Collective charge excitations of strongly correlated electrons, vertex corrections, and gauge invariance

Hartmut Hafermann, Erik G. C. P. van Loon, Mikhail I. Katsnelson, Alexander I. Lichtenstein, Olivier Parcollet

Collective charge excitations of strongly correlated electrons, vertex corrections, and gauge invariance

Hartmut Hafermann,[1] Erik G. C. P. van Loon,[2] Mikhail I. Katsnelson,[2] Alexander I. Lichtenstein,[3] and Olivier Parcollet[1]
[1]*Institut de Physique Théorique (IPhT), CEA, CNRS, 91191 Gif-sur-Yvette, France*
[2]*Radboud University Nijmegen, Institute for Molecules and Materials, NL-6525 AJ Nijmegen, The Netherlands*
[3]*I. Institut für Theoretische Physik, Universität Hamburg, Jungiusstraße 9, D-20355 Hamburg, Germany*
(Received 30 June 2014; revised manuscript received 3 October 2014; published 1 December 2014)

We consider the collective, long-wavelength charge excitations in correlated media in presence of short- and long-range forces. As an example for the case of a short-range interaction, we examine the two-dimensional Hubbard model within dynamical mean-field theory (DMFT). It is shown explicitly that the DMFT susceptibility including vertex corrections respects the Ward identity and yields a manifestly gauge-invariant response in finite dimensions. For computing the susceptibility, we use a different expression and establish its formal equivalence to the standard DMFT formula. It allows for a more stable analytical continuation. We find a zero-sound mode expected for short-range forces. The relation between the vertex corrections, gauge invariance, and the appearance of the collective modes is discussed. Long-range forces are treated within extended dynamical mean-field theory. In order to obtain a gauge-invariant response, it is necessary to additionally incorporate some *nonlocal* vertex corrections into the polarization. In doing so, we obtain plasmons in the three-dimensional Hubbard model. The plasma frequency is determined by the (single-particle) density distribution as a consequence of gauge invariance. We compare this result with the plasma frequency extracted from the analytical continuation of the susceptibility. It is in good agreement with the prediction from the gauge-invariance condition.

DOI: 10.1103/PhysRevB.90.235105 PACS number(s): 71.45.Gm, 71.27.+a, 71.10.Fd

I. INTRODUCTION

In the last two decades, dynamical mean-field theory (DMFT) [1] and its various cluster [2] and diagrammatic [3–7] extensions have emerged as promising and useful tools to solve several aspects of strongly correlated fermion problems. For example, DMFT has shed new light on the Mott transition problem [1] and cluster extensions of DMFT have successfully described some aspects of high-temperature superconductors (see Ref. [8] for a recent example). Furthermore, DMFT is now routinely used in combination with density functional theory to provide an *ab initio* electronic-structure method for strongly correlated systems [9]. For long-range interactions, the extended DMFT (EDMFT) [10–13] as well as more refined GW+DMFT approaches [14–17] have been developed.

Two-particle quantities and response functions can also be computed within the DMFT theoretical framework [1]. For example, systematic computations of antiferromagnetic or superconducting susceptibilities from high to low temperatures have been used to locate continuous phase transitions in cluster extensions of DMFT (see, e.g., Refs. [18–20]). More recently, interest in two-particle vertex functions is increasing [21–23]. A new generation of approaches [3,4,6,7] has emerged which uses certain two-particle functions in the self-consistency condition itself. From a technical point of view, the task of computing these two-particle quantities is significantly more challenging than the computation of the single-particle quantities used in the DMFT self-consistency loop, because of the need to include vertex corrections. However, due to the advent of continuous-time quantum Monte Carlo solvers [24–26], they can now be computed reliably and up to high precision [27].

In this paper, we study in particular the long-wavelength collective charge excitations in the Hubbard and the extended Hubbard models in presence of short- and long-range interactions, respectively. For short-range interactions, we use a regular DMFT scheme and obtain a zero sound mode in the metallic regime which persists up to the Mott transition. For long-range interactions, we use a simplified version of the dual boson method and obtain a plasmon mode. In the latter case, we discuss the failure of EDMFT to properly describe the low-energy excitations, even at a qualitative level. In both the short- and the long-range cases, we obtain a low-energy analysis similar to the standard textbook weak-coupling random phase approximation (RPA) analysis.

From a technical point of view, in order to compute the two-particle response, we employ a formula inspired from recent work on the dual boson approach [6]. While we prove it to be mathematically equivalent to the standard computation procedure of computing the DMFT susceptibility [1], it turns out to yield a much better numerical accuracy in practice.

Moreover, we show that a proper and complete treatment of the (nonlocal) vertex corrections in the DMFT framework is essential for the correctness of the result at low energy. Simplified approximations (like a simple bubble approximation) lead to qualitatively wrong results. We trace the origin of this difficulty in constructing simple approximations to the role of gauge invariance and the associated Ward identities. It has been known since the 1960s that gauge invariance is closely related to the collective modes [28] and the criteria for obtaining conserving approximations that respect gauge invariance have been formulated at the time [29,30]. For weak coupling, these requirements are fulfilled within the RPA. However, the description of correlated systems requires a frequency-dependent self-energy. Designing gauge-invariant approximations is much less straightforward in this case because dynamical vertex corrections are required. We check explicitly that the DMFT susceptibility, which includes these corrections, fulfills the Ward identity in finite dimensions. While our methodology can straightforwardly be generalized

to treat the magnetic (spin) excitations and we expect vertex corrections to be important in general, we focus here on the charge excitations.

The paper is organized as follows: We first introduce the model in Sec. II. In Sec. III, we consider the case of short-range forces on the level of DMFT. We first recall DMFT and the standard calculation of susceptibilities and then introduce the new formula for the response function that follows from the dual boson approach. Results for the charge response are discussed in detail in Sec. III D and are compared to the RPA. We then discuss gauge invariance and show explicitly that it is fulfilled within DMFT. In Sec. IV, we address the case of a long-range interaction. We show that EDMFT does not provide a valid description of plasmons. By including nonlocal vertex corrections into the polarization within the dual boson approximation, we demonstrate that the polarization obtains the proper momentum dependence required by gauge invariance. The energy of the appearing collective mode is compared with the plasma frequency and thereby identified as a plasmon mode. A detailed derivation of the employed relations and a proof of the equivalence of the dual boson formula and the DMFT susceptibility are provided in the Appendix.

II. MODEL

In the following, we consider the extended Hubbard model in finite dimensions. In particular, we focus on the two-dimensional square and three-dimensional cubic lattices. The model is described by the Hamiltonian

$$H = -\tilde{t}\sum_{\mathbf{r}\boldsymbol{\delta}\sigma}(c^{\dagger}_{\mathbf{r}\sigma}c_{\mathbf{r}-\boldsymbol{\delta}\sigma} + c^{\dagger}_{\mathbf{r}-\boldsymbol{\delta}\sigma}c_{\mathbf{r}\sigma}) + U\sum_{\mathbf{r}} n_{\mathbf{r}\uparrow}n_{\mathbf{r}\downarrow} + \frac{1}{2}\sum_{\mathbf{r}\mathbf{r}'} V(\mathbf{r}-\mathbf{r}')n_{\mathbf{r}}n_{\mathbf{r}'}. \quad (1)$$

Here, $\mathbf{r}$ denote the *discrete* positions of the lattice sites and the sum over $\boldsymbol{\delta}$ implies a sum over the displacement vectors $\boldsymbol{\delta} = a(1,0,0), a(0,1,0)$ in two dimensions and additionally $a(0,0,1)$ in the three-dimensional case, respectively. For simplicity, we restrict ourselves to nearest-neighbor hopping $\tilde{t}$ only. The tilde is used to distinguish from the symbol t which is used for time. The lattice spacing a is set to unity in the following. We further denote spin by $\sigma = \uparrow,\downarrow$ and $n = n_{\uparrow} + n_{\downarrow}$. In the above, we have written the local Hubbard interaction with Coulomb repulsion U explicitly. The last term contains the nonlocal part of the interaction, which may be long ranged. Its Fourier transform will be denoted $V(\mathbf{q})$. For the Hubbard model, $V(\mathbf{q}) = 0$. The energy unit is chosen such that $4\tilde{t} = 1$ in both two and three dimensions and all results are obtained at temperature $T = 0.02$.

III. SHORT-RANGE FORCES

In order to address the collective excitations in presence of short-range forces and strong correlations, we consider the two-dimensional Hubbard model, which can be treated within dynamical mean-field theory (DMFT). First, we briefly recall the DMFT procedure and the calculation of the susceptibilities in DMFT, as they can be found in the review of Ref. [1].

A. Recollection of DMFT

In DMFT, the lattice problem (1) with $V \equiv 0$ is mapped onto a local quantum impurity problem subject to a self-consistency condition. The lattice Green's function has the form

$$G_{\nu}(\mathbf{k}) = \frac{1}{i\nu + \mu - \varepsilon_{\mathbf{k}} - \Sigma_{\nu}}, \quad (2)$$

where $\varepsilon_{\mathbf{k}}$ is the Fourier transform of the hopping, Σ_{ν} is the local but frequency-dependent electronic self-energy, and ν stands for the discrete Matsubara frequencies $\nu_n = (2n+1)\pi T$ with T denoting temperature. Here and in the following, it is convenient to write frequency labels as subscripts to obtain a more condensed notation. We further consider the paramagnetic case and spin labels are omitted. In DMFT, the self-energy is a functional of the local Green's function only and has to be determined self-consistently. In practice, it is obtained from the solution of an Anderson impurity model, which, starting from an initial guess, is solved repeatedly until the following self-consistency condition is fulfilled:

$$g_{\nu} = \frac{1}{N}\sum_{\mathbf{k}} G_{\nu}(\mathbf{k}). \quad (3)$$

It relates the local part of the lattice Green's function to the impurity Green's function denoted g_{ν} and implicitly determines the self-energy.

B. DMFT susceptibility

Response functions may be computed once a self-consistent solution to the DMFT equations has been found. Here and in the remainder of the paper, we will focus on the charge response. The charge susceptibility is given by the connected part of the density-density correlation function[1]

$$\chi_{\omega}(\mathbf{q}) := -\langle n_{\omega}(\mathbf{q})n_{\omega}(\mathbf{q})\rangle_{\text{con.}}, \quad (4)$$

where $n = \sum_{\sigma} n_{\sigma}$ is the operator of the total density. The charge susceptibility is expressed in terms of the generalized susceptibility $\chi_{\nu\nu'\omega}(\mathbf{q})$ as

$$\chi_{\omega}(\mathbf{q}) = 2T^2\sum_{\nu\nu'}\chi_{\nu\nu'\omega}(\mathbf{q}), \quad (5)$$

where the factor of 2 stems from the spin degeneracy. The generalized susceptibility in turn is the solution to an integral equation which involves an irreducible vertex. Defining

$$\chi^{0}_{\nu\omega}(\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}} G_{\nu+\omega}(\mathbf{k}+\mathbf{q})G_{\nu}(\mathbf{k}), \quad (6)$$

this equation reads as

$$\chi_{\nu\nu'\omega}(\mathbf{q}) = \frac{1}{T}\chi^{0}_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'} - \chi^{0}_{\nu\omega}(\mathbf{q})\,T\sum_{\nu''}\Gamma^{\text{irr}}_{\nu\nu''\omega}\chi_{\nu''\nu'\omega}(\mathbf{q}). \quad (7)$$

It is depicted graphically in Fig. 1(a). In DMFT, the *irreducible* vertex Γ^{irr} is given by the irreducible vertex of the impurity γ^{irr}, i.e., $\Gamma^{\text{irr}} \equiv \gamma^{\text{irr}}$, and hence local. In practice, it is extracted from

[1]We define the susceptibility with a minus sign relative to the convention used in Ref. [1].

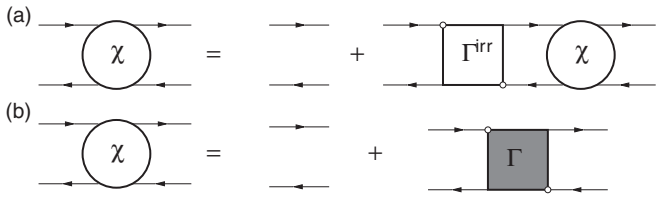

FIG. 1. Diagrammatic representation of (a) the Bethe-Salpeter equation for the generalized susceptibility χ and (b) the relation of the generalized susceptibility and the reducible vertex function Γ. Lines are fully dressed propagators.

the impurity model on the final DMFT iteration by inverting the local Bethe-Salpeter equation

$$\left[\gamma_\omega^{-1}\right]_{\nu\nu'} = \left[\gamma_\omega^{\mathrm{irr}\,-1}\right]_{\nu\nu'} + T\chi^0_{\nu\omega}\delta_{\nu\nu'}. \tag{8}$$

Here, $\chi^0_{\nu\omega} = g_{\nu+\omega}g_\nu$ and the reducible impurity vertex γ is defined through

$$\gamma^{\sigma\sigma'}_{\nu\nu'\omega} := \frac{\langle c_{\nu\sigma}c^*_{\nu+\omega,\sigma}c_{\nu'+\omega,\sigma'}c^*_{\nu'\sigma'}\rangle - \chi^{0\sigma\sigma'}_{\nu\nu'\omega}}{g_{\nu\sigma}g_{\nu+\omega,\sigma}g_{\nu'+\omega\sigma'}g_{\nu'\sigma'}} \tag{9}$$

and

$$\chi^{0\sigma\sigma'}_{\nu\nu'\omega} := (g_{\nu\sigma}g_{\nu'\sigma'}\delta_\omega - g_{\nu+\omega,\sigma}g_{\nu\sigma}\delta_{\nu\nu'}\delta_{\sigma\sigma'})/T \tag{10}$$

(see also Appendix H).

In an equivalent formulation, the susceptibility is expressed in terms of the (reducible) vertex function of the lattice as follows:

$$\chi_\omega(\mathbf{q}) = 2T\sum_\nu \chi^0_{\nu\omega}(\mathbf{q}) - 2T^2\sum_{\nu\nu'}\chi^0_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\chi^0_{\nu'\omega}(\mathbf{q}). \tag{11}$$

This relation is graphically depicted in Fig. 2(a). The vertex function is obtained as the solution to the integral equation

$$\Gamma_{\nu\nu'\omega}(\mathbf{q}) = \Gamma^{\mathrm{irr}}_{\nu\nu'\omega} - T\sum_{\nu''}\Gamma^{\mathrm{irr}}_{\nu\nu''\omega}\,\chi^0_{\nu''\omega}(\mathbf{q})\Gamma_{\nu''\nu'\omega}(\mathbf{q}), \tag{12}$$

which is called the Bethe-Salpeter equation [BSE, see Fig. 2(b)]. Diagrammatically, the BSE corresponds to the infinite sum of ladderlike diagrams to the vertex function with DMFT Green's functions as rails and the irreducible vertex appearing as rungs of the ladders. This can be seen by iterating it. Its physical content is the repeated particle-hole scattering processes which give rise to the collective excitations of the system. The BSE has the formal solution

$$\left[\Gamma_\omega^{-1}(\mathbf{q})\right]_{\nu\nu'} = \left[\Gamma_\omega^{\mathrm{irr}\,-1}\right]_{\nu\nu'} + T\chi^0_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'}. \tag{13}$$

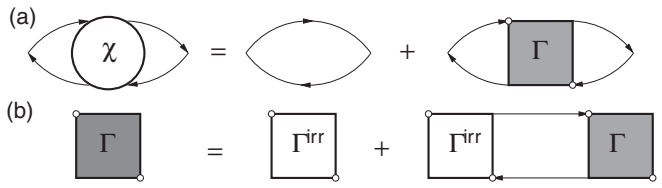

FIG. 2. Diagrammatical representation of (a) the susceptibility and (b) the Bethe-Salpeter equation for the vertex function. Lines are fully dressed propagators.

The lattice vertex depends on the transferred momentum $\mathbf{q}$ only due to the locality of the irreducible vertex.

The equivalence between the two approaches is readily established using the representation of the generalized susceptibility in terms of the vertex function

$$\chi_{\nu\nu'\omega}(\mathbf{q}) = \frac{1}{T}\chi^0_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'} - \chi^0_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\chi^0_{\nu'\omega}(\mathbf{q}) \tag{14}$$

as shown in Fig. 1(b). Inserting (14) into the right-hand side of Eq. (7) and using the Bethe-Salpeter equation for the vertex function (12) again recovers (14).

