PAPER • OPEN ACCESS

Polarons in spinless metals—a variational solution

Journal of Physics: Materials

PAPER

Polarons in spinless metals—a variational solution

M Berciu

Department of Physics and Astronomy, University of British Columbia, Vancouver V6T 1Z1, BC, Canada
Stewart Blusson Quantum Matter Institute, University of British Columbia, Vancouver V6T 1Z4, BC, Canada

E-mail: berciu@phas.ubc.ca

Abstract

We propose a simple variational solution for calculating one-particle spectral functions in lattice models of spinless metals with strong electron-phonon coupling. It is based on a generalization of the Momentum Average variational approximation for single polarons, combined with the assumption that the other fermions in the system are locked into an inert Fermi sea. We expect the method to be accurate for fermion addition spectral functions in metals with a small Fermi energy (nearly empty band), and for fermion removal spectral functions in metals with a large Fermi energy (nearly full band), provided that the characteristic phonon frequency is not too small. Both these regions are far from the region where the Migdal theorem holds, thus our results offer new insights into polaronic behavior in a largely unexplored part of the parameter space. Here, we show results for the Holstein coupling in one-dimension and present ways to gauge their accuracy, but ultimately this will need to be verified against numerical calculations. This variational method can be extended straightforwardly to higher dimensions and other forms of electron-phonon coupling.
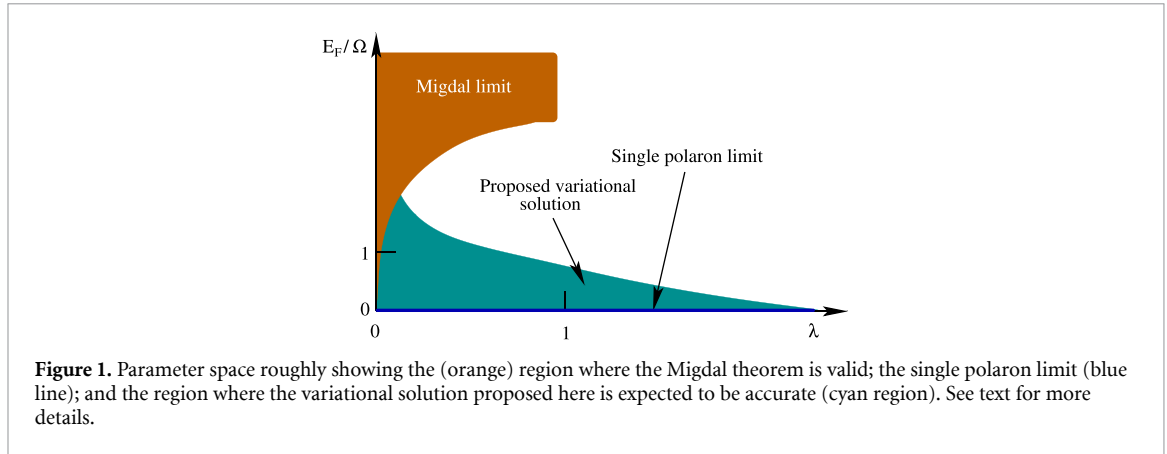
1. Introduction

One of the important challenges in condensed matter physics is to fully understand the effects of electron-phonon coupling on the properties of quasiparticles in metals. To set the stage for the following discussion, we choose E_F/Ω and λ to span the interesting parameter space, where E_F is the Fermi energy of the bare carriers of concentration x, Ω is the characteristic energy of the phonon branch of interest, and λ is a dimensionless effective electron-phonon coupling defined so as to distinguish weak coupling $\lambda \ll 1$ from strong coupling $\lambda \geqslant 1$. (Explicit expressions are provided below for our specific model). This choice means that other ingredients relevant in real materials—coupling to multiple phonons, presence of disorder, bare electron–electron interactions, *etc*—will be ignored hereafter, as a reasonable first step in attempting to solve this difficult problem.

Furthermore, we emphasize that we are only concerned here with *metallic* ground-states and their quasiparticles. We do not consider insulating ground-states due to a Peierls instability [1, 2] and/or charge density waves [3], known to arise at special carrier concentrations x whose corresponding metallic Fermi surface would be nested.

Another complication we wish to ignore is an instability to a superconducting ground-state, whether of the well established Bardeen–Cooper–Schrieffer type at weak electron-phonon coupling [4], or a more exotic Bose–Einstein condensation of a bipolaron liquid that might occur at strong electron-phonon couplings [5, 6]. To achieve this, we further restrict the following discussion to a *liquid of spinless fermions* (alternatively, one can think of fully spin-polarized charge carriers).

Even with all these multiple simplifications, the only regions of the parameter space where we have a good understanding of the properties of the resulting quasiparticles are (i) the single polaron limit corresponding to $E_F/\Omega = 0$, at any λ; and (ii) the Migdal region, where the Migdal theorem holds. The precise boundaries of the latter are not well established. Migdal argued that crossed diagrams can be ignored in the electronic self-energy for $E_F/\Omega \gg 1$, because phase space considerations lead to a decrease of order Ω/E_F of their contribution compared to that of non-crossed diagrams of the same order [7, 8]. This

Figure 1. Parameter space roughly showing the (orange) region where the Migdal theorem is valid; the single polaron limit (blue line); and the region where the variational solution proposed here is expected to be accurate (cyan region). See text for more details.

argument becomes questionable when the number of relevant crossed diagrams vastly exceeds the number of uncrossed diagrams, as is the case at sufficiently strong couplings where one needs to sum up to high orders before convergence is achieved. This explains why $E_F/\Omega \gg 1$ is supplemented by $\lambda < 1$, restricting the Migdal region to the top-left of the parameter space shown in figure 1. We 'tapered' this region all the way to $E_F/\Omega \to 0$ for sufficiently small λ, because if the coupling is weak enough, then first order perturbation theory becomes sufficiently accurate and the self-energy reduces to the lowest order diagram. As a result, crossed diagrams can still be ignored (even though here they are not small compared to non-crossed diagrams of same order) because they only appear at higher orders. We note again that the precise shape of this region is not known. In fact, as we argue below, it likely depends on additional parameters.

Nevertheless, it is clear that so far, we do not have good knowledge regarding the quasiparticles of spinless metals with $E_F/\Omega \leqslant 1$ at medium and strong electron-phonon coupling $\lambda \sim 1$. We begin to address this challenge here, by generalizing the momentum average (MA) approximation—a variational approximation known to work well at all λ in the single polaron limit [9–11]—to low concentrations x. The rough area where we believe its predictions to be accurate is colored cyan in figure 1, although the precise location of the top border depends on additional parameters.

The work presented here is limited to one-dimension (1D) and to the Holstein electron-phonon coupling, but can be straightforwardly extended to higher dimensions [9, 10] and different electron-phonon couplings [12–15], as well as account for disorder [16, 17], for multiple phonon modes [18] and for non-linear coupling [19], following the corresponding single polaron MA solutions.

The paper is organized as follows: section 2 briefly reviews the model, and its variational solution is discussed in section 3. Section 4 shows representative results, and section 5 discusses them and offers perspectives for further generalizations.

2. Model

As already mentioned, we study a 1D metal of spinless fermions interacting with an Einstein branch of phonons through a Holstein coupling:

$$\mathcal{H} = T + \mathcal{H}_{\text{ph}} + \mathcal{H}_{\text{e-ph}}. \tag{1}$$

Here $T = -t \sum_i (c_i^\dagger c_{i+1} + h.c.)$ is the hopping of the spinless fermions, c_i being the annihilation operator for a fermion at site i of the chain with $N \to \infty$ sites. We set the lattice constant $a = 1$ and also take $\hbar = 1$. For a concentration $x = N_f/N$, the Fermi wavevector is $k_F = \pi x$ and the Fermi energy is $E_F = \epsilon(k_F) - \epsilon(0) = 2t[1 - \cos(\pi x)]$, where $\epsilon(k) = -2t \cos k$ is the bare fermion dispersion.

