Second-order dual fermion approach to the Mott transition in the two-dimensional Hubbard model

Erik G. C. P. van Loon,¹ Mikhail I. Katsnelson,¹ and Hartmut Hafermann²

¹*Radboud University, Institute for Molecules and Materials, NL-6525 AJ Nijmegen, The Netherlands*

²*Mathematical and Algorithmic Sciences Lab, Paris Research Center, Huawei Technologies France SASU, 92100 Boulogne Billancourt, France*

(Received 24 May 2018; revised manuscript received 24 September 2018; published 8 October 2018)

We apply the dual fermion approach with a second-order approximation to the self-energy to the Mott transition in the two-dimensional Hubbard model. The approximation captures nonlocal dynamical short-range correlations as well as several features observed in studies using cluster dynamical mean-field theory. This includes a strong reduction of the critical interaction and inversion of the slope of the transition lines with respect to single-site dynamical mean-field theory. We show that these effects coincide with a much smaller momentum differentiation compared to cluster methods. We further discuss the role of the self-consistency condition and show that the approximation behaves as an asymptotic series at low temperature.

DOI: [10.1103/PhysRevB.98.155117](https://doi.org/10.1103/PhysRevB.98.155117)

I. INTRODUCTION

The Mott metal-insulator transition (MIT) is a fundamental problem in condensed matter physics. Dynamical mean-field theory (DMFT) has provided important insights, because the low-energy coherent quasiparticle excitations as well as the incoherent high-energy excitations can be treated on the same footing [1]. The transition was later analyzed using cluster extensions of DMFT. Cellular DMFT results [2,3] revealed a substantial reduction of the critical interaction and a reversal of the slope of the transition lines. Subsequent DCA studies on larger clusters found a similar reduction of the critical interaction, but mainly focused on a different aspect of the transition [4–6]: For both the interaction-driven and doping-driven transition, an intermediate sector-selective phase was found, in which some momentum sectors become insulating while others remain metallic. In the studies on larger clusters (≥ 8 sites) the momentum resolution was sufficiently high to clearly separate the nodal and antinodal directions. All of them consistently showed that quasiparticles are first destroyed in the antinodal direction, leading to an “orbital-selective transition in momentum space.”

In cluster methods, the self-energy is treated rigorously, but correlations are restricted to the size of the cluster. Even though the cluster size acts as a control parameter, convergence is seldom reached in practice due to the exponential growth in computational complexity. Diagrammatic extensions of DMFT [7], like the dynamical vertex approximation (D Γ A) [8] and dual fermion (DF) [9], provide a complementary viewpoint. In these approaches, nonlocal corrections to the DMFT self-energy are included diagrammatically. The local interaction vertices are given by (reducible or irreducible) vertex functions of an impurity model. A significantly higher momentum resolution can be reached, while the self-energy in practice remains approximate at any scale. In general the results necessarily depend on the diagrammatic approximation. For small systems, all diagrammatic corrections obtained from a two-particle vertex can be obtained by solving the full set of parquet equations [10]. Another means to avoid a

possible bias is to sample diagrams irrespective of their topology using a diagrammatic Monte Carlo approach [11,12]. When diagrams in a given fluctuation channel can be expected to dominate for physical reasons, the ladder approximation is a more sophisticated alternative. Examples are the ladder dynamical vertex approximation (D Γ A) [8] or the ladder DF approximation (LDFA) [13]. These approximation can even yield quantitatively accurate results in regimes where they capture the dominant fluctuations [12,13]. Recent systematic studies based on these approaches have revealed precursors of the Mott insulator at low temperature and interaction strengths significantly smaller than the critical value obtained in CDMFT [14,15].

The summation of the infinite diagram series in the ladder approximation can be computationally demanding for inhomogeneous systems [16,17] or multiorbital systems. In addition, it requires a regularization procedure if the series initially diverges in the dynamical-mean solution [18], whose computational complexity typically increases with the strength of the fluctuations. A much simpler approximation is the second-order dual fermion approximation [9,19], which we refer to as DF⁽²⁾. The approximation has been shown to capture short-range spatial correlations beyond DMFT. So far, a systematic study based on the DF⁽²⁾ approximation and the role of the self-consistency condition has been lacking. For applications it is useful to understand its physical content and limitations. In this paper we aim to close this gap by reexamining the Mott transition in the two-dimensional Hubbard model within DF⁽²⁾ as a benchmark. We also report results obtained with an alternative self-consistency condition.