C. Alternative expression for the susceptibility

In the following, we are primarily interested in the susceptibility on the real frequency axis. This requires an analytical continuation from Matsubara frequencies. The details of the analytical continuation are summarized in Appendix A. Here, we discuss the calculation of the susceptibility on Matsubara frequencies. Instead of a straightforward implementation of the equations discussed in the previous section, we employ the following approach, which leads to better results and less artifacts. It is a reformulation of the above equations and was first derived in the context of the dual boson approach [6]. In Appendix H, we show that this formulation is exactly equivalent to the DMFT susceptibility [Eqs. (11) and (12)].

In the alternative formulation, the susceptibility is separated into a local impurity and a lattice contribution:

$$\chi_\omega(\mathbf{q}) = \chi_\omega + 2T^2\sum_{\nu\nu'}\chi_\omega\lambda_{\nu\omega}\tilde{\chi}_{\nu\nu'\omega}(\mathbf{q})\lambda_{\nu'+\omega,-\omega}\chi_\omega. \tag{15}$$

Here, $\chi_\omega := -\langle n_\omega n_\omega\rangle_{\mathrm{con.}}$ denotes the impurity susceptibility which includes local vertex corrections. The lattice vertex Γ includes nonlocal vertex corrections. The expression further involves the three-leg vertex of the impurity model λ (see Appendix H):

$$\lambda^\sigma_{\nu\omega} := \frac{-\langle c_{\nu\sigma}c^*_{\nu+\omega,\sigma}n_\omega\rangle - g_{\nu\sigma}\,\langle n\rangle\,\delta_\omega/T}{g_{\nu\sigma}g_{\nu+\omega,\sigma}\,\chi_\omega} \tag{16}$$

and we have defined

$$\tilde{\chi}_{\nu\nu'\omega}(\mathbf{q}) = \frac{1}{T}\tilde{\chi}^0_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'} - \tilde{\chi}^0_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\tilde{\chi}^0_{\nu'\omega}(\mathbf{q}). \tag{17}$$

In the above, $\tilde{\chi}_{\nu\omega}(\mathbf{q})$ denotes the *nonlocal part* of the bubble:

$$\tilde{\chi}^0_{\nu\omega}(\mathbf{q}) = \chi^0_{\nu\omega}(\mathbf{q}) - \chi^0_{\nu\omega}, \tag{18}$$

where in turn the impurity bubble is defined as $\chi^0_{\nu\omega} := g_\nu g_{\nu+\omega}$. We note that the lattice contribution in Eq. (15) contains a local part, which includes contributions from long-range collective excitations. When summed over $\mathbf{q}$, the first term in Eq. (17) vanishes,[2] while the second in general does not. In DMFT, the local part of the lattice susceptibility and the impurity susceptibility differ.

The lattice vertex Γ is the same as in the DMFT susceptibility [Eq. (11)] and can be obtained from the irreducible vertex

[2]The local part of the dual bubble [Eq. (18)] vanishes, which can be seen by summing (6) over $\mathbf{q}$ and replacing $(1/N)\sum_{\mathbf{k}} G_\nu(\mathbf{k})$ by g_ν according to (3).

through the Bethe-Salpeter equation (12). It is convenient to combine the latter with the impurity BSE for the irreducible vertex. Using the the nonlocal part of the bubble, the resulting BSE reads as

$$\left[\Gamma_\omega^{-1}(\mathbf{q})\right]_{\nu\nu'} = \left[\gamma_\omega^{-1}\right]_{\nu\nu'} + T\,\tilde{\chi}^0_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'}, \tag{19}$$

where γ is the *reducible* impurity vertex. There is hence no need to explicitly compute the irreducible impurity vertex γ^{irr}. This is similar to the dual fermion approach [31]. Such a reformulation is important to avoid unphysical singularities in the low-frequency behavior of the irreducible two-particle vertex [32], which occur in the proximity of the metal-insulator transition [33].

D. Results

Let us now turn to the results for the charge susceptibility in the two-dimensional Hubbard model. To set the stage for the discussion, we examine the noninteracting case first. In the upper panel of Fig. 3 we plot the (negative) imaginary part of the noninteracting susceptibility

$$\chi^0_\omega(\mathbf{q}) = \frac{T}{N}\sum_{\mathbf{k}\nu} G^0_{\nu+\omega}(\mathbf{k}+\mathbf{q})G^0_\nu(\mathbf{k}) \tag{20}$$

on real frequencies. Its features are best understood in terms of the noninteracting dispersion $\epsilon_{\mathbf{k}}$, which we plot in the panel below. The maximum energy up to which one can see significant spectral weight is found at the M point [$\mathbf{q} = (\pi,\pi)$]. This wave vector connects maximum and minimum in the dispersion and its energy is correspondingly given by the bandwidth $W = 2$. One can also see a structure of high intensity at very low energy and in the vicinity of the M point, which is due to the nesting of the Fermi surface. The maximal intensity at the M point is found here. The strongest overall response occurs at the X point. Its dominant contribution in the convolution stems from those k points for which the wave vector X connects two extremal points (Γ-X and X-M,

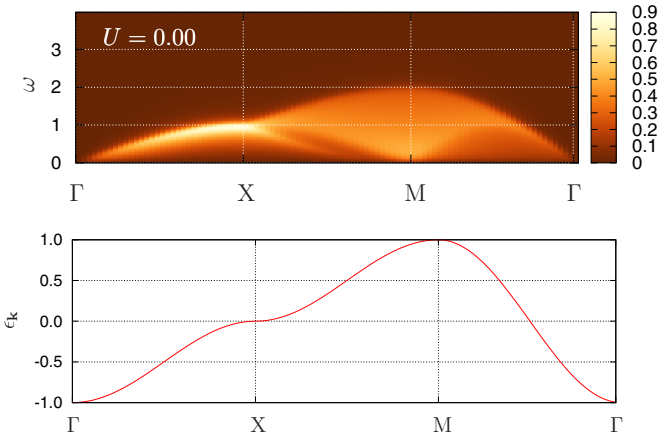

FIG. 3. (Color online) Top panel: Lattice susceptibility $-\frac{1}{\pi}\,\mathrm{Im}\,\chi_\omega(\mathbf{q})$ for the half-filled noninteracting two-dimensional model (i.e., $U = 0$) with full bandwidth $W = 2$ and at finite temperature $T = 0.02$, along a high-symmetry path in momentum space. Γ, X, and M denote the wave vectors $(0,0)$, $(0,\pi)$, and (π,π), respectively. Lower panel: Noninteracting dispersion.

respectively) and therefore corresponds to the energy $\omega = 1$, which is equal to the half bandwidth. One can further see that the energy of the particle-hole excitations approaches zero in the long-wavelength limit.

In Fig. 4, we show results for the interacting case obtained from a standard random phase approximation (RPA) calculation. In RPA, the charge correlation function reads as

$$\chi^{\mathrm{RPA}}_\omega(\mathbf{q}) = \frac{\chi^0_\omega(\mathbf{q})}{1 - U\,\chi^0_\omega(\mathbf{q})}, \tag{21}$$

with χ^0 defined in (20) [the minus sign in the denominator comes from the minus sign in the definition of $\chi_\omega(\mathbf{q})$, Eq. (4)]. By construction, the RPA is of course only reliable for small values of U. Here, we plot RPA results for larger values of the interaction for a comparison with the correlated case. For $U = 0.5$, we observe a picture that is similar to the noninteracting case. The same structures are present also for larger values of U, albeit some of them, e.g., the structure which has an energy minimum at the M point, become less visible. As U increases, the largest overall response shifts from X to the to M point above $U = 1$. The collective excitation becomes better defined. Because the interaction is short ranged, this collective mode, the zero-sound mode, goes to zero energy in the long-wavelength limit for all values of U. For values of U larger than the bandwidth, the maximum energy at the M point is determined by the energy scale U instead of W.

Using DMFT, we can now investigate whether and how this physical picture is modified in a strongly correlated metal, close to a Mott transition. For completeness, let us start by briefly showing some well-known aspects of the Mott transition in DMFT, as illustrated in Figs. 5 and 6. Figure 5 shows the corresponding quasiparticle weight $Z = (1 - d\,\mathrm{Re}\,\Sigma_\omega/d\omega)^{-1}$ as a function of U computed using a polynomial extrapolation of the self-energy on Matsubara frequencies.[3] On this lattice, the transition occurs at $U_c = 2.36 \pm 0.01$. Z vanishes at the Mott transition, corresponding to a divergent effective mass [1] $m^*/m \sim 1/Z$. The static homogeneous charge susceptibility $\lim_{\mathbf{q}\to 0}\lim_{\omega\to 0}\chi(\omega,\mathbf{q}) = -dn/d\mu$ is shown in the same figure. It is proportional to the compressibility and therefore vanishes in the insulator. In Fig. 6, we plot the local density of states for different values of U. One can see a well-defined quasiparticle peak and the Hubbard bands at $\omega \sim U$. For values above the transition, the density of states exhibits a gap.

The DMFT susceptibility including vertex corrections is shown in Fig. 7 for the same parameters as the RPA results in Fig. 4. We obtain it by analytical continuation from Matsubara frequencies. In the weakly correlated regime, for U up to about 1, the results are similar to the RPA, albeit we observe a somewhat broader spectrum. In particular, one can see the minimum at the M point, which is present up to at least $U \sim 1$. In the moderately correlated regime $U \sim 1.5$, this feature is no longer resolved in our data, but the spectra retain a similar shape as in RPA, showing a well-defined mode for all wave vectors. Its maximum at M is still approximately equal to

[3]We have used polynomials up to degree six. The results are converged for all values of U for polynomials of degree five.

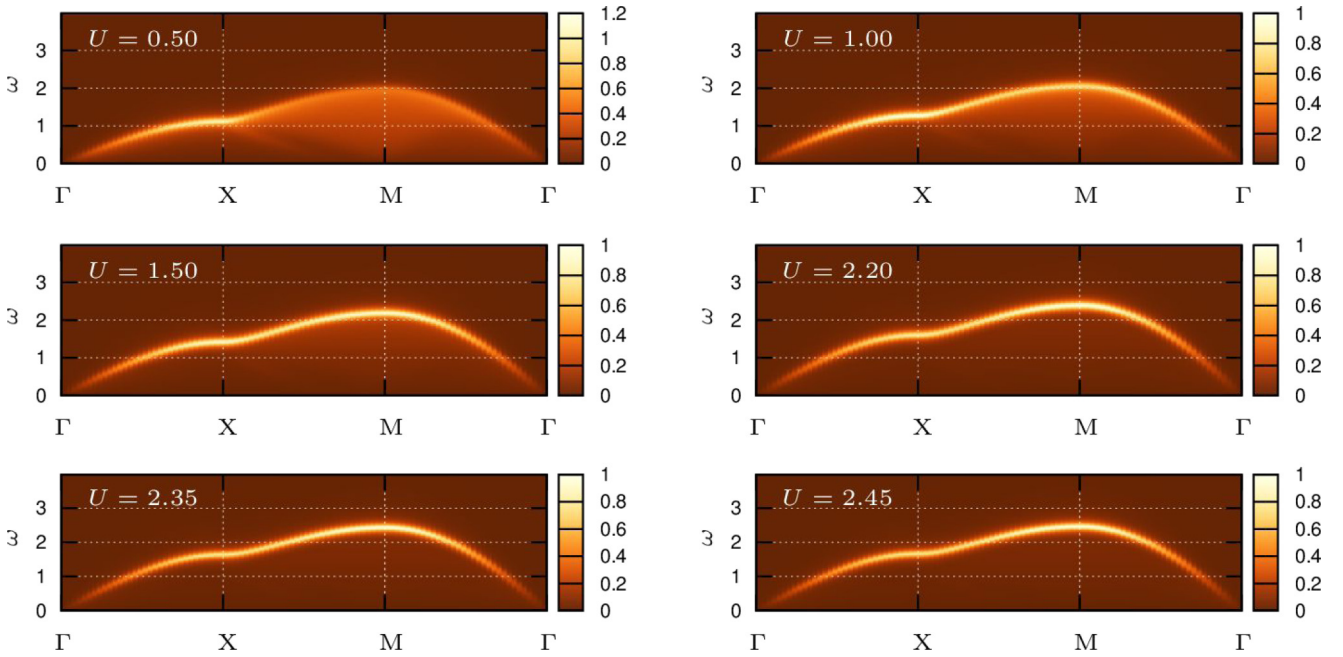

FIG. 4. (Color online) RPA results for the imaginary part of the charge susceptibility $-\frac{1}{\pi}\,\mathrm{Im}\,\chi_\omega(\mathbf{q})$ of the half-filled Hubbard model for various values of U and $T = 0.02$.

the bandwidth. As the transition is approached, however, the spectrum changes substantially. It is considerably broadened and damped at the X point, while at the M point it gains relative intensity. For large interaction, the maximum at the M point occurs at the scale of U.

The collective mode is visible all the way to the transition. Its frequency vanishes in the long-wavelength limit. This is expected for reasons we shall explain below. We cannot strictly exclude even qualitative changes in particular of the high-energy features in the spectra because of the ill-conditioned nature of the analytical continuation problem (see Appendix A for a discussion of the analytical continuation procedure). We can, however, further substantiate the results for the low-energy collective mode directly from the Matsubara data: The polarization $\Pi_\omega(\mathbf{q})$ defined by

$$\chi_\omega(\mathbf{q}) = \frac{-\Pi_\omega(\mathbf{q})}{1 + U\Pi_\omega(\mathbf{q})} \tag{22}$$

is a function of q/ω for small q, as explained in Sec. III E. We fit the Matsubara data with an expression of the form $\Pi_\omega(\mathbf{q}) = -b(q/i\omega_m)^2/[1 + c^2(q/i\omega_m)^2]$, for a small Matsubara frequency $m = 3$, where b and c are the free parameters. We can readily analytically continue this function by letting $i\omega \to \omega + i0^+$. The dispersion is defined by the zeros of the denominator in (22) and can be expressed in terms of the fit parameters as $\omega(q) = q\sqrt{c^2 + bU}$. The thus obtained dispersion is indicated by a white line in Fig. 7. In Fig. 8, we plot the maximum of the charge susceptibility for small wave vectors together with the linear dispersion obtained from the fits. The data are in good agreement showing that the Padé approximation is reliable within this energy range. There is no appreciable change in slope of the mode with increasing interaction. Although it is less visible in Fig. 7 due to decreasing contrast close to the transition, it remains well defined as can be seen from a fixed momentum cut of the susceptibility shown in the inset of Fig. 8. In the insulator,

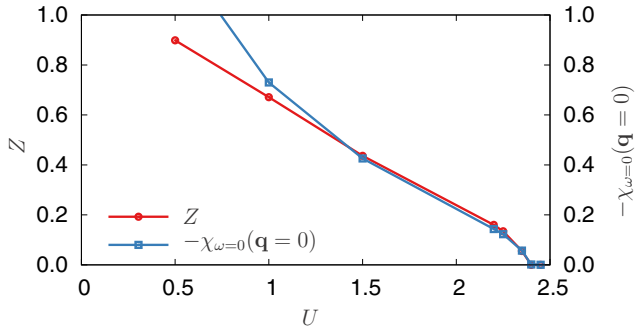

FIG. 5. (Color online) Quasiparticle weight Z and $\lim_{q\to 0}\lim_{\omega\to 0}\chi_\omega(\mathbf{q}) = -dn/d\mu$ as a function of U. Both vanish in the insulating phase. Close to the transition Z is proportional to $dn/d\mu$.

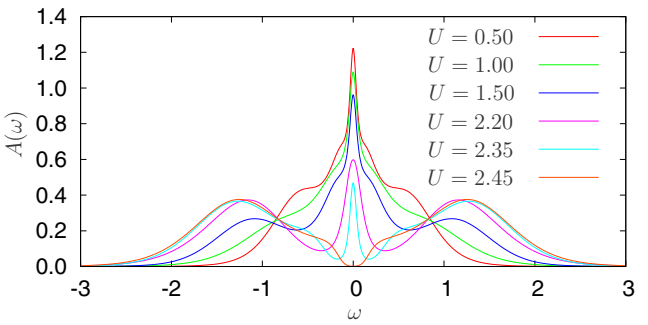

FIG. 6. (Color online) The finite-temperature maximum entropy density of states for various values of U at $T = 0.02$, showing the evolution from a weakly to a moderately and strongly correlated system. The Mott transition occurs at $U \sim 2.36$.