Ignoring the zero-point energy, the phonon energy is $\mathcal{H}_{\text{ph}} = \Omega \sum_i b_i^\dagger b_i$, where b_i is an annihilation operator for a phonon at site i. The Holstein coupling [20, 21] has the well-known form $\mathcal{H}_{\text{e-ph}} = g \sum_i c_i^\dagger c_i (b_i + b_i^\dagger)$.

In the single particle limit, it is customary to define the dimensionless effective coupling as:

$$\lambda = \frac{g^2}{2 \, t\Omega}. \tag{2}$$

We will use this definition, and we will analyze its suitability to characterize the finite-x region below.

At finite concentrations, it is convenient to rewrite the Hamiltonian in terms of dressed boson operators $B_i = b_i + gx/\Omega$, so as to account for the average distortion due to the uniform fermion density. We note that

this is an exact reformulation:

$$\mathcal{H} = -N\frac{g^2 x^2}{\Omega} + T + \tilde{\mathcal{H}}_{\mathrm{ph}} + \tilde{\mathcal{H}}_{\mathrm{e\text{-}ph}} \tag{3}$$

where now $\tilde{\mathcal{H}}_{\mathrm{ph}} = \Omega \sum_i B_i^\dagger B_i$ and

$$\tilde{\mathcal{H}}_{\mathrm{e\text{-}ph}} = g \sum_i (c_i^\dagger c_i - x)(B_i + B_i^\dagger). \tag{4}$$

The mean-field solution corresponds to setting $\tilde{\mathcal{H}}_{\mathrm{e\text{-}ph}} \to 0$, i.e. to ignoring the small fluctuations of both the fermion concentration, and of the dressed phonon displacements. The mean-field ground-state is then $|\mathrm{mf}\rangle = |FS\rangle \otimes |\tilde{0}\rangle$, where

$$|FS\rangle = \prod_{|k| \leqslant k_F} c_k^\dagger |0\rangle \tag{5}$$

is the fermion Fermi sea, and

$$|\tilde{0}\rangle = \prod_i e^{-\frac{g^2 x^2}{2\Omega^2} - \frac{gx}{\Omega}b_i^\dagger} |0\rangle \tag{6}$$

describes the uniform coherent distortion that is the vacuum for the dressed phonon operators, $B_i|\tilde{0}\rangle = 0$. The corresponding mean-field energy is:

$$E_{\mathrm{mf}} = -N\frac{g^2 x^2}{\Omega} + \sum_{|k| \leqslant k_F} \epsilon(k). \tag{7}$$

Thus, at the mean-field level, the fermions' dispersion is not affected by the electron-phonon coupling. The only effect of the latter is a lowering of the total energy due to a uniform lattice distortion driven by the uniform fermion density.

3. Variational solution for one-particle propagators at finite carrier concentrations

To move beyond the underwhelming mean-field solution, we use a variational method to calculate fermion addition and removal propagators, to try to better capture the effects of the electron-phonon coupling on the formation of quasiparticles and their dispersion. Because we work in a *canonical formulation*, we treat the two propagators separately as they characterize Hilbert spaces with different numbers of fermions $N_f \pm 1$.

3.1. Fermion addition
The fermion addition propagator is defined as:

$$G_A(k,z) = \langle \mathrm{GS}|c_k \hat{G}_A(z + E_{\mathrm{GS}})c_k^\dagger|\mathrm{GS}\rangle \tag{8}$$

where $|\mathrm{GS}\rangle$ is the ground-state of the system with N_f fermions, E_{GS} is its energy, $\hat{G}_A(z) = [z - \mathcal{H}]^{-1}$ is the resolvent and $z = \omega + i\eta$ is the energy plus an artificial broadening $\eta \to 0$.

The first obvious obstacle in trying to calculate this propagator is that we do not know $|\mathrm{GS}\rangle$ and E_{GS}. To make progress, we limit our calculation to small concentrations $x = N_f/N$, where the probability of two fermions to meet is already low. Additionally, we note that for spinless fermions we expect the phonon-mediated effective interactions to be repulsive because Pauli's principle prevents two fermions from efficiently sharing the same distortion cloud. As a result, it is a reasonable starting point to assume that the lattice distortion dressing the additional fermion is entirely due to its coupling to phonons, while the other fermions act simply to block states in the Fermi sea and thus limit its scattering phase-space.

Thus, we impose the variational constraint that N_f of the fermions are *always* in the Fermi sea state $|FS\rangle$ and only allow the additional fermion to turn into a quasiparticle:

$$G_A(k,z) \approx \langle \mathrm{mf}|c_k \hat{G}_A(z + E_{\mathrm{mf}})c_k^\dagger|\mathrm{mf}\rangle. \tag{9}$$

This propagator can be calculated similarly to the MA solution for the single carrier limit, as we show next.

The MA variational space is characterized by:

(a) the size of the phonon cloud, i.e. the number of consecutive sites over which it is allowed to spread. There are no constraints on how many phonons are at any of these cloud sites, nor on the distance between the cloud and the fermion that is creating it. In the single carrier limit we have implemented calculations with up to a three-site cloud [13]. For the Holstein model, a one-site cloud is very accurate for $\Omega/t \geqslant 0.5$ or so, and a three-site cloud further improves accuracy down to $\Omega/t \geqslant 0.1$ or so [10, 11, 15]. Here we will implement both one-site and three-site cloud solutions, and only show results for $\Omega/t \geqslant 0.5$.

(b) the number n of free phonons which are not part of the polaron cloud defines the version $\mathrm{MA}^{(n)}$ of the approximation [11]. Here we implement both $n = 0$ and $n = 1$ solutions. The former describes well the bottom of the polaron band but misplaces the location of the polaron+one-phonon continuum defining the top of the band, as those states imply the existence of at least one free phonon.

We now briefly review the main steps in the derivation of the one-site $\mathrm{MA}^{(0)}$ solution, to show that it mirrors the single polaron result, except that the momentum-averaged free propagator is replaced with the one for the given fermion concentration x. The same holds true for all other MA versions, hence their corresponding expressions can be taken directly from previous work.

To generate an equation of motion (EOM) we use Dyson's identity $\hat{G}_A(z) = \hat{G}_A^0(z) + \hat{G}_A^0(z)\tilde{\mathcal{H}}_{\text{e-ph}}\hat{G}_A(z)$, where $\hat{G}_A^0(z) = [z - \mathcal{H}_0]^{-1}$ is the resolvent for $\mathcal{H}_0 = \mathcal{H} - \tilde{\mathcal{H}}_{\text{e-ph}}$. Next we use:

$$\hat{G}_A^0(z + E_{\text{mf}})c_k^\dagger|\text{mf}\rangle = \Theta(|k| - k_F)G_0(k,z)c_k^\dagger|\text{mf}\rangle,$$

where the Heaviside function $\Theta(x)$ insures that the additional fermion's momentum is above the Fermi sea, and $G_0(k,z) = 1/[z - \epsilon(k)]$ is the free fermion propagator. The EOM becomes $G_A(k,z) = \Theta(|k| - k_F)G_0(k,z)$ $\left[1 + \frac{g}{\sqrt{N}}\sum_i e^{ikR_i}\langle\text{mf}|c_k\hat{G}_A(z + E_{\text{mf}})\tilde{\mathcal{H}}_{\text{e-ph}}c_i^\dagger|\text{mf}\rangle\right]$. Next, we replace $\tilde{\mathcal{H}}_{\text{e-ph}}c_i^\dagger|\text{mf}\rangle \to [\tilde{\mathcal{H}}_{\text{e-ph}}, c_i^\dagger]|\text{mf}\rangle$ after noting that $c_i^\dagger\tilde{\mathcal{H}}_{\text{e-ph}}|\text{mf}\rangle \equiv 0$ if we do not allow $\tilde{\mathcal{H}}_{\text{e-ph}}$ to excite particle-hole pairs from the Fermi sea. This leads to the first EOM:

$$G_A(k,z) = \Theta(|k| - k_F)G_0(k,z)\left[1 + gF_1(k,z)\right] \qquad (10)$$

where we defined the generalized propagators:

$$F_n(k,z) = \sum_i \frac{e^{ikR_i}}{\sqrt{N}}\langle\text{mf}|c_k\hat{G}_A(z + E_{\text{mf}})c_i^\dagger B_i^{\dagger n}|\text{mf}\rangle. \qquad (11)$$