II. MODEL

We apply our method to the single-band Hubbard model on the two-dimensional square lattice

$$H = \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^{\dagger} c_{\mathbf{k}\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}. \quad (1)$$

The noninteracting dispersion

$$\epsilon_{\mathbf{k}} = -2t(\cos k_x + \cos k_y) - 4t'(\cos k_x \cos k_y) \quad (2)$$

is determined by nearest- (t) and next-nearest-neighbor hopping t' . We take $t = 1$ as the energy unit. In the above, latin indices denote lattice sites, $\sigma = \uparrow, \downarrow$ labels spin, $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$, and U is the local Hubbard repulsion.

In particular the two-dimensional model is well studied and often used as a benchmark for new computational approaches [20]. An exact solution to the interacting Hubbard model is known in the limit of infinite dimensions [21] provided by DMFT [1]. The DMFT solutions has provided important insights into the Mott metal-insulator transition. Cluster extensions of DMFT [22] treat short-range correlations exactly and refine this picture: In cellular DMFT the critical interaction is considerably reduced due to the effect of spatial correlations [2,3]. The dynamical cluster approximation indicates a strong momentum space differentiation and a momentum sector-selective opening of the gap. Below we address how these aspects are reflected in the DF⁽²⁾ approximation. Diagrammatic extensions [7] provide a complementary viewpoint, because they can treat also long-range spatial correlations, albeit perturbatively.

III. METHOD

In the following we briefly sketch the derivation of the dual fermion method to provide some intuition and to define all necessary quantities. A modern introduction and derivation can be found in review of Ref. [7]. The dual fermion approach is an action-based method. The basic idea is to introduce an auxiliary quantum impurity model at each lattice site (we assume translational invariance in the following). This may be done leaving the original action unaltered by formally adding and subtracting a hybridization function at each site. The result can be written in the form

$$S_{\text{lat}} = \sum_i S_{\text{imp}}[c_i^*, c_i] - \sum_{\mathbf{k}\nu\sigma} c_{\mathbf{k}\nu\sigma}^* (\Delta_{\nu\sigma} - \epsilon_{\mathbf{k}}) c_{\mathbf{k}\nu\sigma}, \quad (3)$$

where the impurity action S_{imp} is given by

$$S_{\text{imp}}[c^*, c] = - \sum_{\nu\sigma} c_{\nu\sigma}^* [\iota\nu + \mu - \Delta_{\nu\sigma}] c_{\nu\sigma} + U \sum_{\omega} n_{\omega\uparrow} n_{-\omega\downarrow}. \quad (4)$$

Here ν (ω) denote fermionic (bosonic) Matsubara frequencies.

In order to formulate a Feynman-type diagrammatic expansion around the non-Gaussian impurity model, the second term in (3) is decoupled by introducing auxiliary (dual) degrees of freedom through a Hubbard-Stratonovich transformation. It is important that these couple locally to the original fermions, for two reasons: First, this guarantees that the mapping to dual fermions preserves the topology of the diagrams. For example, a dual self-energy diagram connecting two neighboring sites will also introduce nearest-neighbor correlations in terms of the real fermions. And second, the original fermions can be integrated out for each site separately. This produces the connected correlation functions of the impurity coupled to dual fermions. The latter enter the resulting

FIG. 1. Second-order approximation to the dual self-energy. The vertex function γ is depicted by a square. Lines are fully dressed $\tilde{G}$ propagators.

dual action, which reads

$$\tilde{S} = - \sum_{\mathbf{k}\nu\sigma} f_{\mathbf{k}\nu\sigma}^* (\tilde{G}_{\mathbf{k}\nu\sigma}^0)^{-1} f_{\mathbf{k}\nu\sigma} + V[f^*, f]. \quad (5)$$

Here the bare dual Green's function is defined as

$$\tilde{G}_{\mathbf{k}\nu\sigma}^0 = G_{\mathbf{k}\nu\sigma}^{\text{DMFT}} - g_{\nu\sigma}, \quad (6)$$

where $g_{\nu\sigma} := -\langle c_{\nu\sigma} c_{\nu\sigma}^* \rangle$ denotes the interacting impurity Green's function. The interacting DMFT lattice Green's function is given by

$$(G_{\mathbf{k}\nu\sigma}^{\text{DMFT}})^{-1} = \iota\nu + \mu - \epsilon_{\mathbf{k}} - \Sigma_{\nu} \quad (7)$$

and, to leading order, the dual interaction V is

$$V[f^*, f] = -\frac{1}{4} \sum_{\nu\nu'\omega} \sum_{\sigma_i} \gamma_{\nu\nu'\omega}^{\sigma_1\sigma_2\sigma_3\sigma_4} f_{\nu\sigma_1}^* f_{\nu+\omega,\sigma_2} f_{\nu'+\omega,\sigma_3}^* f_{\nu'\sigma_4}. \quad (8)$$