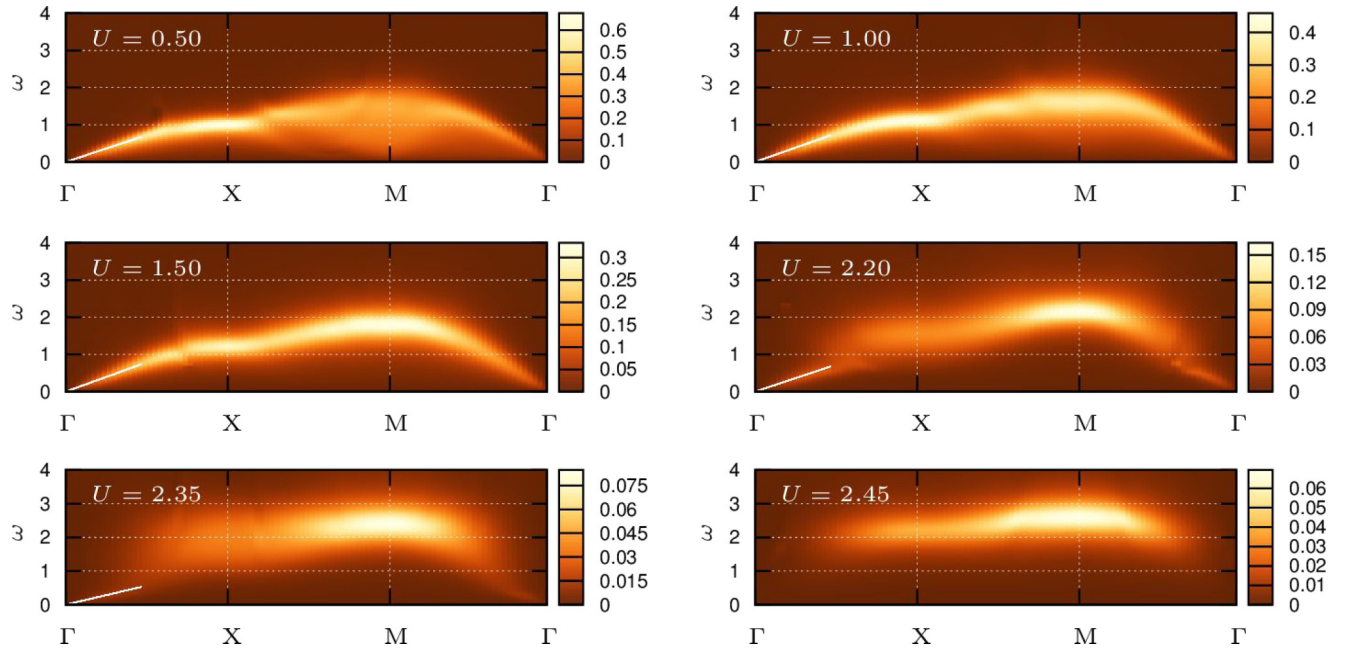

FIG. 7. (Color online) Imaginary part $-\frac{1}{\pi}\,\mathrm{Im}\,\chi_\omega(\mathbf{q})$ of the DMFT charge susceptibility including vertex corrections obtained via analytical continuation using Padé approximants. The low-energy dispersion obtained from a fit of the Matsubara data is indicated by a white line (cf. text).

this mode disappears as expected. All excitations acquire a minimum energy $\sim U$.

We are now going to show that the *nonlocal* (reducible) vertex corrections play an essential role to obtain a correct description of the low-energy physics, *even at a very qualitative level*. Because incorporating the full vertex corrections is technically demanding, they are often neglected and the susceptibility is often approximated by a simple bubble approximation, i.e., by a product of *interacting* DMFT Green's functions $\chi^0_\omega(q) = (T/N)\sum_{\mathbf{k}\nu} G_{\nu+\omega}(\mathbf{k}+\mathbf{q})G_\nu(\mathbf{k})$. In Fig. 9, we plot the susceptibility for $U = 2.2$, obtained within this bubble approximation. It is essentially featureless for all wave vectors. In the long-wavelength limit, it exhibits spectral weight at *finite* energy in contradiction to the foregoing and the standard textbook RPA approximation. For small but finite U, the bubble approximation still exhibits a mode which goes, at least approximately, to zero in the long-wavelength limit. In the correlated regime (larger U), however, the vertex corrections are essential for a qualitatively correct description of the collective excitations. In the next section, we will relate the existence of the zero sound to the gauge invariance and the associated Ward identities. The failure of the bubble approximation can be traced back to a violation of gauge invariance.

We note that in contrast to the susceptibility, the optical conductivity in DMFT is unaffected by the vertex corrections in the long-wavelength limit. Vertex corrections drop out of the conductivity as a consequence of the locality of the irreducible vertex and the inversion symmetry of the lattice (see Refs. [1,34]).

A natural question to ask at this stage is whether an approximation with only *local* vertex corrections could be sufficient to capture the zero-sound mode. To gain more insight into this question, we consider a more advanced approximation than the bubble. Namely, we compute the susceptibility within an approximation that includes short-range vertex corrections,

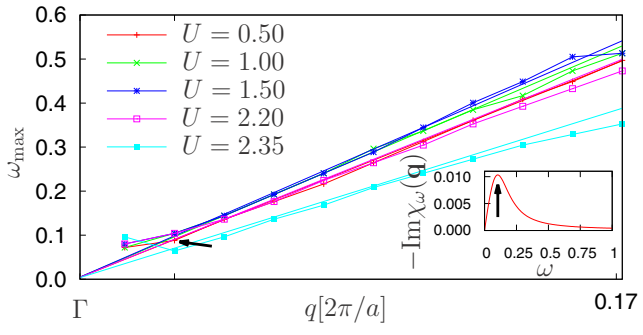

FIG. 8. (Color online) Maxima of the charge susceptibility of Fig. 7 (lines with symbols) and dispersion obtained from a fit of the Matsubara data (lines, cf. text) for small $\mathbf{q}$ vectors up to about one third the way to the X point indicating the slope of the zero-sound mode. Inset: cut at small fixed momentum $q = 0.03125\,(2\pi/a)$ through the charge susceptibility for $U = 2.2$. The arrow indicates the maximum which corresponds to the point marked by an arrow in the main panel.

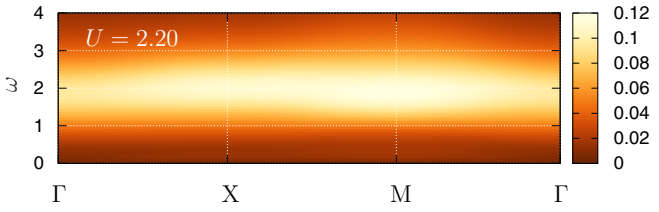

FIG. 9. (Color online) $-\frac{1}{\pi}\,\mathrm{Im}\,\chi^0_\omega(\mathbf{q})$ in the bubble approximation, i.e., *without vertex corrections* for local interaction $U = 2.2$ and otherwise the same parameters as in Fig. 7. In the bubble approximation, the zero-sound mode disappears.

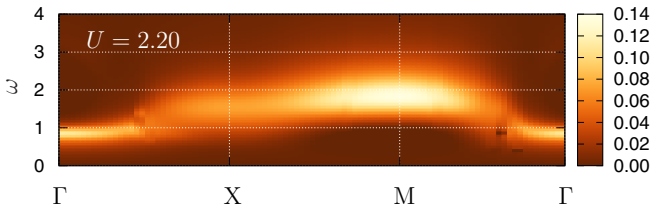

FIG. 10. (Color online) Lattice susceptibility $-\frac{1}{\pi}\,\mathrm{Im}\,\chi_\omega(\mathbf{q})$ obtained by neglecting nonlocal, long-range vertex corrections, more precisely by neglecting the second term in Eq. (17). Parameters are the same as in Fig. 9. The plasmonlike mode at finite energy and long wavelengths is an artifact of the approximation. Local vertex corrections are not sufficient to restore the zero-sound mode.

but neglects correlations from the *reducible* vertex of the lattice Γ. To be precise, we compute it using Eqs. (15)–(18), however neglecting the second term containing the lattice vertex Γ in Eq. (17). We note that this approximation is not exactly the same as a bubble approximation with local vertex corrections added. The important point here is that such an approximation neglects ladder diagrams containing many repeated particle-hole scattering processes on different lattice sites contained in Γ. The result is shown in Fig. 10. As expected, the long-wavelength properties are clearly not reproduced correctly in this approximation. It wrongly predicts a finite-energy collective mode. We note that in the long-wavelength limit, the ladder diagrams contribute at all orders. Low-order diagrams that describe charge correlation contain independent particle-hole propagation (described by the bubble) over large distances, which is unlikely. Therefore, it is physically clear that higher-order diagrams are important. Features at finite wave vectors sufficiently far from the Γ point are, however, remarkably well captured. Here, the short-range vertex corrections are necessary, but also appear to be sufficient to reproduce qualitative features.

E. Gauge invariance

In this section, we discuss the consequences of gauge invariance. The failure of the bubble approximation to describe the low-energy collective modes can explicitly be traced back to its violation. We will first check that the standard DMFT susceptibility [1] as described in Sec. III B [Eqs. (11) and (12)] and in the review Ref. [1] yields a response that obeys local charge conservation and leads to gauge-invariant results. We recall that for the response to be conserving, two conditions have to be fulfilled [29,30]: (i) The self-energy and the irreducible vertex function have to be given by functional derivatives of the Luttinger-Ward functional $\Phi[G_{ij}]$, i.e., $\Sigma_{ij} = \delta\Phi/\delta G_{ji}$ and $\Gamma^{\mathrm{irr}}_{ijkl} = \delta^2\Phi/\delta G_{ji}\delta G_{lk}$. (ii) The generalized susceptibility L has to be constructed such that it gives the change in G through a perturbation A to linear order: $L := -\delta G/\delta A = G(\delta G^{-1}/\delta A)G$. With $G^{-1} = G_0^{-1} - A - \Sigma[G]$ and for approximations for which Σ depends on the perturbation through G only, this leads to the integral equation

$$L = -GG - GG\frac{\delta\Sigma}{\delta A} = -GG - GG\frac{\delta\Sigma}{\delta G}\frac{\delta G}{\delta A} = -GG + GG\frac{\delta\Sigma}{\delta G}L. \qquad (23)$$

Combined with condition (i), one identifies $\Gamma^{\mathrm{irr}} := \delta\Sigma/\delta G = \delta^2\Phi/\delta G^2$. Condition (ii) is evidently fulfilled in DMFT since the above equation is equivalent to the integral equation (7) (written in terms of $\chi := -L$ instead of L). To address the first condition, we recall that in DMFT, the Luttinger-Ward functional depends on the site-diagonal Green's functions only [1], allowing for a decomposition in terms of local functionals $\Phi[G_{i'j'}] = \sum_{i'}\phi[G_{i'i'}]$. This implies that the self-energy is local:

$$\Sigma_{ij} = \frac{\delta\Phi[G_{i'j'}]}{\delta G_{ji}} = \frac{\delta\phi[G_{i'i'}]}{\delta G_{ii}}\delta_{ji}. \qquad (24)$$

The same holds for the irreducible vertex function

$$\Gamma^{\mathrm{irr}}_{ijkl} = \frac{\delta^2\Phi[G_{i'j'}]}{\delta G_{ji}\delta G_{lk}} = \frac{\delta^2\phi[G_{l'l'}]}{\delta G_{ll}^2}\delta_{li}\delta_{lj}\delta_{lk}. \qquad (25)$$

The functional $\phi[G_{i'i'}]$ as well as Σ and Γ^{irr} are generated from the impurity model subject to the self-consistency condition (3). If the self-energy and vertex function are obtained numerically exactly from the solution of the impurity model, condition (i) is hence fulfilled. Note that the dimensionality of the lattice does not enter this argument. It therefore holds on finite-dimensional lattices, where DMFT is an approximation. It also holds for cluster extensions of dynamical mean-field theory.

We can make the argument more explicit by recalling that charge conservation is commonly expressed in terms of a Ward identity. It can be viewed as the Green's function analog of the continuity equation and relates the single-particle Green's function to a vertex function. The above two conditions are sufficient for the Ward identity to be fulfilled, as can be shown by considering a perturbation which corresponds to a gauge transformation [30].

On a discrete lattice, gauge invariance and charge conservation can be preserved exactly, even for finite lattice spacing [35]. Projecting the continuum system onto a discrete Wannier basis under the assumption of weak and slowly varying fields leads to the gauge theory described here (see Appendix B and, e.g., Refs. [36,37]). In Appendix D, we show that with the proper definition of the current and a suitable generalization of the notion of the derivative to the lattice, the Ward identity can be written

$$q_\mu^F\Gamma_\mu(k,q) = G^{-1}(k) - G^{-1}(k+q), \qquad (26)$$

where we have introduced four-vector notation for clarity (only in this section). Summation over the time ($\mu = 0$) and spatial components $\mu = 1,2,3$ is implied using the metric $(-1,1,1,1)$. The corresponding continuity equation is $\partial n/\partial t + \nabla^F\cdot\mathbf{j} = 0$, where ∇^F denotes a *forward* derivative. It corresponds to a finite difference expression owing to the discrete structure of the lattice. In the above, the main difference to the continuum case is the appearance of the momentum $q_\mu^F \equiv (i\omega,\mathbf{q}^F)$ associated with a forward derivative. On a finite lattice of N sites with periodic boundary conditions it has the spatial components $q_\alpha^F = -(i/a)[\exp(iq_\alpha a) - 1]$, where the index q_α takes on the discrete values $q_\alpha^{(n)} = 2\pi n/N$, with n integer. In the above equation, Γ_μ is the renormalized current vertex, which describes the interaction of the interacting electrons with the electromagnetic field. The Ward identity hence relates

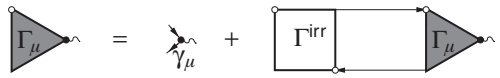

FIG. 11. Diagrammatic representation of the Bethe-Salpeter equation for the renormalized current vertex Γ_μ. Lines are fully dressed propagators.

a vertex function to the single-particle properties described by the Green's function G.

Let us now explicitly check the Ward identity in DMFT. The current vertex obeys a ladder equation, which follows from the Bethe-Salpeter equation for the susceptibility [Eq. (7)]. In DMFT, the ladder equation for the current vertex reads as (see Fig. 11 and Appendix E)

$$\Gamma_{\mu;\nu}(\mathbf{k},\mathbf{q}) = \gamma_\mu(\mathbf{k},\mathbf{q}) - \frac{T}{N}\sum_{\nu'\mathbf{k}'} \Gamma^{\mathrm{irr}}_{\nu\nu'\omega} \times G_{\nu'\sigma'}(\mathbf{k}')G_{\nu'+\omega}(\mathbf{k}'+\mathbf{q})\Gamma_{\mu;\nu'}(\mathbf{k}',\mathbf{q}). \quad (27)$$

Here, the *bare* current vertex is given by

$$\gamma_\mu = \begin{cases} i\tilde{t}a(e^{-i(k_\alpha+q_\alpha)} - e^{-ik_\alpha}), & \mu = \alpha = x,y,z \\ 1, & \mu = 0 \end{cases} \quad (28)$$

which itself obeys the Ward identity $q^F_\mu \gamma_\mu(k,q) = G_0^{-1}(k) - G_0^{-1}(k+q)$ with the noninteracting Green's function $G_0^{-1}(k) = i\nu + \mu - \varepsilon_{\mathbf{k}}$. In order to demonstrate that the DMFT susceptibility is conserving, we have to show that the current vertex obeys the Ward identity (26). To this end, we form the quantity $q^F_\mu \Gamma_\mu(k,q)$ using (27):

$$q^F_\mu \Gamma_{\mu;\nu}(\mathbf{k},\mathbf{q}) = q^F_\mu \gamma_\mu(\mathbf{k},\mathbf{q}) - \frac{T}{N}\sum_{\nu'\mathbf{k}'} \Gamma^{\mathrm{irr}}_{\nu\nu'\omega} \times G_{\nu'}(\mathbf{k}')G_{\nu'+\omega}(\mathbf{k}'+\mathbf{q})\left[q^F_\mu \Gamma_{\mu;\nu'}(\mathbf{k}',\mathbf{q})\right]. \quad (29)$$

Inserting the Ward identity (26) for the interacting and noninteracting current vertices on both sides and using the definition of the Green's function (2), all momentum dependence cancels exactly. Since the irreducible vertex is local, the DMFT self-consistency condition (3) further allows us to express the local part of the lattice Green's function in terms of the impurity Green's function. We thus obtain the purely local equation

$$\Sigma_{\nu+\omega} - \Sigma_\nu = -T\sum_{\nu'} \Gamma^{\mathrm{irr}}_{\nu\nu'\omega}\left[g_{\nu'+\omega} - g_{\nu'}\right], \quad (30)$$

which involves impurity quantities only.[4] We therefore conclude that the Ward identity is fulfilled if this equation is satisfied. It can be viewed as a local version of the Ward identity,[5] which is fulfilled for the impurity model. Gauge invariance and local charge conservation are hence guaranteed for the susceptibilities. This is completely in line with the previous argument: Charge conservation of the DMFT susceptibility follows if the self-energy and irreducible vertex function are determined from the exactly solvable impurity model.[6] While being an approximation in finite dimensions, the DMFT susceptibility preserves local charge conservation exactly, even on a finite-dimensional lattice.