Their EOM are generated similarly. For a one-site $\mathrm{MA}^{(0)}$ variational approximation, once the first phonon is created at a site i we do not allow phonons to be created elsewhere in the system. Combining this with always freezing N_f of the fermions in the Fermi sea, leads to the recurrence equation:

$$F_n(k,z) = gg_0^A(z - n\Omega)\left[nF_{n-1}(k,z) + F_{n+1}(k,z)\right] \qquad (12)$$

where we define the real-space free propagator:

$$g_\delta^A(z) = \frac{1}{N}\sum_i \langle\text{mf}|c_{i+\delta}\hat{G}_A^0(z + E_{\text{mf}})c_i^\dagger|\text{mf}\rangle$$

$$= \frac{1}{N}\sum_{|k|>k_F} e^{-ik\delta}G_0(k,z). \qquad (13)$$

These 1D free propagators have analytic expressions. For completeness, these are presented in appendix A, where we also present additional arguments for the expected accuracy of the MA solution.

These EOM are thus formally identical to those obtained for a single polaron, and reduce to them as $x \to 0$ and $g_\delta^A(z) \to g_\delta(z) = \frac{1}{N}\sum_k e^{-ik\delta}G_0(k,z)$. The solution is then formally similar:

$$G_A(k,z) = \frac{\Theta(|k| - k_F)}{z - \epsilon(k) - gA_1(z)} \qquad (14)$$

where $A_n(z)$ are continuous fractions defined by $A_n(z) = ngg_0^A(z - n\Omega)/[1 - gg_0^A(z - n\Omega)A_{n+1}(z)]$ and $A_M(z) = 0$ for a sufficiently large $M \gg 1$. More details on the meaning of this variational approximation, as well as arguments for its expected quantitative accuracy at small x, are provided in appendix B.

The one-site $\mathrm{MA}^{(1)}$ solution and the three-site $\mathrm{MA}^{(0)}$ solutions proceed similarly but within the corresponding enlarged variational spaces. The corresponding expressions for the self-energy are like those for the single polaron counterparts, upon replacing $g_0(z) \to g_0^A(z)$ [11, 15].

3.2. Fermion removal

The fermion removal propagator is defined as:

$$G_R(k,z) = \langle \mathrm{GS}|c_k^\dagger \hat{G}_R(z - E_{\mathrm{GS}})c_k|\mathrm{GS}\rangle \tag{15}$$

where now $\hat{G}_A(z) = [z + \mathcal{H}]^{-1}$ and $z = \omega - i\eta$.

Its calculation proceeds similarly to that of the fermion addition propagator, with some care needed with details such as the changed Dyson identity $\hat{G}_R(z) = \hat{G}_R^0(z) - \hat{G}_R^0(z)\tilde{\mathcal{H}}_{\mathrm{e\text{-}ph}}\hat{G}_R(z)$ and the relevant free fermion propagator:

$$
\begin{aligned}
g_\delta^R(z) &= \frac{1}{N}\sum_i \langle \mathrm{mf}|c_{i+\delta}^\dagger \hat{G}_R^0(z - E_{\mathrm{mf}})c_i|\mathrm{mf}\rangle \\
&= \frac{1}{N}\sum_{|k|<k_F} e^{-ik\delta} G_0(k,z),
\end{aligned}
\tag{16}
$$

whose expression is provided in appendix A.

The one-site MA$^{(0)}$ solution is found to be:

$$G_R(k,z) = \frac{\Theta(k_F - |k|)}{z - \epsilon(k) - g\tilde{A}_1(z)} \tag{17}$$

where these continued fractions are defined as $\tilde{A}_n(z) = ngg_0^R(z + n\Omega)/[1 - gg_0^R(z + n\Omega)\tilde{A}_{n+1}(z)]$ and $\tilde{A}_M(z) = 0$ for a sufficiently large $M \gg 1$.

This approximation and its more accurate variants are expected to hold for $x \to 1$, i.e. for a small concentration of holes away from a full band. This is because one can use particle-hole symmetry and map the removal propagator at a concentration x onto the addition propagator at the concentration $1 - x$, upon also changing the energy $\omega \to -\omega$. This is why the accuracy considerations mirror those for the fermion addition propagator (see appendix B), and why in the following we only show results for electron addition spectra at a small x.
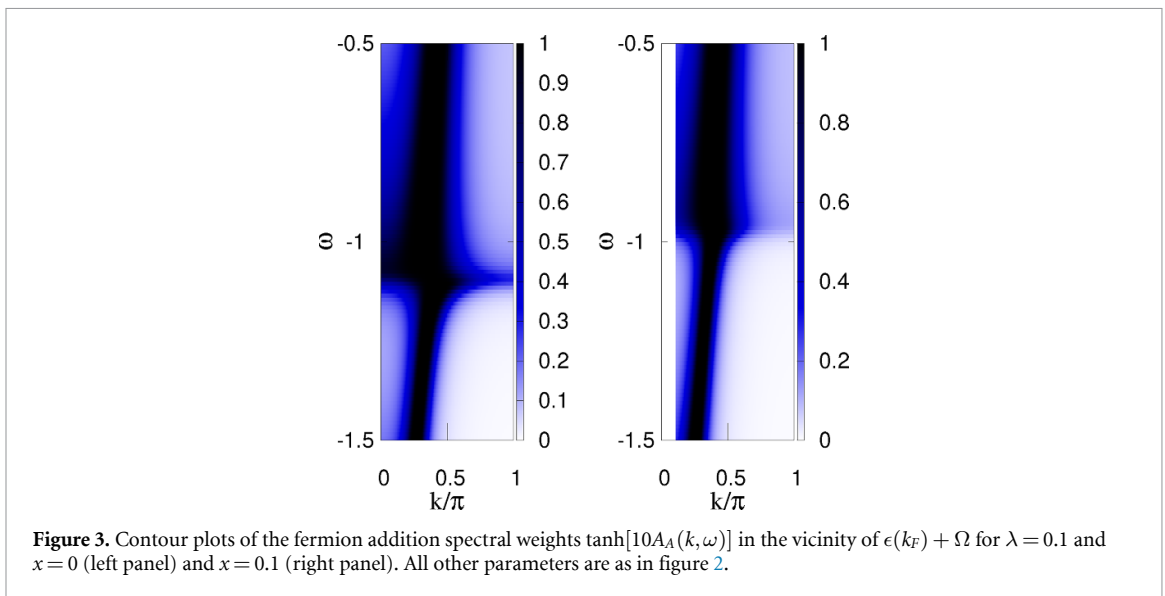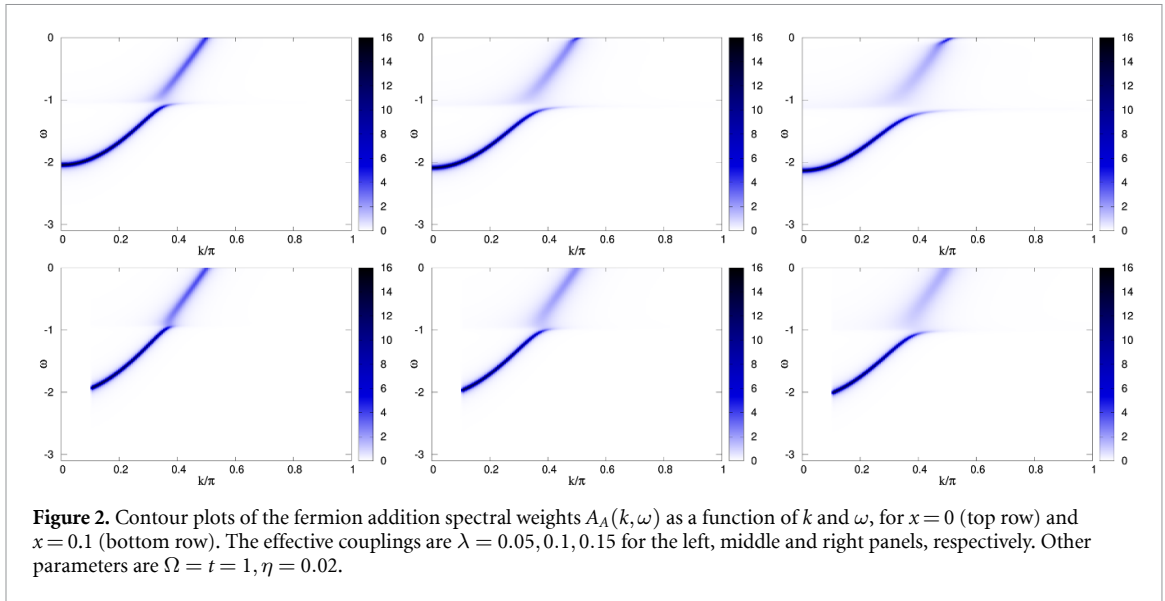
4. Results

