac{1}{N}\Sigma_{\mathbf{q}_3} f_3^{(2)}(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3)$ depends on $f_2^{(2)}(\mathbf{q}_1,\mathbf{q}_2)$. Using this in Eq. (20) allows us to solve it in conjunction with Eq. (19) to find $f_1^{(2)}(\mathbf{q}_1)$ and, therefore, Σ.

We introduce various partial momentum averages:

$$\bar{\bar{f}}_n(\mathbf{q}_1, \mathbf{q}_2) = \frac{1}{N^{n-2}} \sum_{\mathbf{q}_3,\dots,\mathbf{q}_n} f_n^{(2)}(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n),$$

$$\bar{f}_n(\mathbf{q}_1) = \frac{1}{N^{n-1}} \sum_{\mathbf{q}_2,\dots,\mathbf{q}_n} f_n^{(2)}(\mathbf{q}_1, \dots, \mathbf{q}_n),$$

where we take $\bar{\bar{f}}_1(\mathbf{q}_1) = \bar{f}_1(\mathbf{q}_1) = f_1^{(2)}(\mathbf{q}_1)$ and $\bar{\bar{f}}_2(\mathbf{q}_1, \mathbf{q}_2) = f_2^{(2)}(\mathbf{q}_1, \mathbf{q}_2)$. We also define the full momentum averages:

$$\mathcal{F}_n = \frac{1}{N}\sum_{\mathbf{q}_1,\dots,\mathbf{q}_n} f_n^{(2)}(\mathbf{q}_1, \dots, \mathbf{q}_n) = \frac{1}{N}\sum_{\mathbf{q}_1} \bar{f}_n(\mathbf{q}_1). \qquad \text{(C1)}$$

The functions $\bar{\bar{f}}_n$, $\bar{f}_n$, and $\mathcal{F}_n$ should also carry the upper label (2) since their values are different from those obtained within MA$^{(1)}$, however, we do not write it explicitly to simplify the notation somewhat.

The solutions for $\mathcal{F}_n$ and $\bar{f}_n$ proceed just as in the MA$^{(1)}$ case, with one difference. Averaging Eq. (21) over all corresponding momenta, we find that now only for $n \geqslant 3$, $\mathcal{F}_n = \bar{g}_0(\omega - n\Omega)[ng^2\mathcal{F}_{n-1} + \mathcal{F}_{n+1}]$; therefore, this recurrence relation now ends with

$$\mathcal{F}_3 = g^2 A_3(\omega) \mathcal{F}_2. \tag{C2}$$

Similarly, after using the same approach as for MA$^{(1)}$, one finds that $\delta\bar{f}_3(\mathbf{q}_1) = \bar{f}_3(\mathbf{q}_1) - \mathcal{F}_3 = g^2 A_2(\omega - \Omega) \delta\bar{f}_2(\mathbf{q}_1)$ so that $\bar{f}_3(\mathbf{q}_1) = g^2 A_2(\omega - \Omega) \bar{f}_2(\mathbf{q}_1) + g^2[A_3(\omega) - A_2(\omega - \Omega)]\mathcal{F}_2$. Of course, the functions $\mathcal{F}_2$ and $\bar{f}_2(\mathbf{q}_1)$ are still unknown. Finally, we can proceed to calculate $\bar{\bar{f}}_3(\mathbf{q}_1, \mathbf{q}_2)$, which is needed in Eq. (20).

We momentum average Eq. (21) over all momenta except $\mathbf{q}_1$ and $\mathbf{q}_2$ to find

$$\bar{\bar{f}}_n(\mathbf{q}_1, \mathbf{q}_2) = \bar{g}_0(\omega - n\Omega)[g^2[\bar{f}_{n-1}(\mathbf{q}_1) + \bar{f}_{n-1}(\mathbf{q}_2)] \\ + (n-2)g^2\bar{\bar{f}}_{n-1}(\mathbf{q}_1, \mathbf{q}_2) + \bar{\bar{f}}_{n+1}(\mathbf{q}_1, \mathbf{q}_2)].$$