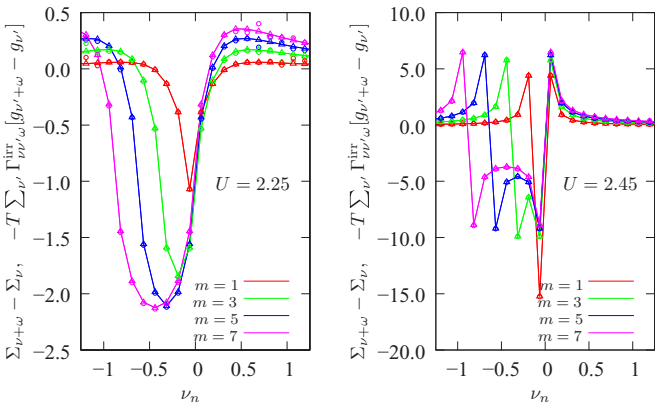

FIG. 12. (Color online) Numerical illustration of the fulfillment of the local Ward identity. Both sides of the equation (30) are plotted for different bosonic frequencies $\omega_m = 2m\pi/\beta$ for the two-dimensional Hubbard model in DMFT for two different values of U somewhat below (left) and above (right) the Mott transition. Results for the left-hand side of the equation $\Sigma_{\nu+\omega} - \Sigma_\nu$ are marked by lines with triangles and for the right-hand side $-T\sum_{\nu'}\Gamma^{\mathrm{irr}}_{\nu\nu'\omega}[g_{\nu'+\omega} - g_{\nu'}]$ by circles, showing good agreement. Note that for $\omega_m = 0$, the equation is identically fulfilled as both sides vanish.

In our calculations we obtain the vertex, self-energy, and Green's function by solving the impurity model numerically exactly. That the Ward identity is indeed fulfilled numerically is illustrated in Fig. 12, where we plot both sides of Eq. (30) for different bosonic frequencies. In order to evaluate the frequency sum on the right-hand side, we have replaced the irreducible vertex by $-U$ above the frequency cutoff up to which it is calculated explicitly. The identity is evidently well fulfilled. For high frequencies, deviations occur which partly originate from the numerical noise which increases with frequency, as well as from the finite frequency cutoff of the vertex function. The latter is computed by inverting a local BSE, which is affected by the finite frequency cutoff. In general, for a correct description of the collective excitations, the low-energy behavior of the vertex function is decisive, which is well captured in our calculations. Note that the numerical error seen in this figure does not propagate into the calculation of the lattice susceptibility because the irreducible vertex does not have to be computed (see Sec. III C).

We would like to point out the relation of the above to the self-consistent Hartree-Fock approximation and RPA [38]. In static mean-field theory, the above equations still hold with $\Sigma = U\langle n\rangle$ and $\Gamma^{\mathrm{irr}} = -U$. As a consequence, Eq. (30) is identically fulfilled, showing that the Hartree-Fock

[4]The minus sign on the right-hand side of Eq. (30) stems from the fact that our convention for the irreducible vertex is such that $\Gamma^{\mathrm{irr}} = -U$ to lowest order in U.

[5]By writing Eq. (30) in differential form, Γ^{irr} is identified as the functional derivative $\delta\Sigma/\delta g$: The variation of a self-energy diagram is given by the functional derivative of Σ with respect to g times the variation in g.

[6]For an approximate impurity solver, the susceptibility is conserving as long as Eq. (30) is fulfilled.

approximation is conserving. The Bethe-Salpeter equation for the response function is then equivalent to the random phase approximation (RPA) for the susceptibility [Eq. (21)].

As mentioned previously, it is known that the zero-sound mode is a consequence of gauge invariance. To see this, consider the electromagnetic response kernel $K_{\mu\nu}$ defined through

$$J_\mu(q) = K_{\mu\nu}(q)A_\nu(q), \tag{31}$$

where J_μ is the expectation value of the current. Gauge invariance and local charge conservation imply the two conditions (see Appendix F)

$$K_{\mu\nu}(q)q^F_\nu = 0, \tag{32}$$

$$q^F_\mu K_{\mu\nu}(q) = 0. \tag{33}$$

For the longitudinal response ($\mathbf{q}$ pointing along direction 3), one obtains $K_{03}q - i\omega K_{00} = 0$ and $K_{33}q - i\omega K_{03} = 0$. (In the following, we are interested in the long-wavelength behavior for which we may replace q^F by q.) The other components do not contribute by symmetry. Combining these two equations yields the continuity equation

$$K_{00}(q) = \frac{q^2}{(i\omega)^2}K_{33}(q). \tag{34}$$

In the long-wavelength limit, only the diamagnetic contribution to the response kernel contributes to K_{33}, which is independent of q and ω (see Appendix F and Ref. [39]). Hence, the susceptibility $\chi_\omega(\mathbf{q}) = -K_{00}(q)/e^2$ is a function of the ratio q/ω. The same holds for the polarization since it is related to χ through a simple geometric series:

$$\chi_\omega(\mathbf{q}) = \frac{-\Pi_\omega(\mathbf{q})}{1 + U\Pi_\omega(\mathbf{q})}. \tag{35}$$

The dispersion of the collective mode is determined by the poles of χ, or as the solution to the equation $1 + U\Pi_\omega(\mathbf{q}) = 0$. Since the interaction U is constant (in general, a short-range interaction remains finite in the limit $q \to 0$), the solution to the above equation must be a kind of sound, i.e., $q/\omega = \text{const}$. The analysis is the same as in the textbook RPA case, except that the polarization $\Pi_\omega(\mathbf{q})$ is a function produced by the DMFT calculation instead of the Lindhardt function in the RPA case. The key fact is that for small momentum, $\Pi_\omega(\mathbf{q})$ is a function of q/ω and hence not a continuous function of (ω,q) at (0,0). As shown above, this is a consequence of gauge invariance. Moreover, every approximation that violates the Ward identity is likely to miss this singularity of the function, and will not be able to reproduce the correct low-energy behavior.

Since we use a quantum Monte Carlo impurity solver, it is useful being able to observe the restriction imposed by gauge invariance also on the level of the imaginary-time data. To this end, we rewrite (34) as

$$(i\omega)^2 K_{00}(q) = (q)^2 K_{33}(q). \tag{36}$$

Using again that in the limit $q \to 0$ K_{33} remains finite, we find the following condition imposed on the susceptibility:

$$(i\omega)^2 \chi_\omega(\mathbf{q}) \underset{q\to 0}{=} 0. \tag{37}$$

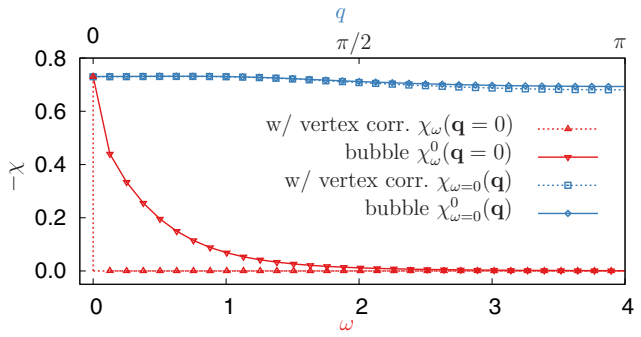

FIG. 13. (Color online) The DMFT susceptibility $\chi_\omega(\mathbf{q})$ for $\mathbf{q} = 0$ (red curve) and $\omega = 0$ (blue curve), for $U = 1$, $T = 0.02$. As a consequence of gauge invariance, the limits $\omega \to 0$ and $\mathbf{q} \to 0$ of $\chi_\omega(\mathbf{q})$ do not commute. For a frequency-dependent self-energy, this condition is respected by including vertex corrections. It is violated for the bubble approximation for which the limits commute.

The charge susceptibility has to vanish for any finite frequency in this limit and hence also for subsequently taking the limit $\omega \to 0$. On the other hand, taking the limit $\omega \to 0$ first leads to the the static response $\lim_{q\to 0}\lim_{\omega\to 0}\chi_\omega(\mathbf{q}) = -dn/d\mu$ which is finite in the metallic phase. Hence, the limits $\lim_{\omega\to 0}$ and $\lim_{q\to 0}$ do not commute, which implies a discontinuous jump in the susceptibility. The Lindhardt bubble, corresponding to the noninteracting result, has the required property [40,41]. The noncommutativity of the susceptibility comes from the product of Green's functions $\chi^0_{kq} = G_k G_{k+q}$, which enters the vertex through the Bethe-Salpeter equation for the vertex function [Eq. (12)]. It is singular in this limit because the poles of the Green's functions merge when $q \to 0$ (see, e.g., Ref. [39], Chap. 6, Sec. 4).

Figure 13 shows that the susceptibility including vertex corrections indeed vanishes for finite frequencies and displays a discontinuity.[7] The bubble approximation computed from interacting Green's functions, on the other hand, clearly violates gauge invariance. The result is continuous, which explains the failure of the bubble approximation observed previously.

One can expect that the discontinuity (which can be used as a rigorous test for the implementation) will only be restored by summing an infinite number of diagrams beyond the bubble. Physically, it is clear that in order to describe the long-wavelength behavior of the two-particle excitations and the response functions, repeated particle-hole scattering generated through the Bethe-Salpeter equation is essential. At the same time, the Bethe-Salpeter equation accounts for the collective mode and ensures gauge invariance. In this sense, the collective excitations are key to the gauge-invariant character of the theory. A different way of seeing this is the fact that the effective quasiparticle interaction determined by the vertex function generates the backflow of electrons around a quasiparticle moving through the medium [39]. This backflow

[7]The condition (37) can be used to benchmark the accuracy of the simulation. Deviations from zero occur due to Monte Carlo noise in the vertex function and an insufficient frequency cutoff.

is necessary to fulfill the continuity equation and hence to assure local charge conservation. In the insulator, $dn/d\mu = 0$ because of the gap so that the discontinuity disappears and with it the zero-sound mode.

IV. LONG-RANGE INTERACTION: THE DUAL BOSON APPROACH

Plasmons are long-wavelength excitations of the electron gas with a finite energy, which appear in presence of a Coulomb potential. Here, we use the following expression for a Coulomb-type potential in three dimensions,

$$V(\mathbf{q}) = \begin{cases} e^2 \frac{V}{q^2} & q \neq 0, \\ 0 & q = 0, \end{cases} \tag{38}$$

where e is the electron charge. Setting the interaction to zero for $q = 0$ corresponds to adding a homogeneous positively charged background which compensates the negative charge of the electron gas.

Nonlocal interactions can be treated on the basis of extended dynamical mean-field theory [11,12,42]. In EDMFT, the lattice model is mapped to a local impurity problem which contains a local, but retarded interaction W_ω. The self-energy in EDMFT is determined through the self-consistency condition (3) as in DMFT. The dynamical interaction accounts for the dynamical screening of the local charge due to the nonlocal interaction $V(\mathbf{q})$. It is determined through an additional self-consistency condition

$$\chi_\omega = \sum_{\mathbf{k}} X_\omega(\mathbf{k}), \tag{39}$$

which is written in terms of the two-particle propagator (the lattice susceptibility)

$$X_\omega(\mathbf{q}) = \frac{1}{\chi_\omega^{-1} + W_\omega - V(\mathbf{q})}. \tag{40}$$

We denote it by $X_\omega(\mathbf{q})$ in order to distinguish it from the DMFT susceptibility $\chi_\omega(\mathbf{q})$. In the above, χ_ω is the impurity charge susceptibility. This equation can be understood as follows. Consider the representation of the susceptibility in terms of the polarization Π:

$$\begin{aligned} X_\omega(\mathbf{q}) &= -\Pi_\omega(\mathbf{q}) + \Pi_\omega(\mathbf{q})V(\mathbf{q})\Pi_\omega(\mathbf{q}) \mp \cdots \\ &= \frac{-\Pi_\omega(\mathbf{q})}{1 + V(\mathbf{q})\Pi_\omega(\mathbf{q})} = \frac{1}{-\Pi_\omega(\mathbf{q})^{-1} - V(\mathbf{q})}, \end{aligned} \tag{41}$$

which is a simple geometric series. Π contains all diagrams irreducible with respect to the interaction $V_\mathbf{q}$. In EDMFT, the polarization is obtained from the impurity model, which yields a nonperturbative, albeit local result. The impurity susceptibility χ_ω contains polarization diagrams. However, it cannot directly be used as the polarization because it contains diagrams reducible in W. Let us denote $\Pi_\omega^W = -\chi_\omega$. We can easily take out these reducible contributions by writing Π_ω^W as a geometric series: $\Pi_\omega^W = \Pi/(1 + W_\omega \Pi_\omega)$ or $\Pi_\omega^{-1} = (\Pi_\omega^W)^{-1} - W_\omega = -\chi_\omega^{-1} - W_\omega$. Inserting this into (41) recovers (40).

In case of a Coulomb potential, the collective charge excitations are plasmons, whose dispersion relation $\omega(q)$ is

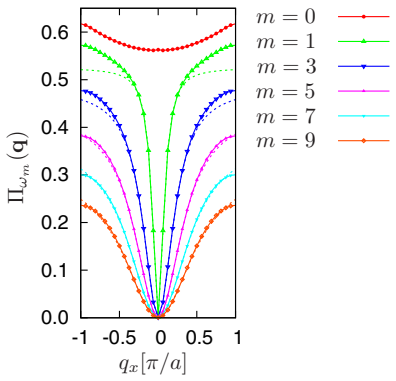

FIG. 14. (Color online) Polarization $\Pi_{\omega_m}(\mathbf{q})$ as a function of momentum q_x for $q_y = 0$ and different Matsubara frequencies ω_m. For $\omega_m > 0$ and small q, we clearly observe the $\sim q^2$ behavior of the polarization as required by gauge invariance. Dashed lines show fits to a $-b(q/i\omega_m)^2/[1 + c^2(q/i\omega_m)^2]$ behavior for $\omega_m > 0$ (b and c are fit parameters). When q is small compared to the frequency, the q^2 behavior is clearly visible. For small ω_m, it is less visible due to the finite momentum resolution.

solution of

$$1 + V(\mathbf{q})\Pi_\omega(\mathbf{q}) = 0. \tag{42}$$

In the standard RPA analysis (where Π is just the Lindhardt function), for $q \to 0$, ω finite, one has (with f some function, and g the coupling constant)

$$\Pi_\omega(\mathbf{q}) \sim g q^2 f(\omega) + \mathcal{O}(q^4), \tag{43}$$

which yields the plasmon dispersion relation $\omega(\mathbf{q}) = \omega_p + aq^2$ at small q, where a is a constant and ω_p the plasma frequency. ω_p is the solution of

$$1 + g e^2 V f(\omega_p) = 0. \tag{44}$$

From the previous discussion of gauge invariance, we know that the susceptibility vanishes in the long-wavelength limit for finite frequencies $X_\omega(\mathbf{q} \to 0) = 0$. By virtue of (41), we expect the same behavior for the polarization, i.e., $\Pi(\mathbf{q} \to 0, \omega) = 0$. This behavior can be observed directly on the Matsubara data, as shown in Fig. 14: for finite Matsubara frequencies, the polarization vanishes in the long-wavelength limit. At small momenta, the data are well described by a function of the form $-b(q/i\omega_m)^2/[1 + c^2(q/i\omega_m)^2]$. For q small compared to $i\omega_m$, the polarization is clearly proportional to q^2. One therefore expects that the standard RPA analysis still holds in the correlated regime.