4.1. Fermion addition spectral weights

We start by considering the spectral weights for fermion addition $A_A(k,\omega) = -\frac{1}{\pi}\mathrm{Im}G_A(k,\omega + i\eta)$ for low fermion concentrations and at weak effective couplings. Representative results obtained with MA$^{(1)}$ for $x = 0.1$ are shown in the bottom row of figure 2 for $\lambda = 0.05, 0.1, 0.15$. For comparison, the corresponding $x = 0$ single polaron results are shown in the top row.

In both cases, we see a polaron band at low energies—this is a discrete eigenstate (here broadened by η) located below a much broader polaron + phonon continuum that starts at Ω above the bottom of the polaron band. At first sight, the only difference appears to be that at finite x, the spectral weight for fermion addition vanishes for $k < k_F = \pi x$.

However, a more careful look shows a qualitative difference for the finite x results as compared to the single polaron results, related to the end point of the polaron band. It is well-known that in 1D, the single polaron band extends across the entire Brillouin zone for any $\lambda > 0$, i.e. there is a discrete polaron eigenstates below the continuum at any k. Its quasiparticle weight decreases very fast with increasing k, which is why this state is difficult to see on the scale of figure 2. To make it more apparent, in figure 3 we plot instead $\tanh[10A_A(k,\omega)]$ for $\lambda = 0.1$. The left panel shows the $x = 0$ result, and indeed reveals the entire polaron band, flattened just below the continuum but extending across the entire Brillouin zone. In contrast, the right panel shows that at finite x, the polaron state instead 'joins' the continuum, i.e. there is a $k_{\max}$ so that for $k > k_{\max}$ there is no discrete polaron state, the lowest feature being the continuum.

In other words, for low λ at any finite x we observe the famous 'kink' that is textbook knowledge in the Migdal limit. As already discussed, for these parameters the Migdal theorem does not hold, however at such small λ the lowest self-energy diagram is dominant so that $\Sigma_A(k,\omega) \approx g^2 g_0^A(z - \Omega)$. This agrees with the Migdal result for $g \to 0$. For $x = 0$, there is a divergence at $\omega = -2t$ in the density of states (DOS) of the free fermion band. The DOS is proportional to the imaginary part of $g_0(z)$, so the DOS divergence is responsible for a divergence in the real part of $g_0(z)$ at $\omega = -2t$, through the Kramers-Krönig relations; the latter divergence, in turn, explains why there is a discrete low-energy pole at the polaron energy $E_P(k)$ satisfying $E_P(k) = \epsilon_k + \Sigma_A(k, E_P(k))$ for any k. For any finite x, on the other hand, the bottom of the free fermion band (energies below $\epsilon(k_F)$) is blocked out and there is no singularity in the DOS at $\epsilon(k_F)$. As a result, the real part

Figure 2. Contour plots of the fermion addition spectral weights $A_A(k,\omega)$ as a function of k and ω, for $x=0$ (top row) and $x=0.1$ (bottom row). The effective couplings are $\lambda = 0.05, 0.1, 0.15$ for the left, middle and right panels, respectively. Other parameters are $\Omega = t = 1, \eta = 0.02$.

Figure 3. Contour plots of the fermion addition spectral weights $\tanh[10 A_A(k,\omega)]$ in the vicinity of $\epsilon(k_F) + \Omega$ for $\lambda = 0.1$ and $x = 0$ (left panel) and $x = 0.1$ (right panel). All other parameters are as in figure 2.

of $g_0^A(z)$ has only a finite peak at $\omega = \epsilon(k_F)$, therefore the equation $\omega = \epsilon_k + \Sigma_A(k,\omega)$ only has a discrete solution at $\omega = E_P(k)$ below the continuum for a restricted range $k_F \leqslant k \leqslant k_{\max}$.

Spectral weights for larger values $\lambda = 0.4, 0.7$ and 1 are shown in figure 4 for the single polaron (top row) and at $x = 0.1$ (bottom row). Here the polaron band becomes fully visible in the single polaron case, but we can also observe it extending over the full Brillouin zone for the finite x. This is a consequence of the fact that $k_{\max}$ increases with λ, as can be discerned from figure 2, and eventually reaches the Brillouin zone boundary. For $\Omega/t = 1, x = 0.1$, this happens somewhere between $\lambda = 0.15$ and $\lambda = 0.4$, i.e. still at weak couplings. Insofar as we use the existence of a full polaron band as a sign of 'polaronic physics', it is clear that in 1D and for $E_F/\Omega < 1$, this sets in fast. This is to be contrasted with the Migdal limit $E_F/\Omega \gg 1$ where 'polaronic physics' is expected to become relevant only for $\lambda \sim 1$ [8], although that analysis is in 3D. Furthermore, it is worth noting that in 3D there is a critical value of λ even for the single Holstein polaron band to extend over the whole Brillouin zone (there is no singularity at the bottom of the 3D DOS, unlike in 1D) [22]. Thus, it is not clear how much of this difference is due to dimensionality, and how much to having very different E_F/Ω ratios.