Let $\delta\bar{\bar{f}}_n(\mathbf{q}_1, \mathbf{q}_2) = \bar{\bar{f}}_n(\mathbf{q}_1, \mathbf{q}_2) - \bar{f}_n(\mathbf{q}_1) - \bar{f}_n(\mathbf{q}_2) + \mathcal{F}_n$. For these, the recurrence relations are $\delta\bar{\bar{f}}_n(\mathbf{q}_1, \mathbf{q}_2) = \bar{g}_0(\omega - n\Omega)[(n-2)g^2\delta\bar{\bar{f}}_{n-1}(\mathbf{q}_1, \mathbf{q}_2) + \delta\bar{\bar{f}}_{n+1}(\mathbf{q}_1, \mathbf{q}_2)]$ with the solution $\delta\bar{\bar{f}}_3(\mathbf{q}_1, \mathbf{q}_2) = g^2 A_1(\omega - 2\Omega) \delta\bar{\bar{f}}_2(\mathbf{q}_1, \mathbf{q}_2)$, from which we find $\bar{\bar{f}}_3(\mathbf{q}_1, \mathbf{q}_2) = g^2 A_1 f_2^{(2)}(\mathbf{q}_1, \mathbf{q}_2) + g^2(A_2 - A_1)[\delta\bar{f}_2(\mathbf{q}_1) + \delta\bar{f}_2(\mathbf{q}_2)] + g^2[A_3 - A_1]\mathcal{F}_2$, i.e., only in terms of various partial averages of $f_2^{(2)}(\mathbf{q}_1, \mathbf{q}_2)$. Throughout we use the shorthand notation

$$A_1 \equiv A_1(\omega - 2\Omega), \quad A_2 \equiv A_2(\omega - \Omega), \quad A_3 \equiv A_3(\omega).$$

We now substitute $\bar{\bar{f}}_3(\mathbf{q}_1, \mathbf{q}_2)$ in Eq. (20), which gives

$$f_2(\mathbf{q}_1, \mathbf{q}_2) = g^2 G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \tilde{\omega})[f_1^{(2)}(\mathbf{q}_1) + f_1^{(2)}(\mathbf{q}_2) \\ + (A_2 - A_1)[\delta\bar{f}_2(\mathbf{q}_1) + \delta\bar{f}_2(\mathbf{q}_2)] + [A_3 - A_1]\mathcal{F}_2]. \tag{C3}$$

We used the fact that $[G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \omega - 2\Omega)]^{-1} - g^2 A_1 = [G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \tilde{\omega})]^{-1}$, with $\tilde{\omega} = \omega - 2\Omega - g^2 A_1$.

Momentum averaging Eq. (C3) over both momenta leads immediately to $\mathcal{F}_2 = g^2 \bar{g}_0(\tilde{\omega})[2\mathcal{F}_1 + (A_3 - A_1)\mathcal{F}_2]$, and therefore,

$$\mathcal{F}_2 = \frac{2g^2 \bar{g}_0(\tilde{\omega})\mathcal{F}_1}{1 - g^2 \bar{g}_0(\tilde{\omega})[A_3 - A_1]}. \tag{C4}$$

It follows that all higher order total averages $\mathcal{F}_n$ are proportional to $\mathcal{F}_1 = \Sigma$.

Since only $\bar{f}_2(\mathbf{q}_1)$ is needed in Eq. (19), we can obtain it by momentum averaging Eq. (C3) over $\mathbf{q}_2$. This leads directly to the closed system of coupled equations:

$$f_1^{(2)}(\mathbf{q}_1) = G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega)[g^2 + \delta\bar{f}_2(\mathbf{q}_1) + \mathcal{F}_2],$$

$$\delta\bar{f}_2(\mathbf{q}_1) = g^2 \bar{g}_0(\tilde{\omega})[f_1^{(2)}(\mathbf{q}_1) + (A_2 - A_1)\delta\bar{f}_2(\mathbf{q}_1) - 2\mathcal{F}_1] \\ + \frac{g^2}{N}\sum_{\mathbf{q}_2} G_0(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2, \tilde{\omega})[f_1^{(2)}(\mathbf{q}_2) \\ + (A_2 - A_1)\delta\bar{f}_2(\mathbf{q}_2)].$$

To solve it, we first rewrite these as a single equation in terms of the unknowns $x_\mathbf{q} = f_1^{(2)}(\mathbf{q}) + (A_2 - A_1)\delta\bar{f}_2(\mathbf{q})$. This is obtained by multiplying the first equation by $[A_2 - A_1 + G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega)]$ and then adding the result to the second equation. We introduce the shorthand notation:

$$a_{ij} \equiv a_{ij}(\omega) = 1 - g^2 \bar{g}_0(\tilde{\omega})[A_i - A_j]$$

and use the identities $[G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega)]^{-1} - g^2 \bar{g}_0(\tilde{\omega})\{1 + (A_2 - A_1)[G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega)]^{-1}\} = [1 - g^2 \bar{g}_0(\tilde{\omega})(A_2 - A_1)][G_0(\mathbf{k} - \mathbf{q}_1, \tilde{\tilde{\omega}})]^{-1}$, where

$$\tilde{\tilde{\omega}} = \omega - \Omega - \frac{g^2 \bar{g}_0(\tilde{\omega})}{a_{21}},$$

and $G_0(\mathbf{k} - \mathbf{q}_1, \tilde{\tilde{\omega}})\{1 + (A_2 - A_1)[G_0(\mathbf{k} - \mathbf{q}_1, \omega - \Omega)]^{-1}\} = A_2 - A_1 + G_0(\mathbf{k} - \mathbf{q}_1, \tilde{\tilde{\omega}})/a_{21}$. Both identities are based on the definition $[G_0(\mathbf{k}, \omega)]^{-1} = \omega - \epsilon_\mathbf{k} + i\eta$.