In the EDMFT approximation, however, the polarization is computed from the local susceptibility and hence momentum independent. Therefore, the equation for the dispersion relation is

$$1 + \frac{e^2 V \Pi_\omega}{q^2} = 0. \tag{45}$$

The polarization has to decrease as a function of large frequencies and to leading order we expect it to behave as $\Pi(\omega) \sim 1/\omega^\alpha$, with some $\alpha > 0$. As a consequence, the frequency of the excitation will diverge in the long-wavelength limit: $\omega \sim 1/q^{2/\alpha}$. This behavior is shown in the left panel of Fig. 15. We note that *a priori* a solution with a finite

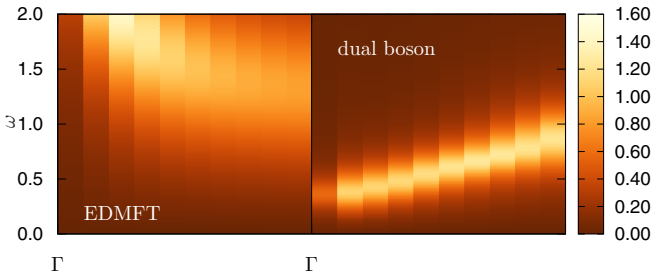

FIG. 15. (Color online) $-\mathrm{Im}\,\epsilon^{-1}(E,\mathbf{q})$ for small values wave vectors (up to $|\mathbf{q}| \sim 2\pi \times 0.34$) for $U = V = 1$ and $T = 0.02$. The energy of the collective mode diverges in EDMFT for $|\mathbf{q}| \to 0$ due to violation of gauge invariance. This unphysical behavior is corrected by including nonlocal vertex corrections, yielding a finite-energy plasmon mode.

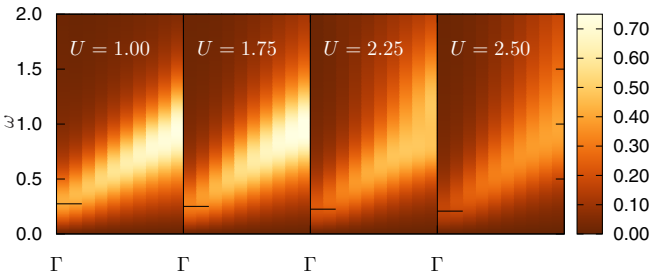

FIG. 16. (Color online) $-\mathrm{Im}\,\epsilon^{-1}(E,\mathbf{q})$ for small values wave vectors up to $|\mathbf{q}| \sim 2\pi \times 0.34$ showing the plasmon mode in the three-dimensional Hubbard model with Coulomb interaction $V(\mathbf{q}) \propto \frac{1}{\mathbf{q}^2}$. Data are shown for $V = 0.5$ and different values of U at $T = 0.02$. The plasma frequency corresponding to the maximum intensity of the peak for $q \to 0$ (indicated by the horizontal line) decreases with increasing U.

plasma frequency could exist if the polarization were to vanish for a finite ω. We do not observe this in our calculations, however. We therefore find that EDMFT does not provide a valid description of plasmons.

Let us now consider the dual boson approach [6], which can be viewed as a diagrammatic expansion around extended dynamical mean-field theory. Additional details on the approach and its convergence properties can be found in Refs. [6,43]. In this approach, the description of the collective modes amounts to replacing (40) by

$$X_\omega(\mathbf{q}) = \frac{1}{\chi_\omega^{-1}(\mathbf{q}) + W_\omega - V(\mathbf{q})}, \tag{46}$$

where $\chi_\omega(\mathbf{q})$ is given by Eqs. (15)–(18), which includes nonlocal vertex corrections into the EDMFT susceptibility. In EDMFT, there is an ambiguity of calculating the susceptibility. It is either given by the bosonic propagator (40), or can be obtained similarly as in DMFT. In the latter case, one computes the susceptibility from (46) with $\chi_\omega(\mathbf{q})$ computed from the DMFT expressions of Sec. III B. In dual boson, the bosonic propagator is also given by (46), however, with $\chi_\omega(\mathbf{q})$ given by the alternative expressions described in Sec. III C. As we have shown (see Appendix H), the two expressions for $\chi_\omega(\mathbf{q})$ are equivalent, so that the dual boson approach resolves this ambiguity.

The resulting polarization $\Pi_\omega(\mathbf{q}) = [-\chi_\omega^{-1}(\mathbf{q}) - W_\omega]^{-1}$ depends on momentum, in contrast to EDMFT. It can be proven that this approach yields a gauge-invariant response in the long-wavelength limit [6]. Let us now show that this approach indeed describes plasmons.

A. Results

In this section, we present results for the three-dimensional extended Hubbard model with the infinite-range potential (38) of strength V. We use a momentum-space discretization of $32 \times 32 \times 32$ k points.[8] All results are for temperature $T = 0.02$.

[8]Momentum-dependent quantities including the vertex function are calculated on the irreducible part of the Brillouin zone only and the vertex function is stored for a single momentum at a time.

To begin with, we compare the physical content of the two approximations for the polarization operator which enter Eqs. (40) and (46), respectively. To this end, we examine the *inverse dielectric function*

$$\epsilon_\omega^{-1}(\mathbf{q}) = 1 + V(\mathbf{q})X_\omega(\mathbf{q}). \tag{47}$$

Here, $X_\omega(\mathbf{q})$ is analytically continued to the real axis using Padé approximants. The dielectric function is experimentally accessible through electron energy-loss spectroscopy (EELS). In Fig. 15, we plot the inverse dielectric function for small momenta and on real frequencies. As mentioned before, the energy of the collective mode diverges in the long-wavelength limit in EDMFT (left part of the figure). Including vertex corrections into the polarization corrects this unphysical result and we observe a plasmon mode at finite energy in the long-wavelength limit (right panel). The dispersion of this mode is roughly consistent with a q^2 behavior.

We can study the dependence of this mode on the interaction strength. In Fig. 16, we plot the inverse dielectric function for different values of the local interaction U but fixed strength of the long-range potential set to $V = 0.5$. We see that with increasing onsite interaction, the mode shifts to lower energies, as indicated by the horizontal bars which mark the center of the peak at low q. The spectral weight also decreases with increasing interaction.

B. Plasma frequency

In order to show that the observed mode is indeed the collective plasmon excitation, we can compare the maximum of the spectral intensity to the plasma frequency for this model. The derivation of the plasma frequency is essentially based on gauge invariance. One can obtain the expression either starting from the continuity equation (34) for the response kernel or, alternatively, from the electrical conductivity and using the f-sum rule [44]. Details of the derivation are given in Appendix G. The result is

$$\omega_p^2 = e^2 a^2 \tilde{t} V \mathcal{N}, \tag{48}$$

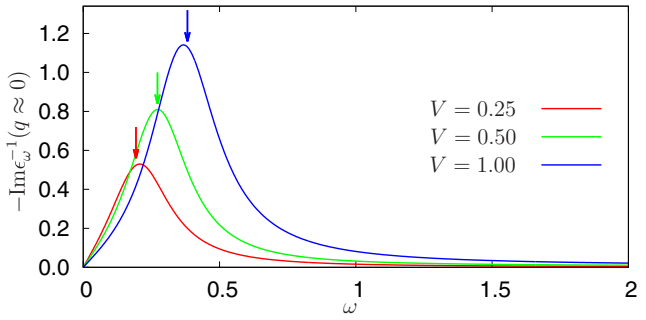

FIG. 17. (Color online) Fixed momentum cuts of the imaginary part of the inverse dielectric function for the smallest momentum $q > 0$ and for $U = 1, T = 0.02$, and different values of V. The arrows indicate the plasma frequency computed from Eq. (48).

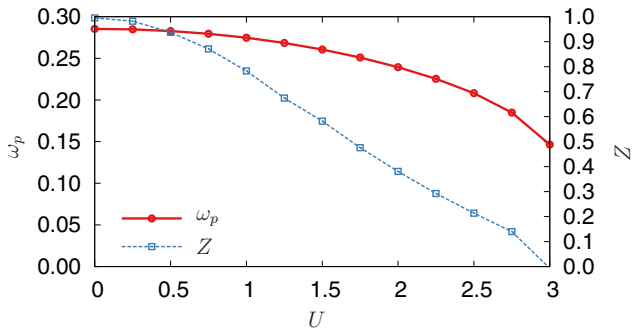

FIG. 18. (Color online) Dependence of the plasma frequency on the local interaction U for fixed $V = 0.5$ and $T = 0.02$. The quasiparticle weight Z is shown for comparison. The plasma frequency remains finite at the Mott transition.

where we have defined

$$\mathcal{N} := \frac{2}{N} \sum_{\mathbf{k}\sigma} \cos(k_z a) \langle n_{\mathbf{k}\sigma} \rangle . \tag{49}$$

Here, we have assumed the field to be oriented along the z axis. In order to rationalize this expression, we note that by linearizing the dispersion in the vicinity of the Fermi level and identifying the coefficient (the velocity) with k/m, we see that $\tilde{t} \sim 1/m$, where m is the bare band mass. Further letting $a = 1$, $V = 4\pi$, and replacing $\mathcal{N}$ with n, it formally takes the same form as the plasma frequency in the continuum $\omega_p^2 = 4\pi n e^2/m$. The appearance of $\mathcal{N}$ instead of the local density is a peculiarity of the lattice model and a consequence of the fact that the electromagnetic potential couples to the bonds rather than to the local charge density (see Appendix B).

In Fig. 17, we show low-momentum cuts of the inverse dielectric function (the data for $V = 0.5$ are the same as in the first panel of Fig. 16). They exhibit a well-defined peak, the position of which is well captured by the expression for the plasma frequency [Eq. (48)]. The agreement is remarkable given that the dielectric function has been obtained by analytical continuation. We emphasize that this coincidence is nontrivial: the position of the peak (the energy of the collective mode) is determined by the two-particle properties of the system, while the plasma frequency is computed from single-particle properties (i.e., the density distribution) only. This relation is a consequence of gauge invariance, which is seen to be fulfilled in our calculation. The connection of single- and two-particle properties is reminiscent of the Ward identity. We further note that the result for the plasma frequency is not restricted to our particular approximation, but applies to any approximation on a discrete lattice which respects gauge invariance (including RPA).

The plasma frequency depends on the local interaction through $\mathcal{N}$. In Fig. 18, we plot the dependence of ω_p on U. The plasma frequency decreases with increasing interaction as observed in Fig. 17. In a simplified picture, the plasma frequency decreases because it is inversely proportional to the square root of the effective mass and the effective mass increases with interaction. According to (49), the frequency decreases because the density distribution becomes less momentum dependent as the interaction increases. In the insulator $\langle n_{\mathbf{k}\sigma} \rangle$ remains momentum dependent and hence the frequency remains finite, but the spectral weight drops to zero.

Clearly, the plasma frequency does not scale with the quasiparticle Z (which also holds for a short-range interaction). This has the important implication that plasmons in strongly correlated systems are beyond Fermi-liquid theory where the quasiparticle contributions are considered dominant [39,40,45]. From the general theory of interacting Fermi systems [39,45] it is known that there are two contributions to the occupation number $\langle n_{\mathbf{k}\sigma} \rangle$: the quasiparticle contribution, which originates from the pole of the electron Green's function and is proportional to Z, and a nonquasiparticle one, which stems from the branch cut of the Green's function and relates to its "incoherent" part. One can clearly see from Fig. 18 that in the vicinity of the metal-insulator transition, the plasma frequency is mostly associated with the incoherent (nonquasiparticle) properties.[9] The plasmons, however, remain well defined as one can see from Figs. 16 and 17.

In Fig. 19, we finally plot the doping dependence of the plasma frequency. ω_p is seen to decrease with doping but only appreciably so for an almost empty (filled) band. For

[9]Within DMFT, the MIT is first order [1] so that strictly speaking the point $Z = 0$ cannot be reached, but Z near the transition is small so our conclusion remains valid.

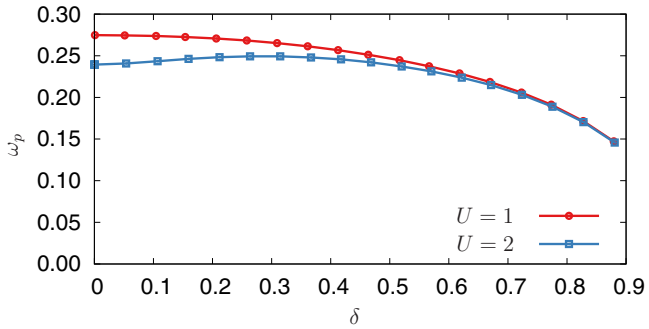

FIG. 19. (Color online) Doping (δ) dependence of the plasma frequency fixed interaction $V = 0.5$, $T = 0.02$, and two values of U.

sufficiently large U there appears to be a shallow maximum in the doping dependence of the plasma frequency.

V. CONCLUSIONS

In this paper, we have addressed the collective charge excitations of strongly correlated electrons in presence of short- and long-range interactions and discussed the relation between gauge invariance and vertex corrections. Nonlocal vertex corrections have been shown to be essential for a qualitatively correct description of the collective modes in both cases. Both the zero-sound mode in the case of a short-range interaction and the plasmon mode emerge through an RPA-like mechanism and are present up to the Mott transition. Our results emphasize the importance of including vertex corrections from a fully frequency-dependent irreducible vertex when working with a frequency-dependent self-energy. Respecting gauge invariance is necessary in order to obtain a proper description of the collective modes in correlated media.

On the technical side, we have proven that the DMFT susceptibility including vertex corrections yields a gauge-invariant charge response in finite dimensions. We have further shown that an alternative expression for the susceptibility that emerges in the dual boson approach is equivalent to the DMFT susceptibility. Such a formulation has the advantage that it circumvents numerical problems due to a divergence of the irreducible vertex close to a metal-insulator transition. It also resolves the ambiguity of calculating the susceptibility in EDMFT. The approach is straightforwardly generalized to treat spin excitations.

ACKNOWLEDGMENTS

We thank T. Ayral, S. Brener, S. Biermann, J. Otsuki, and A. Rubtsov for valuable discussions. H.H. would further like to thank F. Gélis for useful discussions on lattice gauge theory. We also thank L. Boehnke for providing his maximum entropy code. E.G.C.P.v.L. and M.I.K. acknowledge support from ERC Advanced Grant No. 338957 FEMTO/NANO, A.I.L. from the DFG (Grant No. FOR1346), and H.H. and O.P. from the ERC Grant No. 278472–*MottMetals*. This research used high-performance computing resources of GENCI-CCRT under Grant No. t2014056112. The simulations employed a modified version of an open source implementation of the hybridization-expansion continuous-time quantum impurity solver [46], based on the ALPS libraries [47].

APPENDIX A: ANALYTICAL CONTINUATION

The analytical continuation requires accurate input data. For solving the quantum impurity model, we utilize the numerically exact hybridization-expansion continuous-time quantum Monte Carlo method [25], which can treat a local retarded interaction without approximation [48,49]. To maximize accuracy, we employ improved measurements for the local susceptibility χ, self-energy, three-leg vertex λ, and four-leg vertex function γ in the simulation [50].

For the analytical continuation of the imaginary-time data itself, we use a straightforward implementation of the Padé algorithm as presented in Ref. [51]. Because the analytical continuation is an ill-posed mathematical problem, we performed consistency checks. We compared the analytical continuation of the local part of the susceptibility to the local part of the analytically continued result with good agreement for all values of U (the momentum sum and analytical continuation should commute if the latter was exact). As an additional check for the Padé algorithm, we numerically integrated the imaginary part of the analytically continued susceptibility to verify that the Kramers-Kronig relation

$$\chi'(\omega) = -\mathcal{P}\int_{-\infty}^{\infty} \frac{d\omega'}{\pi} \frac{\chi''(\omega')}{\omega - \omega'} \tag{A1}$$

is well fulfilled for $\omega = 0$. In the calculations, the vertex function was determined for up to 128 fermionic Matsubara frequencies (including positive and negative), which corresponds to a cutoff of four times the bandwidth. We took the same number of positive bosonic frequencies and checked that the results do not change appreciably for a smaller number of frequencies. For the single-particle Green's function and self-energy, we took a larger cutoff of 192 frequencies, which is sufficient at this temperature. The analytical continuation of the Matsubara data turns out to be robust when varying the number of input frequencies. Despite these checks, we cannot exclude a qualitative deviation from the real spectra, such as a splitting into multiple peaks. The analytical continuation by Padé approximants tends to give a single-peak spectrum when used in conjunction with data afflicted with statistical errors [52]. For the long-wavelength excitations we are mainly interested in, the Padé results are verified independently: by computing the dispersion from a fit of the Matsubara data in the case of short-range forces (Sec. III D) and by evaluating the plasma frequency from the density in the case of long-range interaction (Sec. IV A).