The other observations apparent from figure 4 are that at finite x, the polaron bandwidth is broader (even in the absence of the $k < k_F$ region) and the quasiparticle weight at the bottom of the band is larger than for the corresponding single polaron band (see different scales for $\lambda = 1$ contours). We can understand qualitatively the former observation as follows. As mentioned, the continuum starts at $\min_{k \geqslant k_F} E_P(k) + \Omega$. For the single polaron, the minimum is at $k = 0$ whereas at finite x it is as $k = k_F$, leaving more 'space' for a broader dispersion in the latter case.
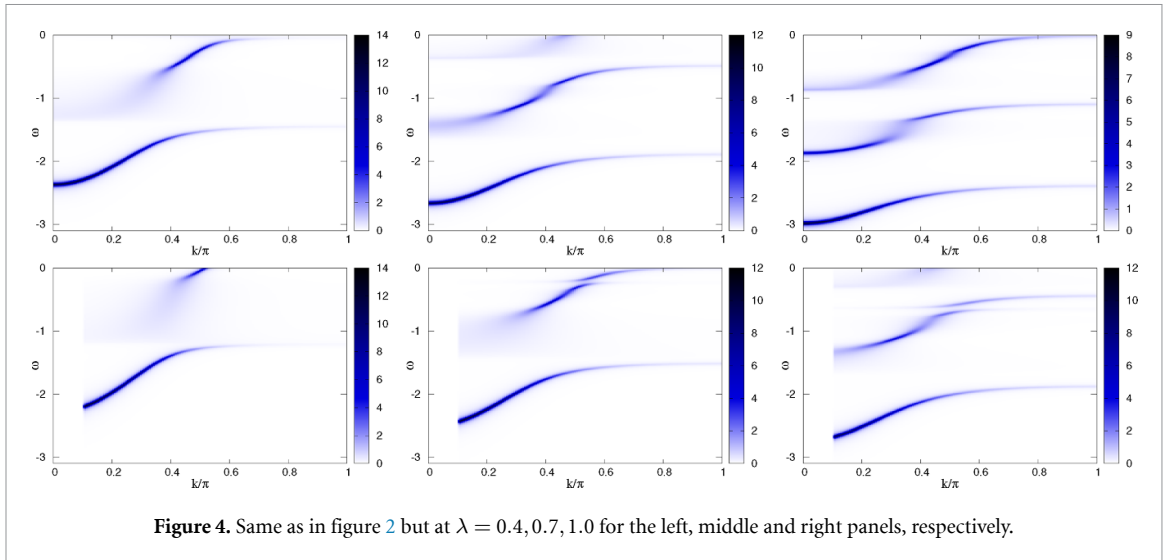
Figure 4. Same as in figure 2 but at $\lambda = 0.4, 0.7, 1.0$ for the left, middle and right panels, respectively.

The latter observation is encouraging as it bodes well for the internal self-consistency of this variational scheme even at such strong couplings. Apart from the variational assumptions about the structure of the cloud dressing the additional fermion (whose validity improves with increasing λ), we were also forced to assume that the other fermions are undressed and locked in an inert Fermi sea. Of course, in reality each fermion will create its own cloud. As a result, even if the resulting object has a well defined momentum $k < k_F$, with some probability the fermion could have any momentum k' with the cloud carrying $k - k'$. In other words, in the presence of these clouds, our assumption that the original fermions only block the states in the Fermi sea becomes suspect. The quasiparticle weight equals the probability that there is no cloud so the higher the former, the lower the latter, and the more valid is the inert Fermi sea approximation. We further discuss these issues in the last section.

4.2. Fermi speed and effective mass renormalization

A more quantitative way to characterize the quasiparticle (polaron) dispersion and its renormalization from the free fermion band is by considering the values of the polaron speed at the Fermi momentum

$$v_F^* = \left. \frac{dE_P(k)}{dk} \right|_{k=k_F} \tag{18}$$

and corresponding effective mass:

$$m^*(k_F) = \frac{k_F}{v_F^*}. \tag{19}$$

The ratio defining the effective mass is well defined for $k_F > 0$. In the limit $k_F = 0$ (for a single polaron), applying the L'Hôpital rule gives the usual link between the single polaron effective mass and the inverse of the 2nd derivative of the energy. Below, we calculate the ratios between these quantities and their bare fermion counterparts $v_F = 2t \sin k_F$ and $m(k_F) = k_F/v_F$. Given that $v_F^*/v_F = m(k_F)/m^*(k_F)$, it suffices to show only one of them. We choose to present the effective mass renormalization as it is a quantity more studied in the single polaron limit.

The derivative $dE_P(k)/dk$ can be calculated directly, by extracting the energy $E_P(k)$ of the lowest pole over a small momentum range above k_F, and fitting it to extract the slope at k_F. Because the MA self-energies are momentum-independent at the levels implemented here (for the Holstein coupling, weak momentum dependence only appears in the self-energy for MA$^{(2)}$ and higher levels), a simpler approach is to use the fact that for a local self-energy, $v_F^* = v_F / \left(1 - \frac{d\Sigma}{d\omega} |_{\omega=E_P(k_F)} \right) = v_F Z(k_F)$, where $Z(k_F)$ is the quasiparticle weight. It can be extracted by fitting $A_A(k_F, \omega)$ near $\omega \approx E_P(k_F)$ with a Lorentzian of width η, to find the corresponding weight $Z(k_F)$. We implemented both approaches and found agreement to better that 0.1% accuracy. In the following we show results obtained with the second method, i.e. we use:
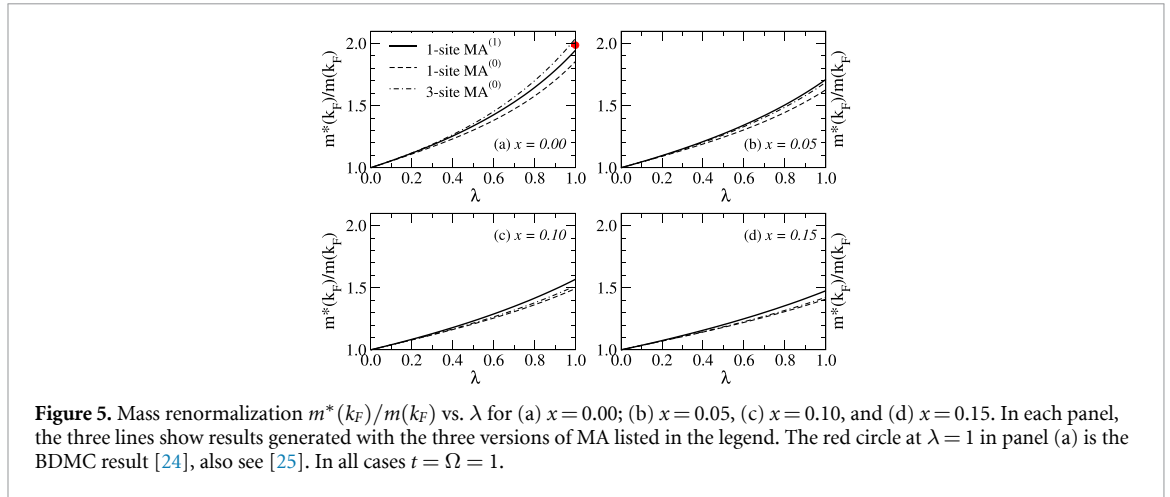
$$\frac{m^*(k_F)}{m(k_F)} = \frac{1}{Z(k_F)}. \tag{20}$$