With these, the final equation for $x_\mathbf{q}$ becomes

$$a_{21}x_\mathbf{q} = G_0(\mathbf{k} - \mathbf{q}_1, \tilde{\tilde{\omega}})\left[g^2 + \frac{a_{21} - a_{31}}{a_{21}}\mathcal{F}_2\right] - a_{31}(A_2 - A_1)\mathcal{F}_2 \\ + \left[A_2 - A_1 + \frac{G_0(\mathbf{k} - \mathbf{q}_1, \tilde{\tilde{\omega}})}{a_{21}}\right] \\ \times \frac{g^2}{N}\sum_{\mathbf{q}_2} G_0(\mathbf{k} - \mathbf{q} - \mathbf{q}_2, \tilde{\omega})x_{\mathbf{q}_2}.$$

We now Fourier transform $x(i) = \frac{1}{N}\sum_\mathbf{q} e^{i\mathbf{q}\cdot\mathbf{R}_i}x_\mathbf{q}$. From the definition of $x_\mathbf{q}$, it follows immediately that $x(0) = \mathcal{F}_1 = \Sigma$. Since $\mathcal{F}_2 \sim \mathcal{F}_1 = x(0)$ [see Eq. (C4)], the resulting system of inhomogeneous equations is

$$\sum_j M_{ij}(\mathbf{k}, \omega)x(j) = e^{i\mathbf{k}\cdot\mathbf{R}_i}g^2 G_0(-i, \tilde{\omega}),$$

where $G_0(i, \omega) = \frac{1}{N}\sum_\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{R}_i}G_0(\mathbf{k}, \omega)$, and the matrix elements are

$$M_{00} = 1 - g^2 \bar{g}_0(\tilde{\omega})\bar{g}_0(\tilde{\tilde{\omega}})\left(\frac{2}{a_{31}} - \frac{1}{a_{21}}\right), \tag{C5}$$

$$M_{i0} = -g^2 \bar{g}_0(\tilde{\omega})e^{i\mathbf{k}\cdot\mathbf{R}_i}G_0(-i, \tilde{\tilde{\omega}})\left(\frac{2}{a_{31}} - \frac{1}{a_{21}}\right), \tag{C6}$$

for $i \neq 0$, and for both $i, j \neq 0$:

$$M_{ij} = a_{21}\delta_{i,j} - g^2 e^{i\mathbf{k}\cdot\mathbf{R}_i}G_0(j, \tilde{\omega}) \\ \times \left[(A_2 - A_1)\delta_{i,-j} + \frac{G_0(-i-j, \tilde{\tilde{\omega}})}{a_{21}}\right]. \tag{C7}$$

We truncate this system by keeping only sites $\mathbf{R}_i$ within a certain distance from the origin (see discussion in main text) and solve it numerically. The self-energy is

$$\Sigma_{MA^{(2)}}(\mathbf{k}, \omega) = x(0).$$

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

[2] G. L. Goodvin, M. Berciu, and G. A. Sawatzky, Phys. Rev. B **74**, 245104 (2006).

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

[4] V. S. Viswanath and G. Müller, *The Recursion Method* (Springer-Verlag, Berlin, 1994).

[5] The real-space meaning of MA was pointed to us by O. S. Barišić [Phys. Rev. Lett. **98**, 209701 (2007)] and M. Berciu [Phys. Rev. Lett. **98**, 209702 (2007)].

[6] E. N. Economou, *Green's Functions in Quantum Physics* (Springer-Verlag, Berlin, 1983).

[7] For example, see O. S. Barišić and S. Barišić, Eur. Phys. J. B **54**, 1 (2006); O. S. Barišić, Phys. Rev. B **73**, 214304 (2006).

[8] J. Bonča, S. A. Trugman, and I. Batistić, Phys. Rev. B **60**, 1633 (1999).

[9] For a review, see H. Fehske and S. A. Trugman, in *Polarons in Advanced Materials*, edited by A. S. Alexandrov (Canopus, Bath/Springer-Verlag, Bath, 2007).

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

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

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

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

[14] A. S. Mishchenko, N. Nagaosa, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **66**, 020301(R) (2002).

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

[16] H. Fröhlich, Adv. Phys. **3**, 325 (1954).