APPENDIX B: CURRENT OPERATOR

In order to discuss local charge conservation, we require a gauge theory on the lattice. While it can be formulated more generally, it is sufficient for our purposes to consider the case of weak and slowly varying fields.[10] We introduce the coupling of the Hamiltonian (1) to a vector potential via the Peierls substitution [53]:

$$\hat{T} = -\tilde{t}\sum_{\mathbf{r}\boldsymbol{\delta}\sigma} c^{\dagger}_{\mathbf{r}\sigma} e^{ie\mathbf{A}_{\mathbf{r}}\boldsymbol{\delta}} c_{\mathbf{r}-\boldsymbol{\delta}\sigma} + c^{\dagger}_{\mathbf{r}-\boldsymbol{\delta}\sigma} e^{-ie\mathbf{A}_{\mathbf{r}}\boldsymbol{\delta}} c_{\mathbf{r}\sigma}. \tag{B1}$$

For a discussion on the validity of the Peierls substitution and the consequences of the above assumptions, see Refs. [36,54]. The coupling only affects the kinetic energy $\hat{T}$. Under a gauge transformation $\mathbf{A}_{\mathbf{r}}\boldsymbol{\delta} \to \mathbf{A}_{\mathbf{r}}\boldsymbol{\delta} + \Lambda_{\mathbf{r}-\boldsymbol{\delta}} - \Lambda_{\mathbf{r}}$, $c^{\dagger}_{\mathbf{r}} \to c^{\dagger}_{\mathbf{r}} e^{i\Lambda_{\mathbf{r}}}$, $c_{\mathbf{r}} \to c_{\mathbf{r}} e^{-i\Lambda_{\mathbf{r}}}$, the Hamiltonian remains invariant. The current is determined in the usual way as the functional derivative $\mathbf{j}_{\mathbf{r}} = -\delta H/\delta \mathbf{A}_{\mathbf{r}}$. Within linear response, the exponential is expanded up to second order in the vector potential. For the

[10] By slowly varying we mean that the vector potential does not change appreciably over an interatomic distance, so that we can write $\int_{\mathbf{r}}^{\mathbf{r}+\boldsymbol{\delta}} \mathbf{A}\, d\mathbf{r} \approx \mathbf{A}_{\mathbf{r}}\boldsymbol{\delta}$. We only discuss quantities in absence of external fields, i.e., in the limit $\mathbf{A} \to \mathbf{0}$.

current, we obtain

$$\mathbf{j}_\mathbf{r} = ie\tilde{t}\sum_{\delta\sigma}(c^\dagger_{\mathbf{r}\sigma}c_{\mathbf{r}-\delta\sigma} - c^\dagger_{\mathbf{r}-\delta\sigma}c_{\mathbf{r}\sigma})\boldsymbol{\delta} - e^2\tilde{t}\sum_{\delta\sigma}(c^\dagger_{\mathbf{r}\sigma}(\mathbf{A}_\mathbf{r}\boldsymbol{\delta})c_{\mathbf{r}-\delta\sigma} + c^\dagger_{\mathbf{r}-\delta\sigma}(\mathbf{A}_\mathbf{r}\boldsymbol{\delta})c_{\mathbf{r}\sigma})\boldsymbol{\delta}. \quad \text{(B2)}$$

The first term is the paramagnetic current $\mathbf{j}^p_\mathbf{r}$ and the second is the diamagnetic contribution $\mathbf{j}^d_\mathbf{r}$. The momentum representation of the paramagnetic current is

$$\mathbf{j}^p_\mathbf{q} = \sum_\mathbf{r}\mathbf{j}^p_\mathbf{r}e^{-i\mathbf{q}\mathbf{r}} = \frac{ie\tilde{t}}{N}\sum_{\mathbf{k}\delta\sigma}c^\dagger_{\mathbf{k}\sigma}(e^{-i(\mathbf{k}+\mathbf{q})\delta} - e^{i\mathbf{k}\delta})\boldsymbol{\delta}\, c_{\mathbf{k}+\mathbf{q}\sigma}. \quad \text{(B3)}$$

The individual spatial components can be written

$$[\mathbf{j}^p_\mathbf{q}]_\alpha = \frac{ie\tilde{t}a}{N}\sum_{\mathbf{k}\sigma}c^\dagger_{\mathbf{k}\sigma}(e^{-i(k_\alpha+q_\alpha)a} - e^{ik_\alpha a})c_{\mathbf{k}+\mathbf{q}\sigma}. \quad \text{(B4)}$$

This expression can be cast into in the symmetrical form

$$[\mathbf{j}^p_\mathbf{q}]_\alpha = \frac{ie\tilde{t}a}{N}\sum_{\mathbf{k}\sigma}c^\dagger_{\mathbf{k}-\mathbf{q}/2\sigma}(e^{-ik_\alpha a} - e^{ik_\alpha a})e^{-iq_\alpha/2}c_{\mathbf{k}+\mathbf{q}/2\sigma}. \quad \text{(B5)}$$

Symbolically, this can be expressed in terms of a derivative of the dispersion

$$\epsilon_\mathbf{k} = -\tilde{t}\sum_\alpha(e^{-ik_\alpha a} + e^{ik_\alpha a}) \quad \text{(B6)}$$

in the form

$$[\mathbf{j}^p_\mathbf{q}]_\alpha = \tfrac{e}{N}\textstyle\sum_{\mathbf{k}\sigma}c^\dagger_{\mathbf{k}-\mathbf{q}/2\sigma}\frac{\partial\varepsilon_\mathbf{k}}{\partial k_\alpha}e^{-iq_\alpha a/2}c_{\mathbf{k}+\mathbf{q}/2\sigma}. \quad \text{(B7)}$$

In the long-wavelength ($q \to 0$) limit, this reduces to

$$\mathbf{j}^p = \tfrac{e}{N}\textstyle\sum_{\mathbf{k}\sigma}c^\dagger_{\mathbf{k}\sigma}(\boldsymbol{\nabla}\varepsilon_\mathbf{k})c_{\mathbf{k}\sigma}. \quad \text{(B8)}$$

For completeness, we provide the result for the diamagnetic contribution to the current. In momentum space, we obtain

$$\begin{aligned}[\mathbf{j}^d_\mathbf{q}]_\alpha &= -\frac{e^2a^2\tilde{t}}{N^2}\sum_{\mathbf{k}\mathbf{k}'\sigma}c^\dagger_{\mathbf{k}\sigma}A^\alpha(\mathbf{k}-\mathbf{k}')[e^{-i(k'_\alpha+q_\alpha)a} + e^{ik_\alpha a}]c_{\mathbf{k}'+\mathbf{q}\sigma}\\ &= -\frac{e^2a^2\tilde{t}}{N^2}\sum_{\mathbf{k}\mathbf{q}'\sigma}c^\dagger_{\mathbf{k}\sigma}A^\alpha(\mathbf{q}')[e^{-i(k_\alpha+q_\alpha-q'_\alpha)a} + e^{ik_\alpha a}]c_{\mathbf{k}'+\mathbf{q}-\mathbf{q}'\sigma}\\ &= -\frac{e^2a^2\tilde{t}}{N^2}\sum_{\mathbf{k}\mathbf{q}'\sigma}c^\dagger_{\mathbf{k}-\mathbf{q}/2+\mathbf{q}'/2\sigma}A^\alpha(\mathbf{q}')[e^{-ik_\alpha a} + e^{ik_\alpha a}]\\ &\quad\times e^{-i(q_\alpha/2-q'_\alpha/2)a}c_{\mathbf{k}+\mathbf{q}/2-\mathbf{q}'/2\sigma}. \end{aligned} \quad \text{(B9)}$$

Similarly as for the paramagnetic current, we can rewrite this as

$$[j^d_\mathbf{q}]_\alpha = -\frac{e^2}{N^2}\sum_{\mathbf{k}\mathbf{q}'\sigma}c^\dagger_{\mathbf{k}-\mathbf{q}/2+\mathbf{q}'/2\sigma}A^\alpha(\mathbf{q}')\left[\sum_\beta\frac{\partial^2\varepsilon_\mathbf{k}}{\partial k_\alpha\partial k_\beta}\right]\times e^{-i(q_\alpha/2-q'_\alpha/2)a}c_{\mathbf{k}+\mathbf{q}/2-\mathbf{q}'/2\sigma}. \quad \text{(B10)}$$

In the following, we are interested in the continuity equation in absence of an electromagnetic field and will henceforth focus on the paramagnetic contribution to the current. We introduce the four-vector notation $j^\mu_\mathbf{q} = (j^0_\mathbf{q}, \mathbf{j}^\alpha_\mathbf{q})$, with $j^0_\mathbf{q} = en_\mathbf{q}$ and define the bare current vertex through

$$j^\mu_\mathbf{q} = \frac{e}{N}\sum_{\mathbf{k}\sigma}c^\dagger_{\mathbf{k}\sigma}\gamma^\mu(\mathbf{k},\mathbf{q})c_{\mathbf{k}+\mathbf{q}\sigma}. \quad \text{(B11)}$$

Using (B4), we obtain

$$\gamma^\mu = \begin{cases} i\tilde{t}a\left(e^{-i(k_\alpha+q_\alpha)a} - e^{ik_\alpha a}\right), & \mu = \alpha = x,y,z \\ 1, & \mu = 0. \end{cases} \quad \text{(B12)}$$

APPENDIX C: CONTINUITY EQUATION

The Hamiltonian including the coupling to the electromagnetic field fulfills the continuity equation

$$\begin{aligned} e\frac{\partial n_\mathbf{r}}{\partial t} &= -ie[n_\mathbf{r}, H]\\ &= ie\tilde{t}\sum_{\delta\sigma}(c^\dagger_{\mathbf{r}\sigma}c_{\mathbf{r}+\delta\sigma} + c^\dagger_{\mathbf{r}\sigma}c_{\mathbf{r}-\delta\sigma} - c^\dagger_{\mathbf{r}+\delta\sigma}c_{\mathbf{r}\sigma} - c^\dagger_{\mathbf{r}-\delta\sigma}c_{\mathbf{r}\sigma})\\ &\quad - e^2\tilde{t}\sum_{\delta\sigma}(c^\dagger_{\mathbf{r}\sigma}(\mathbf{A}_\mathbf{r}\boldsymbol{\delta})c_{\mathbf{r}+\delta\sigma} + c^\dagger_{\mathbf{r}\sigma}(\mathbf{A}_\mathbf{r}\boldsymbol{\delta})c_{\mathbf{r}-\delta\sigma}\\ &\quad - c^\dagger_{\mathbf{r}+\delta\sigma}(\mathbf{A}_\mathbf{r}\boldsymbol{\delta})c_{\mathbf{r}\sigma} - c^\dagger_{\mathbf{r}-\delta\sigma}(\mathbf{A}_\mathbf{r}\boldsymbol{\delta})c_{\mathbf{r}\sigma}). \end{aligned} \quad \text{(C1)}$$

We can write the right-hand side as a divergence of the current operator (B2). To this end, we define the *forward derivative* of the current

$$\nabla^F\cdot\mathbf{j}_\mathbf{r} := \frac{\mathbf{j}_{\mathbf{r}+\delta} - \mathbf{j}_\mathbf{r}}{a}, \quad \text{(C2)}$$

which should be understood such that the finite difference of the x direction of the current is formed by displacement δ in the x direction, etc. Using Eqs. (C1) and (B2), it is easy to see that the continuity equation can then be written in the form

$$e\frac{\partial n_\mathbf{r}}{\partial t} + \nabla^F\cdot\mathbf{j}_\mathbf{r} = 0. \quad \text{(C3)}$$

In order to recast this equation into momentum space, one defines $\mathbf{q}^F$ as the eigenvalue of the operator $\nabla^F = \sum_\alpha\partial^F_\alpha\mathbf{e}_\alpha$ acting on a plane wave $\phi_\mathbf{q}(\mathbf{r}) \sim e^{i\mathbf{q}\mathbf{r}}$. Denoting $q_\alpha = \mathbf{q}_\alpha\mathbf{e}_\alpha$ as the α component of $\mathbf{q}$, which takes the discrete values $q^{(n)}_\alpha = 2\pi in/(Na)$ and correspondingly $r_\alpha = \mathbf{r}\mathbf{e}_\alpha, r_{l_\alpha} = al_\alpha$, we have

$$\begin{aligned}\nabla^F e^{i\mathbf{q}\mathbf{r}} &= \frac{1}{a}\sum_\alpha(e^{iq_\alpha(l_\alpha+1)a} - e^{iq_\alpha l_\alpha a})\mathbf{e}_\alpha\\ &= \frac{1}{a}\sum_\alpha(e^{iq_\alpha a} - 1)e^{iq_\alpha l_\alpha a} \equiv i\mathbf{q}^F e^{i\mathbf{q}\mathbf{r}} \end{aligned} \quad \text{(C4)}$$

so that

$$\mathbf{q}^F = \sum_\alpha q^F_{k_\alpha}\mathbf{e}_\alpha = -\frac{i}{a}\sum_\alpha(e^{iq_\alpha a} - 1)\mathbf{e}_\alpha. \quad \text{(C5)}$$

In the long-wavelength limit, this simplifies to

$$\mathbf{q}^F \approx -\frac{i}{a}\sum_\alpha(iq_\alpha a)\,\mathbf{e}_\alpha \equiv \mathbf{q}. \quad \text{(C6)}$$

We note that by defining the *backward derivative* as $\nabla^B\cdot\mathbf{j}_\mathbf{r} := (\mathbf{j}_{\mathbf{r}-\delta} - \mathbf{j}_\mathbf{r})/a$, we obtain $\mathbf{q}^B = (\mathbf{q}^F)^*$. The dispersion can be expressed in terms of the product $\mathbf{q}^F\cdot\mathbf{q}^B$.

APPENDIX D: WARD IDENTITY

The derivation of the Ward identity can be found in textbooks (see, e.g., Ref. [55]). Here, we sketch the derivation as required for the subsequent discussion. It is convenient to use four-vector notation throughout.

With $k=(i\omega,\mathbf{k})$, $z=(\tau,\mathbf{r})$, $\partial_\mu^F=(\partial_\tau,\nabla^F)$, $q_\mu^F=(i\omega,\mathbf{q}^F)$, and the metric $(-1,1,1,1)$, we have $k_\mu z_\mu=\mathbf{kr}-i\omega\tau$ and the continuity equation becomes $\partial_\mu^F j_\mu(z)=0$. The Ward identity is obtained by applying the four-divergence to the three-leg correlation function

$$\Lambda_\mu(x,y,z)=\langle T_\tau c(x)c^\dagger(y)j_\mu(z)\rangle. \tag{D1}$$

We have

$$\begin{aligned}\partial_\mu^F\Lambda_\mu=&\left\langle T_\tau c(x)c^\dagger(y)\left[\partial_{z_\alpha}^F j_\alpha+\partial_{z_0}j_0\right]\right\rangle\\&+\langle T_\tau c(x)[j_0(z),c^\dagger(y)]\delta(y_0-z_0)\rangle\\&+\langle T_\tau c^\dagger(y)[c(x),j_0(z)]\delta(x_0-z_0)\rangle,\end{aligned} \tag{D2}$$

where the term in angular brackets in the first line vanishes by virtue of the continuity equation and the two other lines emerge due to time differentiation while accounting for time ordering. With $j_0\equiv en$ the commutators are

$$\begin{aligned}[j_0(z),c^\dagger(y)]\delta(y_0-z_0)&=ec^\dagger(y)\delta(y-z),\\[c(x),j_0(z)]\delta(x_0-z_0)&=ec(x)\delta(x-z).\end{aligned} \tag{D3}$$

Inserting this back into (D2) and using the definition of Green's function $G(x-y):=-\langle T_\tau c(x)c^\dagger(y)\rangle$ yields

$$\partial_\mu^F\Lambda_\mu=e[\delta(x-z)-\delta(y-z)]G(x-y). \tag{D4}$$

Noting that Λ_μ can be expressed in terms of the generalized susceptibility (see following), this equation is recognized as the lattice formulation of Eq. (13) in Ref. [30]. Defining the current vertex Γ_μ through

$$\Lambda_\mu(x,y,z)=e\int dx'\int dy' G(x-x')\Gamma_\mu(x',y',z)G(y'-y) \tag{D5}$$

and introducing the lattice Fourier transform $G(k)=\int dx\, e^{-ikx}G(x)$, where $\int dx:=\sum_{\mathbf{x}}\int_0^\beta d\tau$, one straightforwardly obtains the momentum-space representation of (D4):

$$G(k)q_\mu^F\Gamma_\mu(k,q)G(k+q)=G(k+q)-G(k). \tag{D6}$$

It is commonly written in the form

$$q_\mu^F\Gamma_\mu(k,q)=G^{-1}(k)-G^{-1}(k+q), \tag{D7}$$

which is the Ward identity.