Figure 5. Mass renormalization $m^*(k_F)/m(k_F)$ vs. λ for (a) $x = 0.00$; (b) $x = 0.05$, (c) $x = 0.10$, and (d) $x = 0.15$. In each panel, the three lines show results generated with the three versions of MA listed in the legend. The red circle at $\lambda = 1$ in panel (a) is the BDMC result [24], also see [25]. In all cases $t = \Omega = 1$.
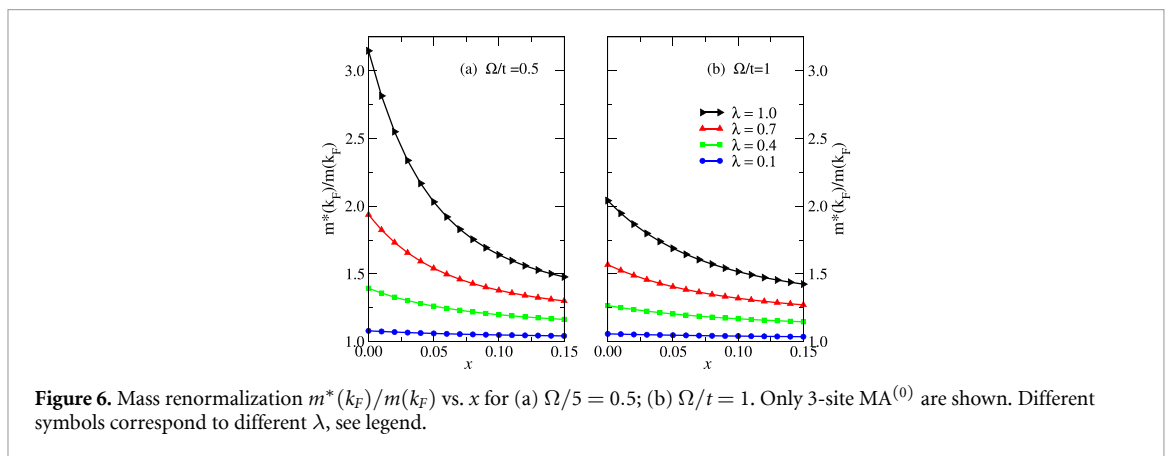

Figure 6. Mass renormalization $m^*(k_F)/m(k_F)$ vs. x for (a) $\Omega/5 = 0.5$; (b) $\Omega/t = 1$. Only 3-site MA$^{(0)}$ are shown. Different symbols correspond to different λ, see legend.

Figure 5 shows this mass renormalization ratio as a function of λ for $x = 0, 0.05, 0.1, 0.15$ and $\Omega/t = 1$, as predicted by the three versions of MA implemented here. The largest variation between the three MA predictions is observed at $\lambda = 1$ in panel (a). This is the single polaron limit where numerical results are available at these parameters, for example in [25]. The red circle shows the bold diagrammatic Monte Carlo (BDMC) result [24] (error bars are smaller than the symbol size). Both the 1-site MA$^{(1)}$ and the 3-site MA$^{(0)}$ predictions are within less than 2% of this BDMC value. The agreement between the three MA results improves with increasing x, showing that these variational results are basically converged for these fermion concentrations.

These results allow us to revisit the issue of how to define a 'proper' effective coupling at finite x. In metals with a large E_F, it is customary to define it as $\lambda_{\text{eff}}(k_F) = m^*(k_F)/m(k_F) - 1$. In the low-$\lambda$ limit we obtain a linear scaling with λ (as expected from perturbation theory), so if we followed that approach we would find $\lambda_{\text{eff}}(k_F) \propto \lambda$. In 3D and the Migdal limit, $\lambda_{\text{eff}} = 1 + \frac{g^2}{\Omega} N(E_F)$, where $N(\epsilon)$ is the free electron DOS at energy ϵ [8]. This result is obtained after assuming that $N(\epsilon) = N(E_F)$ is constant for $\epsilon \in [E_F - \Omega, E_F + \Omega]$, which is a good approximation if $E_F \gg \Omega$ but is clearly inappropriate in the limit $E_F < \Omega$ analyzed here. The analytic relationship valid in our limit can likely be found with some good will, but its usefulness is doubtful considering that nonlinear corrections set in at rather low λ values, as shown in figure 5. This is why defining an effective coupling from the mass renormalization is less useful when stronger couplings are considered, and why instead we continue to use the single polaron effective coupling λ.

These results confirm that indeed the mass renormalization is reduced (i.e. the polaron bandwidth and its quasiparticle weight are increased) as x increases. This is clearly demonstrated in figure 6, where we plot the mass renormalization ratio versus x. The MA prediction is that especially at stronger couplings, the mass renormalization reduces significantly and quickly as the concentration of carriers increases. This is in qualitative agreement with 2D results reported recently in figure 1 of [26], which shows an increasing $Z(k_F)$ with increasing E_F/Ω at small values of the latter parameter (due to a small concentration x). Those results also show an eventual upturn as the Migdal limit is being reached, but we do not expect that regime to be well described by our variational approximation.

As a final comment, we note that we could also plot the results of figure 6 against $E_F/\Omega = 2t$ $[1 - \cos(\pi x)]/\Omega$. Clearly, however, this 'composite' parameter is not sufficient to fully characterize the behavior in the small E_F/Ω limit, as evident from the fact that in the limit $x \to 0$ the results depend strongly on both λ *and* the adiabaticity ratio Ω/t. The situation might improve if a better definition for λ_{eff} was identified, nevertheless it is likely that one needs three dimensionless parameters, *e.g.* $\lambda, x, \Omega/t$ to fully characterize the parameter space at small x, far from the Migdal limit.

5. Discussions and conclusion

We proposed a variational method for studying the effects of strong electron-phonon coupling on polaron formation in spinless metals. One key assumption is to lock the original N_f fermions in an inert Fermi sea, and only allow the additional fermion/hole to create a polaron cloud. This assumption is only likely to be reasonable at small x (nearly empty band) for fermion addition, and large x (nearly full band) for fermion removal. Our specific model has particle-hole symmetry so we only showed results for the fermion addition problem at small x. We continue to focus our discussion on this case but note that it carries over to fermion removal spectral weights if $x \to 1 - x$.

The second key assumption is that the dressing cloud accompanying the quasiparticle at small x is not all that different from that in the $x = 0$ limit, i.e. from that of a single polaron. For single polarons, MA has been shown to be quite accurate at all coupling strengths λ, if Ω/t is not too small. The particular MA versions implemented here allow for a rather compact phonon cloud, of up to three sites; for the 1D single Holstein polaron the results are accurate if $\Omega/t > 0.5$ or so (in the strongly adiabatic limit, the extent of the phonon cloud increases and the variational space must be modified accordingly) [10, 15].

The validity of the second assumption can be gauged by considering the variation of the results upon increasing the cloud size. For the parameters shown here, this leads to changes of few percent, giving us some confidence that the variational space for cloud configurations is quite sufficient (appendix B provides additional insight into this).

The validity of the first approximation is harder to gauge, especially at larger λ where the quasiparticle weights become quite small, i.e. where the fermions are dressed by quite large clouds. This may limit this variational approximation to only being quantitatively accurate at small λ. However, our results predict an increase in the spectral weight with increasing x at all λ, so perhaps the method is applicable up to strong(er) couplings. It will be necessary to compare these MA predictions against numerical results in order to fully understand its range of validity.

In this context, it is important to mention that BDMC results for the 2D version of this problem have been recently published [26], so there is hope that 1D results will become available soon either from BDMC and/or from 1D specific numerical methods such as density matrix renormalization group. Taken together, we hope that they will provide a good starting point for understanding the properties of these quasiparticles in this part of the parameter space that lies far from the Migdal limit.

If the comparison with numerical results finds that this finite-x MA extension performs poorly, then it may be possible to use an iterative approach to improve it, for example by replacing the bare fermions' dispersion with the calculated polaron dispersion $E_P(k)$.

If the comparison with numerical results will validate this finite-x MA extension in some region of the parameter space, it is then straightforward to extend MA to higher dimensions and to other couplings. In particular, MA is known to become more accurate in higher dimensions for single polarons [10]. The main challenge in going to higher dimensions is to calculate the real-space propagators of the bare fermions. Here we presented 1D results because we obtained analytical expressions for the 1D bare propagators, see appendix A. In higher D, numerical calculations will be necessary to get them; we plan to investigate this issue next. Studying other couplings in 1D is straightforward and can be done immediately, but is of dubious value before the method is validated.