APPENDIX E: CURRENT VERTEX

1. Noninteracting case

In the noninteracting case, the bare current vertex (B12) has to fulfill the Ward identity

$$q_\mu^F\gamma_\mu(k,q)=G_0^{-1}(k)-G_0^{-1}(k+q). \tag{E1}$$

In this case, we have $G_0^{-1}(k)=k_0-\varepsilon_{\mathbf{k}}$ so the right-hand side of (26) becomes $\varepsilon_{\mathbf{k+q}}-\varepsilon_{\mathbf{k}}-q_0$, while the left-hand side reads

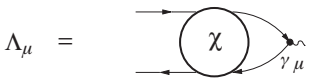

FIG. 20. Diagrammatic representation of the three-leg correlation function Λ_μ in terms of the generalized susceptibility.

as $q_\alpha^F\gamma_\alpha-q_0$. Using (B6) and (C5), we see that indeed

$$\begin{aligned}q_\alpha^F\gamma_\alpha&=\tilde{t}\sum_\alpha(e^{iq_\alpha a}-1)(e^{-i(k_\alpha+q_\alpha)a}-e^{ik_\alpha a})\\&=\tilde{t}\sum_\alpha(e^{-ik_\alpha a}-e^{i(k_\alpha+q_\alpha)a}-e^{-i(k_\alpha+q_\alpha)a}+e^{-ik_\alpha a})\\&=\varepsilon_{\mathbf{k+q}}-\varepsilon_{\mathbf{k}}.\end{aligned} \tag{E2}$$

2. Interacting case

In the interacting case, we have

$$\begin{aligned}\Lambda_\mu^\sigma(k,q)&=\langle c_{k\sigma}c_{k+q,\sigma}^\dagger j_q^\mu\rangle\\&=-\frac{e}{N}\sum_{k'\sigma'}\langle c_{k\sigma}c_{k+q,\sigma}^\dagger c_{k'\sigma'}c_{k'+q,\sigma'}^\dagger\rangle\gamma^\mu(k',q),\end{aligned} \tag{E3}$$

where the correlation function in the second line is the generalized susceptibility (see Fig. 20)

$$\Lambda_\mu^\sigma(k,q)=-\frac{e}{N}\sum_{k'\sigma'}\chi_{kk'q}^{\sigma\sigma'}\gamma^\mu(k',q). \tag{E4}$$

The latter can be expressed in terms of the vertex function as follows:

$$\begin{aligned}\chi_{kk'q}^{\sigma\sigma'}=&-\frac{1}{T}G_{k\sigma}G_{k+q\sigma}\delta_{kk'}\delta_{\sigma\sigma'}\\&+G_{k\sigma}G_{k+q\sigma}\Gamma_{kk'q}^{\sigma\sigma'}G_{k'\sigma'}G_{k'+q\sigma'}.\end{aligned} \tag{E5}$$

Inserting (E5) into (E4) and using the definition of the current vertex $\Lambda_\mu(k,q)=eG_k\Gamma_\mu(k,q)G_{k+q}$, one obtains

$$\Gamma_\mu^\sigma(k,q)=\gamma_\mu(k,q)-\frac{T}{N}\sum_{k'\sigma'}\Gamma_{kk'q}^{\sigma\sigma'}G_{k'\sigma'}G_{k'+q,\sigma'}\gamma_\mu(k',q). \tag{E6}$$

Inserting the Bethe-Salpeter equation for the vertex

$$\Gamma_{kk'q}^{\sigma\sigma'}=\Gamma_{kk'q}^{\mathrm{irr}\sigma\sigma'}-\frac{T}{N}\sum_{k''\sigma''}\Gamma_{kk''q}^{\mathrm{irr}\sigma\sigma''}G_{k''\sigma''}G_{k''+q,\sigma''}\Gamma_{k''k'q}^{\sigma''\sigma'} \tag{E7}$$

into (E6), we obtain the ladder equation for the current vertex

$$\Gamma_\mu^\sigma(k,q)=\gamma_\mu(k,q)-\frac{T}{N}\sum_{k'\sigma'}\Gamma_{kk'q}^{\mathrm{irr}\sigma\sigma'}G_{k'\sigma'}G_{k'+q,\sigma'}\Gamma_\mu^{\sigma'}(k',q). \tag{E8}$$

APPENDIX F: ELECTROMAGNETIC RESPONSE KERNEL

Within linear response, the electromagnetic response kernel is defined by

$$J_\mu(q)=K_{\mu\nu}(q)A_\nu(q), \tag{F1}$$

where $J_\mu(q)$ is the expectation value of the current. Demanding the invariance of the kernel under a gauge transformation

$A_\nu(q) \to A_\nu(q) + iq_\nu^F \Lambda(q)$ implies $K_{\mu\nu} q_\nu^F = 0$. On the other hand, the fact that the expectation value J_μ fulfills the continuity equation $q_\mu^F J_\mu(q) = 0$ implies the condition $q_\mu^F K_{\mu\nu}(q) = 0$ imposed by charge conservation.

Following the standard derivation, i.e., by expressing the kernel as a functional derivative of the current with respect to the vector potential, we obtain the following result for the kernel on the discrete lattice:

$$\begin{aligned} K_{\mu\nu}(\mathbf{r},\tau;\mathbf{r}',\tau') = {} & \langle T_\tau j_\mu(\mathbf{r},\tau) j_\nu(\mathbf{r}',\tau')\rangle \\ & - e^2 a^2 \tilde{t} \delta_{\mu\nu}(1-\delta_{0\mu})\delta(\tau-\tau')\delta_{\mathbf{r}',\mathbf{r}-\boldsymbol{\delta}} \\ & \times \sum_{\boldsymbol{\delta}\sigma} \langle c^\dagger_{\mathbf{r}\sigma\tau} c_{\mathbf{r}-\boldsymbol{\delta}\sigma\tau'} + c^\dagger_{\mathbf{r}-\boldsymbol{\delta}\sigma\tau} c_{\mathbf{r}\sigma\tau'}\rangle, \end{aligned} \tag{F2}$$

where the second term originates from the derivative of the diamagnetic current [see Eq. (B2)] with respect to the vector potential and expectation values are taken in the absence of an external field. Note that the field does not couple directly to the local density $n_\mathbf{r} \equiv \sum_\sigma \langle c^\dagger_{\mathbf{r}\sigma\tau} c_{\mathbf{r}\sigma\tau}\rangle$ because the electromagnetic potential is a link variable. In momentum space, the kernel becomes

$$\begin{aligned} K_{\mu\nu}(\mathbf{q},i\omega) = {} & \langle T_\tau j_\mu(\mathbf{q},i\omega) j_\nu(-\mathbf{q},-i\omega)\rangle \\ & - e^2 a^2 \tilde{t} \delta_{\mu\nu}(1-\delta_{0\nu})\mathcal{N}, \end{aligned} \tag{F3}$$

where for simplicity we have defined

$$\mathcal{N} := \frac{1}{N}\sum_{\mathbf{k}\sigma} (e^{-ik_\nu a} + e^{ik_\nu a}) \langle n_{\mathbf{k}\sigma}\rangle . \tag{F4}$$

In order to show that the exact kernel obeys the constraint imposed by charge conservation, we form the expression

$$\begin{aligned} q_\mu^F K_{\mu\nu}(\mathbf{q},i\omega) &= \left\langle T_\tau \left[q_\mu^F j_\mu(\mathbf{q},i\omega)\right] j_\nu(-\mathbf{q},-i\omega)\right\rangle + \langle [j_0(\mathbf{q}), j_\nu(-\mathbf{q})]\rangle - e^2 a^2 \tilde{t}(1-\delta_{0\nu}) q_\nu^F \frac{1}{N}\sum_{\mathbf{k}\sigma} (e^{-ik_\nu a} + e^{ik_\nu a}) \langle n_{\mathbf{k}\sigma}\rangle \\ &= \langle [j_0(\mathbf{q}), j_\nu(-\mathbf{q})]\rangle + ie^2 a \tilde{t}(1-\delta_{0\nu}) \frac{1}{N}\sum_{\mathbf{k}\sigma} (e^{-i(k_\nu - q_\nu)a} - e^{-ik_\nu a} + e^{i(k_\nu+q_\nu)a} - e^{ik_\nu a}) \langle n_{\mathbf{k}\sigma}\rangle , \end{aligned} \tag{F5}$$

where $\langle n_{\mathbf{k}\sigma}\rangle = \langle c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma}\rangle$. The first term on the right-hand side vanishes for the exact kernel because of the continuity equation $q_\mu^F j_\mu = 0$. To obtain the last line, the explicit expression (C5) was substituted for $\mathbf{q}^F$. The commutator arises because the time derivative does not commute with the time-ordering symbol. It evaluates to

$$\begin{aligned} \langle [j_0(\mathbf{q}), j_\nu(-\mathbf{q})]\rangle &= e^2(1-\delta_{0\nu}) \frac{1}{N}\sum_{\mathbf{k}\sigma} [\gamma_\nu(\mathbf{k}+\mathbf{q},-\mathbf{q}) - \gamma_\nu(\mathbf{k},-\mathbf{q})] \langle n_{\mathbf{k}\sigma}\rangle \\ &= ie^2 a \tilde{t}(1-\delta_{0\nu}) \frac{1}{N}\sum_{\mathbf{k}\sigma} (e^{-ik_\nu a} - e^{i(k_\nu+q_\nu)a} - e^{-i(k_\nu-q_\nu)a} + e^{ik_\nu}) \langle n_{\mathbf{k}\sigma}\rangle , \end{aligned} \tag{F6}$$

so that the terms on the right-hand side of (F5) cancel, leading to the required result $q_\mu^F K_{\mu\nu}(q) = 0$. In the same way, one shows that the kernel is gauge invariant, i.e., $K_{\mu\nu}(q) q_\nu^F = 0$.

APPENDIX G: PLASMA FREQUENCY

The plasma frequency is determined by the uniform response, i.e., in the limit $q_\alpha \to 0$. For small momenta, we replace $q_\alpha^F \to q_\alpha$. From the gauge-invariance condition $K_{\mu\nu} q_\nu^F = 0$ we have $K_{\mu 0} = K_{\mu\alpha} q_\alpha / q_0$. We can take the direction of the field parallel to the z axis. The transverse response vanishes by symmetry. Hence, we have $K_{00} = K_{0z} q/q_0$, where $q \equiv q_z$. Similarly, we obtain $K_{0z} = K_{zz} q/q_0$, which yields the continuity equation

$$K_{00} = \frac{q^2}{q_0^2} K_{zz}. \tag{G1}$$

Now, we use that the same relation holds when the density-density and current-current correlation functions in the kernel, which are reducible in the interaction, are replaced by the corresponding quantities irreducible in the interaction $V(\mathbf{q})$ (see, e.g., Ref. [39]). This is possible because the former are related to the latter through simple geometric series. We denote the resulting quantities $\tilde{K}_{00}$ and $\tilde{K}_{zz}$. Hence, we can identify $\tilde{K}_{00}$ with $e^2\Pi$ where Π contains all polarization diagrams irreducible with respect to $V(\mathbf{q})$. In the long-wavelength limit, only the diamagnetic term in $\tilde{K}_{zz}$ contributes. With the Coulomb interaction $V(\mathbf{q}) = e^2 V/q^2$ for $q \neq 0$, we have in the limit $q_\alpha \to 0$ and for $q_0 \to \omega + i0^+$,

$$\epsilon_\omega = \lim_{q\to 0} 1 + V(\mathbf{q})\Pi_\omega(\mathbf{q}) = 1 - \omega_p^2/\omega^2 \tag{G2}$$

with the plasma frequency

$$\omega_p = ea\sqrt{\tilde{t} V \mathcal{N}}. \tag{G3}$$

Alternatively, the same result may be obtained by relating the response kernel to the conductivity and using the f-sum rule.

APPENDIX H: EQUIVALENCE OF EQS. (11) AND (15)

We would like to establish the equivalence between the usual expression for the susceptibility

$$\chi_\omega(\mathbf{q}) = 2T\sum_\nu \chi^0_{\nu\omega}(\mathbf{q}) - 2T^2 \sum_{\nu\nu'} \chi^0_{\nu\omega}(\mathbf{q}) \Gamma_{\nu\nu'\omega}(\mathbf{q}) \chi^0_{\nu'\omega}(\mathbf{q}) \tag{H1}$$

and the alternative form given by the equations (see Fig. 21)

$$\chi_\omega(\mathbf{q}) = \chi_\omega + \chi_{-\omega} 2T^2 \sum_{\nu\nu'} \lambda_{\nu+\omega,-\omega} \tilde{\chi}_{\nu\nu'\omega}(\mathbf{q}) \lambda_{\nu'\omega} \chi_\omega, \tag{H2}$$

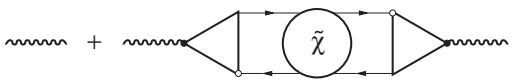

FIG. 21. Diagrammatic representation of Eq. (H2).

$$\tilde{\chi}_{\nu\nu'\omega}(\mathbf{q}) = \frac{1}{T}\tilde{\chi}^0_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'} - \tilde{\chi}^0_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\tilde{\chi}^0_{\nu'\omega}(\mathbf{q}). \tag{H3}$$

For simplicity, we consider the paramagnetic case and the charge susceptibility only. Correspondingly, spin labels are omitted and the vertex functions are taken in the charge channel (e.g., $\Gamma^{\mathrm{ch}} = \Gamma^{\uparrow\uparrow} + \Gamma^{\uparrow\downarrow}$). The label "ch" is also suppressed in the following. We further use the following definitions:

$$\chi^0_{\nu\omega}(\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}} G_\nu(\mathbf{k})G_{\nu+\omega}(\mathbf{k}+\mathbf{q}), \tag{H4}$$

$$\tilde{\chi}^0_{\nu\omega}(\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}} \tilde{G}_\nu(\mathbf{k})\tilde{G}_{\nu+\omega}(\mathbf{k}+\mathbf{q}), \tag{H5}$$

$$\chi^0_{\nu\omega} = g_\nu g_{\nu+\omega}, \tag{H6}$$

$$\tilde{G}_\nu(\mathbf{k}) = G_\nu(\mathbf{k}) - g_\nu, \tag{H7}$$

where G is the lattice Green's function and g is the impurity Green's function. The impurity Green's function, charge susceptibility, as well as of the three-leg charge vertex and the four-leg vertex of the impurity are defined in terms of impurity correlation functions as follows:

$$g_{\nu\sigma} := -\langle c_{\nu\sigma} c^*_{\nu\sigma}\rangle, \tag{H8}$$

$$\chi_\omega := -(\langle n_\omega n_{-\omega}\rangle - \langle n\rangle\langle n\rangle\delta_\omega), \tag{H9}$$

$$\lambda^\sigma_{\nu\omega} := \frac{g^{\sigma(3)}_{\nu\omega} - g_{\nu\sigma}\langle n\rangle\delta_\omega/T}{g_{\nu\sigma}g_{\nu+\omega,\sigma}\chi_\omega}, \tag{H10}$$

$$\gamma^{\sigma\sigma'}_{\nu\nu'\omega} := \frac{g^{(4)\sigma\sigma'}_{\nu\nu'\omega} - (g_{\nu\sigma}g_{\nu'\sigma'}\delta_\omega - g_{\nu+\omega,\sigma}g_{\nu\sigma}\delta_{\nu\nu'}\delta_{\sigma\sigma'})/T}{g_{\nu\sigma}g_{\nu+\omega,\sigma}g_{\nu'+\omega\sigma'}g_{\nu'\sigma'}}. \tag{H11}$$

The three- and four-point functions $g^{(3)}$ and $g^{(4)}$ in turn are given by the averages

$$g^{(3)\sigma}_{\nu\omega} := -\langle c_{\nu\sigma} c^*_{\nu+\omega,\sigma} n_\omega\rangle, \tag{H12}$$

$$g^{(4)\sigma\sigma'}_{\nu\nu'\omega} := +\langle c_{\nu\sigma} c^*_{\nu+\omega,\sigma} c_{\nu'+\omega,\sigma'} c^*_{\nu'\sigma'}\rangle. \tag{H13}$$

In the paramagnetic state, g, $g^{(3)}$, and λ are independent of spin. Here, γ is the reducible impurity charge vertex.