Another obvious future direction is to consider spinful fermions. Here, there are two qualitatively different regimes to consider. The first is when the bare electron-electron repulsion is rather weak. The main new physics expected here is the formation of bipolarons, so it would make better sense to consider propagators for the addition of a singlet or triplet of two particles (again, in the presence of an inert Fermi sea). This is a lattice generalization of the famous Cooper pair calculation [27], except it would again be in the less studied limit of small E_F/Ω and moderate-to-large λ. We note that MA has already been used successfully to study a single bipolaron [28] so this extension can be implemented straightforwardly as well. The other limit is when the bare electron-electron repulsion is large enough to prevent the binding of bipolarons in real-space. In this case, the presence of the opposite-spin carriers is irrelevant because they would neither block states for the additional carrier, nor have a tendency to share a cloud with it. As a result, in this case the approximation presented here can be used upon rescaling $x \to x/2$.

Finally, it is worth commenting on the perspectives of using this approximation to study actual materials. An interesting possibility is related to the recent proposal of spin polaron formation in hole-doped spin-polarized moiré superlattices [29]. The method proposed here can be straigthforwardly extended to study finite hole concentrations, given that carrier-magnon coupling can be treated similarly to carrier-phonon coupling, as demonstrated in the context of hole-doped cuprate layers [30, 31]. For spinful systems, the most intriguing option are $STiO_3$ and other quantum paraelectrics, which are known to have substantial electron-phonon coupling and to superconduct at surprisingly low carrier concentrations x [32–35].

To summarize, we believe that this work opens up a range of interesting possibilities for making progress in the study of this old problem.

Data availability statement

Acknowledgments

Appendix A. Real-space 1D free propagators at finite carrier concentration x

A.1. $x = 0$ expressions

For completness, we briefly review the expression in the single fermion limit:

$$g_\delta(z) = \frac{1}{N} \sum_k e^{-ik\delta} G_0(k, z). \tag{A.1}$$

For $\epsilon(k) = -2t \cos k$, it is straightforward to derive the EOMs $zg_0(z) = 1 - 2tg_1(z)$ and for any $|\delta| \geqslant 1$, $zg_\delta(z) = -2t[g_{\delta-1}(z) + g_{\delta+1}(z)]$. The latter recurrence relation has the solution $g_\delta(z) = [\zeta(z)]^{|\delta|} g_0(z)$ where

$$\zeta(z) = -\frac{z}{2t} + \sqrt{\frac{z}{2t} - 1}\sqrt{\frac{z}{2t} + 1} \tag{A.2}$$

is the characteristic root that satisfies $|\zeta(z)| < 1$. The final result is:

$$g_\delta(z) = \frac{[\zeta(z)]^{|\delta|}}{\sqrt{z - 2t}\sqrt{z + 2t}}. \tag{A.3}$$

A.2. Finite x expressions

Here we present our derivation for the fermion addition propagator:

$$g_\delta^A(z) = \frac{1}{N} \sum_{|k|>k_F} e^{-ik\delta} G_0(k, z) \tag{A.4}$$

from which one can then obtain the fermion removal propagator using $g_\delta^R(z) = g_\delta(z) - g_\delta^A(z)$.

The EOMs now become $zg_0^A(z) = 1 - x - 2tg_1^A(z)$ and for any $|\delta| \geqslant 1$, $zg_\delta^A(z) = \gamma_\delta - 2t[g_{\delta-1}(z) + g_{\delta+1}(z)]$ where $\gamma_\delta = \frac{1}{N} \sum_{|k|>k_F} e^{-ik\delta} = -\sin(\pi x \delta)/(\pi \delta)$ for any $\delta \neq 0$.

For any $\delta \geqslant 1$, the general solution of the latter recurrence relation now takes the form $g_\delta^A(z) = \alpha_\delta(z) + \zeta(z)g_{\delta-1}^A(z)$, where $\zeta(z)$ was defined in equation (A.2) and $\alpha_\delta(z) = -\frac{\zeta(z)}{t}\gamma_\delta + \zeta(z)\alpha_{\delta+1}$ has the explicit solution

$$\alpha_\delta(z) = \frac{[\zeta(z)]^{1-\delta}}{t} \sum_{k=\delta}^{\infty} [\zeta(z)]^k \frac{\sin(\pi x k)}{\pi k}.$$

For $n = 1$, the sum can be carried out and we find:

$$\alpha_1(z) = \frac{i}{2\pi t} \ln \frac{1 - \zeta(z)e^{i\pi x}}{1 - \zeta(z)e^{-i\pi x}}. \tag{A.5}$$

Using this general solution in the $\delta = 0$ EOM then results in:

$$g_0^A(z) = \frac{1 - x - 2t\alpha_1(z)}{\sqrt{z - 2t}\sqrt{z + 2t}}. \tag{A.6}$$

For the one-site MA$^{(0)}$ solution this is all that is needed, but for the more accurate solutions one needs the propagators for a finite distance as well. After some re-organizing, the general solution for any $n \geqslant 1$ is found to be:

$$g_n(z) = g_{-n}(z) = \beta_n(z) + [\zeta(z)]^n g_0^A(z) \tag{A.7}$$

where $\beta_1(z) = \alpha_1(z)$ and for any $n \geqslant 1$:

$$\beta_n(z) = \alpha_1(z) \frac{[\zeta(z)]^{1-n} - [\zeta(z)]^{1+n}}{1 - [\zeta(z)]^2}$$
$$- \sum_{k=1}^{n-1} \frac{\sin(\pi x k)}{t\pi k} \frac{[\zeta(z)]^{1+k-n} - [\zeta(z)]^{1-k+n}}{1 - [\zeta(z)]^2}. \tag{A.8}$$

This is the most compact form we found for the general solution.

Appendix B. On the accuracy of MA

One way to better understand the meaning of the variational MA approximation and its accuracy is in diagrammatic terms (of course, a diagrammatic expansion is meaningful only at weak coupling). This is discussed in [10] for the single polaron case, and it can be directly extended to finite x. For completeness, we briefly review it here for the one-site MA$^{(0)}$ version, but we re-formulate it in real space, as opposed to the original discussion of [10] which is in momentum space (at the time, we did not yet understand the variational meaning of MA [23]).

We focus on the fermion addition propagator; the discussion for the fermion removal propagator is similar. To obtain the diagrammatic expansion, the Dyson identity is expanded iteratively:

$$\hat{G}_A(z) = \hat{G}_A^0(z) + \hat{G}_A^0(z)\tilde{H}_{\text{e-ph}}\hat{G}_A^0(z)$$
$$+ \hat{G}_A^0(z)\tilde{H}_{\text{e-ph}}\hat{G}_A^0(z)\tilde{H}_{\text{e-ph}}\hat{G}_A^0(z) + \dots$$

Calculating these terms within the same variational approximation used to derive the EOMs allows us to identify self-energy diagrams of various orders in the usual way. The one-site MA$^{(0)}$ approach makes approximations starting at the 2nd order self-energy diagrams (the first order diagram is included exactly). There is one crossed and one un-crossed 2nd order diagram. If we impose *only* the constraint of keeping the Fermi sea unchanged, their expressions in real space are:

$$\Sigma_{2,nc}(k,z) = g^4 \sum_n g_n^A(z - \Omega) g_0^A(z - 2\Omega) g_n^A(z - \Omega)$$

$$\Sigma_{2,c}(k,z) = g^4 \sum_n e^{-ikn} g_n^A(z - \Omega) g_n^A(z - 2\Omega) g_n^A(z - \Omega).$$

In both cases the fermion first emits a phonon at some (any) site, let us call it i. With the energy lowered by Ω, it then moves to any other site $j = i + n$ where it emits a second phonon, lowering its energy by 2Ω. In the non-crossed diagram, the fermion remains at site j to absorb the 2nd phonon before returning to site i (a distance n away) to absorb the first phonon. In the crossed diagram the phonons are absorbed in inverse order, so the fermion travels back and forth by n sites before each process.