The vertices λ and γ (see Fig. 22) are closely related [6]. Using $n_\omega = T\sum_{\nu'\sigma'} c^*_{\nu'\sigma'}c_{\nu'+\omega,\sigma'}$, the fact that the Grassmann numbers anticommute, together with the above definitions of the correlation functions, one finds that

$$\begin{aligned} g_{\nu\sigma}g_{\nu+\omega,\sigma}\lambda^\sigma_{\nu\omega}\chi_\omega &= g^{(3)\sigma}_{\nu\omega} - \frac{1}{T}g_{\nu\sigma}\langle n\rangle\delta_\omega \\ &= T\sum_{\nu'\sigma'}\langle c_{\nu\sigma}c^*_{\nu+\omega,\sigma}c_{\nu'+\omega,\sigma'}c^*_{\nu'\sigma'}\rangle + g_{\nu\sigma}\sum_{\nu'\sigma'}\langle c_{\nu'\sigma'}c^*_{\nu'+\omega,\sigma'}\rangle\delta_\omega \\ &= T\sum_{\nu'\sigma'} g^{(4)\sigma\sigma'}_{\nu\nu'\omega} - g_{\nu\sigma}\sum_{\nu'\sigma'} g_{\nu'\sigma'}\delta_\omega. \end{aligned} \tag{H14}$$

Similarly, one obtains

$$\begin{aligned} &g_{\nu\sigma}g_{\nu+\omega,\sigma}T\sum_{\nu'\sigma'}\gamma^{\sigma\sigma'}_{\nu\nu'\omega}g_{\nu'+\omega\sigma'}g_{\nu'\sigma'} \\ &\quad = T\sum_{\nu'\sigma'} g^{(4)\sigma\sigma'}_{\nu\nu'\omega} - g_{\nu\sigma}\sum_{\nu'\sigma'} g_{\nu'\sigma'} + g_{\nu\sigma}g_{\nu+\omega\sigma}. \end{aligned} \tag{H15}$$

Taken together, one finds the relation

$$\begin{aligned} &g_{\nu\sigma}g_{\nu+\omega,\sigma}\left(T\sum_{\nu'\sigma'}\gamma^{\sigma\sigma'}_{\nu\nu'\omega}g_{\nu'+\omega\sigma'}g_{\nu'\sigma'} - 1\right) \\ &\quad = g_{\nu\sigma}g_{\nu+\omega,\sigma}\lambda^\sigma_{\nu\omega}\chi_\omega. \end{aligned} \tag{H16}$$

Further using that

$$n_{-\omega} = T\sum_{\nu'\sigma'} c^*_{\nu'\sigma'}c_{\nu'-\omega,\sigma'} = T\sum_{\nu'\sigma'} c^*_{\nu'+\omega,\sigma'}c_{\nu',\sigma'} \tag{H17}$$

and

$$\langle c_{\nu+\omega,\sigma}c^*_{\nu\sigma}n_{-\omega}\rangle = \langle n_{-\omega}c_{\nu+\omega,\sigma}c^*_{\nu\sigma}\rangle, \tag{H18}$$

we have

$$\begin{aligned} &\chi_{-\omega}\lambda^\sigma_{\nu+\omega,-\omega}g_{\nu+\omega,\sigma}g_{\nu\sigma} \\ &= g^{(3)\sigma}_{\nu+\omega,-\omega} - \frac{1}{T}g_{\nu+\omega\sigma}\langle n\rangle\delta_\omega = T\sum_{\nu'\sigma'}\langle c_{\nu'+\omega,\sigma'}c^*_{\nu'\sigma'}c_{\nu\sigma}c^*_{\nu+\omega,\sigma}\rangle \\ &\quad + g_{\nu+\omega,\sigma}\sum_{\nu'\sigma'}\langle c_{\nu'\sigma'}c^*_{\nu'+\omega,\sigma'}\rangle\delta_\omega \\ &= T\sum_{\nu'\sigma'} g^{(4)\sigma'\sigma}_{\nu'\nu\omega} - g_{\nu+\omega\sigma}\sum_{\nu'\sigma'} g_{\nu'\sigma'}\delta_\omega. \end{aligned} \tag{H19}$$

This result can be expressed in terms of the four-leg vertex similarly to the above. We therefore find the following relations between the three-leg and four-leg vertices:

$$\lambda^\sigma_{\nu\omega}\chi_\omega = T\sum_{\nu'\sigma'}\gamma^{\sigma\sigma'}_{\nu\nu'\omega}\chi^0_{\nu'\omega} - 1, \tag{H20}$$

$$\chi_{-\omega}\lambda^\sigma_{\nu+\omega,-\omega} = T\sum_{\nu'\sigma'}\chi^0_{\nu'\omega}\gamma^{\sigma'\sigma}_{\nu'\nu\omega} - 1. \tag{H21}$$

From Eqs. (H4)–(H7), it is further easy to see that

$$\tilde{\chi}^0_{\nu\omega}(\mathbf{q}) = \chi^0_{\nu\omega}(\mathbf{q}) - \chi^0_{\nu\omega}. \tag{H22}$$

Now recall that the lattice vertex Γ is calculated from the Bethe-Salpeter equation (BSE)

$$\left[\Gamma^{-1}_\omega(\mathbf{q})\right]_{\nu\nu'} = \left[\gamma^{\mathrm{irr}\,-1}_\omega\right]_{\nu\nu'} + T\chi^0_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'}, \tag{H23}$$

where in turn the local irreducible vertex is calculated from the BSE of the impurity

$$\left[\gamma^{-1}_\omega\right]_{\nu\nu'} = \left[\gamma^{\mathrm{irr}\,-1}_\omega\right]_{\nu\nu'} + T\chi^0_{\nu\omega}\delta_{\nu\nu'}. \tag{H24}$$

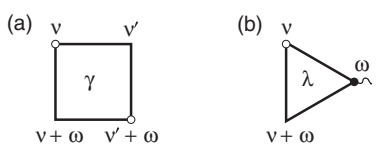

FIG. 22. Definition of the impurity vertex functions.

Combining the BSEs using (H22), one can write

$$\left[\Gamma_{\omega}^{-1}(\mathbf{q})\right]_{\nu\nu'} = \left[\gamma_{\omega}^{-1}\right]_{\nu\nu'} + T\,\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q})\delta_{\nu\nu'}, \tag{H25}$$

from which it follows that

$$\gamma_{\nu\nu'\omega} - \Gamma_{\nu\nu'\omega}(\mathbf{q}) = T\sum_{\nu''}\gamma_{\nu\nu''\omega}\tilde{\chi}^{0}_{\nu''\omega}(\mathbf{q})\Gamma_{\nu''\nu'\omega}(\mathbf{q}) = T\sum_{\nu''}\Gamma_{\nu''\nu'\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu''\omega}(\mathbf{q})\gamma_{\nu\nu''\omega}. \tag{H26}$$

The charge susceptibility can be expressed in terms of the local vertex and Green's functions as

$$\chi_{\omega} = 2T\sum_{\nu}\chi^{0}_{\nu\omega} - 2T^{2}\sum_{\nu\nu'}\chi^{0}_{\nu\omega}\gamma_{\nu\nu'\omega}\chi^{0}_{\nu'\omega}. \tag{H27}$$

Using these relations and inserting (H3) into (H2), one obtains

$$\begin{aligned}\chi_{\omega}(\mathbf{q}) = {} & 2T\sum_{\nu}\left[\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q}) + \chi^{0}_{\nu\omega}\right] - 2T^{2}\sum_{\nu\nu'}\chi^{0}_{\nu\omega}\gamma_{\nu\nu'\omega}\chi^{0}_{\nu'\omega} + 2T^{3}\sum_{\nu\nu'\nu''}\chi^{0}_{\nu\omega}\gamma_{\nu\nu'}\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q})\gamma_{\nu'\nu''}\chi^{0}_{\nu''\omega} \\ & - 2T^{2}\sum_{\nu\nu'}\chi^{0}_{\nu\omega}\gamma_{\nu\nu'}\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q}) - 2T^{2}\sum_{\nu\nu'}\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q})\gamma_{\nu\nu'}\chi^{0}_{\nu'\omega} + 2T^{3}\sum_{\nu\nu'\nu''}\chi^{0}_{\nu\omega}\gamma_{\nu\nu'}\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q})\Gamma_{\nu'\nu''\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu''\omega}(\mathbf{q}) \\ & + 2T^{3}\sum_{\nu\nu'\nu''}\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q})\gamma_{\nu'\nu''}\chi^{0}_{\nu''\omega} - 2T^{4}\sum_{\nu\nu'\nu''\nu'''}\chi^{0}_{\nu\omega}\gamma_{\nu\nu'}\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q})\Gamma_{\nu'\nu''\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu''\omega}(\mathbf{q})\gamma_{\nu''\nu'''}\chi^{0}_{\nu'''\omega} \\ & - 2T^{2}\sum_{\nu\nu'}\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q}).\end{aligned} \tag{H28}$$

Substituting (H26), all terms involving the impurity vertex cancel and one is left with

$$\begin{aligned}\chi_{\omega}(\mathbf{q}) = {} & 2T\sum_{\nu}\left[\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q}) + \chi^{0}_{\nu\omega}\right] - 2T^{2}\sum_{\nu\nu'}\chi^{0}_{\nu\omega}\Gamma_{\nu\nu'\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q}) - 2T^{2}\sum_{\nu\nu'}\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\chi^{0}_{\nu'\omega} \\ & - 2T^{2}\sum_{\nu\nu'}\chi^{0}_{\nu\omega}\Gamma_{\nu\nu'\omega}(\mathbf{q})\chi^{0}_{\nu'\omega} - 2T^{2}\sum_{\nu\nu'}\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q})\Gamma_{\nu\nu'\omega}(\mathbf{q})\tilde{\chi}^{0}_{\nu'\omega}(\mathbf{q})\end{aligned} \tag{H29}$$

$$= 2T\sum_{\nu}\left[\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q}) + \chi^{0}_{\nu\omega}\right] - 2T^{2}\sum_{\nu\nu'}\left[\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q}) + \chi^{0}_{\nu\omega}\right]\Gamma_{\nu\nu'\omega}(\mathbf{q})\left[\tilde{\chi}^{0}_{\nu\omega}(\mathbf{q}) + \chi^{0}_{\nu'\omega}\right]. \tag{H30}$$

Using (H22) this is seen to be equal to (H1).

[1] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[2] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005).
[3] A. Toschi, A. A. Katanin, and K. Held, Phys. Rev. B **75**, 045118 (2007).
[4] A. N. Rubtsov, M. I. Katsnelson, and A. I. Lichtenstein, Phys. Rev. B **77**, 033101 (2008).
[5] A. N. Rubtsov, M. I. Katsnelson, A. I. Lichtenstein, and A. Georges, Phys. Rev. B **79**, 045133 (2009).
[6] A. Rubtsov, M. Katsnelson, and A. Lichtenstein, Ann. Phys. (NY) **327**, 1320 (2012).
[7] G. Rohringer, A. Toschi, H. Hafermann, K. Held, V. I. Anisimov, and A. A. Katanin, Phys. Rev. B **88**, 115112 (2013).
[8] E. Gull, O. Parcollet, and A. J. Millis, Phys. Rev. Lett. **110**, 216405 (2013).
[9] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006).
[10] Q. Si and J. L. Smith, Phys. Rev. Lett. **77**, 3391 (1996).
[11] H. Kajueter, Ph.D. thesis, Rutgers University. 1996.
[12] J. L. Smith and Q. Si, Phys. Rev. B **61**, 5184 (2000).
[13] R. Chitra and G. Kotliar, Phys. Rev. B **63**, 115110 (2001).
[14] P. Sun and G. Kotliar, Phys. Rev. B **66**, 085120 (2002).
[15] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003).
[16] T. Ayral, P. Werner, and S. Biermann, Phys. Rev. Lett. **109**, 226401 (2012).
[17] T. Ayral, S. Biermann, and P. Werner, Phys. Rev. B **87**, 125149 (2013).
[18] M. Jarrell and T. Pruschke, Z. Phys. B: Condens. Matter **90**, 187 (1993).
[19] P. R. C. Kent, M. Jarrell, T. A. Maier, and T. Pruschke, Phys. Rev. B **72**, 060411 (2005).
[20] T. A. Maier, M. Jarrell, T. C. Schulthess, P. R. C. Kent, and J. B. White, Phys. Rev. Lett. **95**, 237001 (2005).
[21] G. Rohringer, A. Valli, and A. Toschi, Phys. Rev. B **86**, 125114 (2012).
[22] M. Kinza and C. Honerkamp, Phys. Rev. B **88**, 195136 (2013).
[23] L. Huang and Y. Wang, arXiv:1303.2818.
[24] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005).
[25] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, Phys. Rev. Lett. **97**, 076405 (2006).

[26] E. Gull, P. Werner, O. Parcollet, and M. Troyer, Europhys. Lett. **82**, 57003 (2008).
[27] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011).
[28] Y. Nambu, Phys. Rev. **117**, 648 (1960).
[29] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
[30] G. Baym, Phys. Rev. **127**, 1391 (1962).
[31] S. Brener, H. Hafermann, A. N. Rubtsov, M. I. Katsnelson, and A. I. Lichtenstein, Phys. Rev. B **77**, 195105 (2008).
[32] V. Janiš and V. Pokorný, Phys. Rev. B **90**, 045143 (2014).
[33] T. Schäfer, G. Rohringer, O. Gunnarsson, S. Ciuchi, G. Sangiovanni, and A. Toschi, Phys. Rev. Lett. **110**, 246405 (2013).
[34] A. Khurana, Phys. Rev. Lett. **64**, 1990 (1990).
[35] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
[36] M. Graf and P. Vogl, Phys. Rev. B **51**, 4940 (1995).
[37] N. Blümer, Ph.D. thesis, Universität Augsburg, 2002.
[38] G. D. Mahan, *Many-Particle Physics*, Physics of Solids and Liquids, 3rd ed. (Plenum, New York, 2000).
[39] P. Nozires, *Theory of Interacting Fermi Systems*, Advanced Book Classics (Addison-Wesley, Reading, MA, 1964).
[40] P. M. Platzman and P. A. Wolff, *Waves and Interactions in Solid State Plasmas* (Academic, New York, 1973).
[41] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, UK, 2005).
[42] R. Chitra and G. Kotliar, Phys. Rev. Lett. **84**, 3678 (2000).
[43] E. G. C. P. van Loon, A. I. Lichtenstein, M. I. Katsnelson, O. Parcollet, and H. Hafermann, arXiv:1408.2150.
[44] P. Anderson, *The Theory of Superconductivity in the High-TC Cuprate Superconductors*, Princeton Series in Physics (Princeton University Press, Princeton, NJ, 1997).
[45] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Pergamon, New York, 1965).
[46] H. Hafermann, P. Werner, and E. Gull, Comput. Phys. Commun. **184**, 1280 (2013).
[47] B. Bauer, L. D. Carr, H. G. Evertz, A. Feiguin, J. Freire, S. Fuchs, L. Gamper, J. Gukelberger, E. Gull, S. Guertler, A. Hehn, R. Igarashi, S. V. Isakov, D. Koop, P. N. Ma, P. Mates, H. Matsuo, O. Parcollet, G. Pawowski, J. D. Picon, L. Pollet, E. Santos, V. W. Scarola, U. Schollwck, C. Silva, B. Surer, S. Todo, S. Trebst, M. Troyer, M. L. Wall, P. Werner, and S. Wessel, J. Stat. Mech.: Theory Exper. (2011) P05001.
[48] P. Werner and A. J. Millis, Phys. Rev. Lett. **99**, 146404 (2007).
[49] P. Werner and A. J. Millis, Phys. Rev. Lett. **104**, 146401 (2010).
[50] H. Hafermann, Phys. Rev. B **89**, 235128 (2014).
[51] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).
[52] L. Huang, T. Ayral, S. Biermann, and P. Werner, Phys. Rev. B **90**, 195114 (2014).
[53] R. E. Peierls, Z. Phys. **80**, 763 (1933).
[54] A. S. Alexandrov and H. Capellmann, Phys. Rev. Lett. **66**, 365 (1991).
[55] J. Schrieffer, *Theory of Superconductivity*, Advanced Book Program Series (Westview Press, Boulder, CO, 1999).