The lowest eigenstate of the quasiparticle will be at or below the lowest possible energy of the added fermion, namely $\epsilon_F = \epsilon(k_F)$. As the equation above shows, in the self-energy, the free fermion propagator is called with even lower energies because it emits one or more phonons. Then, let us consider how $[g_n^A(\epsilon_F - \Omega)]^2$ varies with n and x. The results are shown in figure B1. For $x = 0$, the propagator decays exponentially with n. For small x the decay is slower and one can see the first of the Friedel oscillations at
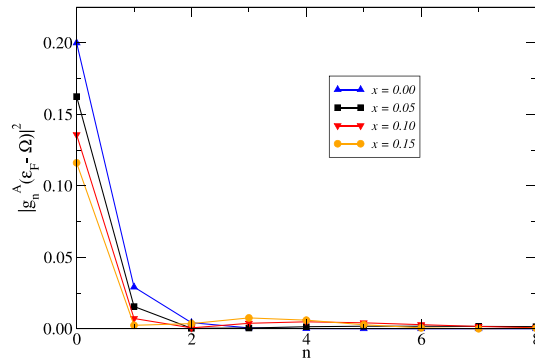
Figure B1. $|g_n^A(\epsilon_F - \Omega)|^2$ vs n for various values of the fermion concentration x. Here, $\epsilon_F = \epsilon(k_F) = -2t\cos(\pi x)$. While at $x = 0$ there is an exponential decay with distance n, at finite x one can see the appearance of Friedel oscillations. Nevertheless, for small x the results are quantitatively rather similar. See text for more details.

larger n, but their amplitude is still small. The decay is even faster as the energy argument becomes $z - m\Omega$, $m > 1$ in higher order diagrams.

The one-site MA$^{(0)}$ is equivalent to ignoring all contributions with $n \neq 0$ from all diagrams (not just 2nd order ones). As discussed in the main text, for $\Omega = t = 1$ this is an excellent approximation at any λ, and the results in figure B1 suggest that the accuracy should remain similarly good for $x < 0.15$ or so.

It is important to note that the $n = 0$ contribution of the crossed and non-crossed diagrams are precisely equal. This is a consequence of the fact that the phonons are emitted at the same site and thus indistinguishable, so the order of re-absorbing them cannot possibly matter. Thus, the Migdal argument is clearly invalid in this small x limit, where we can ignore the $n \neq 0$ contributions.

For completeness, we mention that the increase of the variational space selectively adds more contributions to the self-energy diagrams. For example, the one-site MA$^{(1)}$ sets $g_n^A(z - m\Omega) \to \delta_{n,0} g_0^A$ $(z - m\Omega)$ only for $m \geqslant 2$, so $\Sigma_{2,nc}(k, z)$ is calculated without further approximations. The three-site MA$^{(1)}$ keeps all contributions with $n \leqslant 2$ exactly and only ignores free propagators that move the free fermion by $n \geqslant 3$ sites, *etc.* Comparing their results thus gives us an estimate for how important are these ignored propagators, and therefore of the accuracy of the MA prediction.

ORCID iD

M Berciu ⓘ https://orcid.org/0000-0002-6736-1893

References

[1] Su W P, Schrieffer J R and Heeger A J 1979 *Phys. Rev. Lett.* **42** 1698
[2] Heeger A J, Kivelson S, Schrieffer J R and Su W P 1988 *Rev. Mod. Phys.* **60** 781
[3] Grüner G 1988 *Rev. Mod. Phys.* **60** 1129
[4] Bardeen J, Cooper L N and Schrieffer J R 1957 *Phys. Rev.* **108** 1175
[5] Nocera A, Adrian J S Feiguin E and Berciu M 2021 *Phys. Rev. B* **104** L201109
[6] Zhang C, Sous J, Reichman D R, Berciu M, Millis A J, Prokof'ev N V and Svistunov B V arXiv: 2203.07380
[7] Migdal A B 1958 *J. Exp. Theor. Phys.* **7** 996
[8] Engelsberg S and Schrieffer J 1963 *Phys. Rev* **131** 993
[9] Berciu M 2006 *Phys. Rev. Lett.* **97** 036402
[10] Goodvin G L, Berciu M and Sawatzky G A 2006 *Phys. Rev. B* **74** 245104
[11] Berciu M and Goodvin G L 2007 *Phys. Rev. B* **76** 165109
[12] Goodvin G L and Berciu M 2008 *Phys. Rev. B* **78** 235120
[13] Marchand D J J, De Filippis G, Cataudella V, Berciu M, Nagaosa N, Prokof'ev N V, Mishchenko A S and Stamp P C E 2010 *Phys. Rev. Lett.* **105** 266605
[14] Herrera F, Madison K W, Krems R V and Berciu M 2013 *Phys. Rev. Lett.* **110** 223002
[15] Marchand D J J, Stamp P C E and Berciu M 2017 *Phys. Rev. B* **95** 035117
[16] Berciu M, Mishchenko A S and Nagaosa N 2010 *Europhys. Lett.* **89** 37007
[17] Ebrahimnejad H and Berciu M 2012 *Phys. Rev. B* **85** 165117
    Ebrahimnejad H and Berciu M 2012 *Phys. Rev. B* **86** 205109
[18] Covaci L and Berciu M 2007 *Europhys. Lett.* **80** 67001
[19] Adolphs C P J and Berciu M 2013 *Europhys. Lett.* **102** 47003
[20] Holstein T 1959 *Ann. Phys., NY* **8** 325
[21] Holstein T 1959 *Ann. Phys., NY* **8** 343
[22] Goodvin G L and Berciu M 2010 *Europhys. Lett.* **92** 37006
[23] Barisić O S 2007 *Phys. Rev. Lett.* **98** 209701
    Berciu M 2007 *Phys. Rev. Lett.* 209702

[24] Mishchenko A 2022 private communication
[25] Hague J P, Kornilovitch P E, Alexandrov A S and Samson J H 2006 *Phys. Rev.* B **73** 054303
[26] Mishchenko A S, Tupitsyn I S, Nagaosa N and Prokof'ev N 2021 *Sci. Rep.* **11** 9699
[27] Cooper L N 1956 *Phys. Rev.* B **104** 1189
[28] Sous J, Chakraborty M, Krems R V and Berciu M 2018 *Phys. Rev. Lett.* **121** 247001
     Sous J, Chakraborty M, Adolphs C P J, Krems R V and Berciu M 2017 *Sci. Rep.* **7** 1169
[29] Davydova M, Zhang Y and Liang F arXiv: 2206.01221
[30] Ebrahimnejad H, Sawatzky G A and Berciu M 2014 *Nat. Phys.* **10** 951
[31] Ebrahimnejad H, Sawatzky G A and Berciu M 2016 *J. Phys.: Condens. Matter* **28** 105603
[32] Schooley J F, Hosler W R and Cohen M L 1964 *Phys. Rev. Lett.* **12** 474
[33] van der Marel D, Barantani F and Rischau C W 2019 *Phys. Rev. Res.* **1** 013003
[34] Gastiasoro M N, Ruhman J and Fernandes R M 2020 *Ann. Phys., NY* **417** 168107
[35] Yue Y, Hwang H Y, Raghu S and Chung S B 2022 *npj Quantum Mater.* **7** 36