The reducible impurity vertex hence plays the role of the bare interaction of the dual fermions. It is defined as

$$\gamma_{\nu\nu'\omega}^{\sigma\sigma'} := \frac{g_{\nu\nu'\omega}^{\sigma\sigma'} - \beta g_{\nu\sigma} g_{\nu'\sigma'} \delta_{\omega} + \beta g_{\nu\sigma} g_{\nu+\omega\sigma} \delta_{\nu\nu'} \delta_{\sigma\sigma'}}{g_{\nu\sigma} g_{\nu+\omega,\sigma} g_{\nu'+\omega,\sigma'} g_{\nu'\sigma'}}, \quad (9)$$

which involves the two-particle Green's function

$$g_{\nu\nu'\omega}^{\sigma\sigma'} := \langle c_{\nu\sigma} c_{\nu+\omega,\sigma}^* c_{\nu'+\omega,\sigma'} c_{\nu'\sigma'}^* \rangle. \quad (10)$$

In the above we have introduced the short-hand notation $\gamma^{\sigma\sigma'} := \gamma^{\sigma\sigma\sigma'\sigma'}$.

A. Dual self-energy

The second-order approximation to the dual self-energy is depicted diagrammatically in Fig. 1. In the paramagnetic case, it is convenient to work with the charge (c) and spin (s) components defined by $\gamma^{s(c)} = \gamma^{\uparrow\uparrow(\pm)} \gamma^{\uparrow\downarrow}$. Introducing

$$\tilde{\chi}_{\mathbf{q}\omega\nu}^0 = \frac{1}{N} \sum_{\mathbf{k}'} \tilde{G}_{\mathbf{k}'+\mathbf{q}\nu+\omega} \tilde{G}_{\mathbf{k}'\nu}, \quad (11)$$

the dual self-energy in the DF⁽²⁾ approximation reads

$$\begin{aligned} \tilde{\Sigma}_{\mathbf{k}\nu\sigma} = & -\frac{T}{N} \sum_{\mathbf{k}'\nu'} \gamma_{\nu\nu'\omega=0}^c \tilde{G}_{\mathbf{k}'\nu'} \\ & -\frac{1}{2} \frac{T^2}{N} \sum_{\mathbf{q}\omega} \sum_{\nu'} \left[\frac{1}{2} \gamma_{\nu\nu'\omega}^c \tilde{G}_{\mathbf{k}+\mathbf{q}\nu+\omega} \tilde{\chi}_{\mathbf{q}\omega\nu'}^0 \gamma_{\nu'\nu\omega}^c \right. \\ & \left. + \frac{3}{2} \gamma_{\nu\nu'\omega}^s \tilde{G}_{\mathbf{k}+\mathbf{q}\nu+\omega} \tilde{\chi}_{\mathbf{q}\omega\nu'}^0 \gamma_{\nu'\nu\omega}^s \right]. \end{aligned} \quad (12)$$

The physical self-energy is obtained from its dual counterpart through the relation

$$\Sigma_{\mathbf{k}v} = \Sigma_v^{\text{imp}} + \frac{\tilde{\Sigma}_{\mathbf{k}v}}{1 + \tilde{\Sigma}_{\mathbf{k}v}g_v}. \quad (13)$$

Equivalently, one may obtain the physical Green's function $G_{\mathbf{k}v}$ from the renormalized to dual Green's function $\tilde{G}_{\mathbf{k}v}$ through the expression

$$G_{\mathbf{k}v} = (\Delta_v - \epsilon_{\mathbf{k}})^{-1} + (\Delta_v - \epsilon_{\mathbf{k}})^{-1} g_v^{-1} \tilde{G}_{\mathbf{k}v} g_v^{-1} (\Delta_v - \epsilon_{\mathbf{k}})^{-1}. \quad (14)$$

B. Self-consistency condition and computational scheme

A priori, the choice of hybridization function is arbitrary. It was introduced such that in an exact theory, the solution is independent of its value. In practice however, it acts to improve the starting point of the perturbation theory. There are two apparent self-consistency conditions to fix the hybridization function. Condition (i) has been used exclusively in the literature so far. It is obtained by requiring the local part of the *interacting* dual Green's function to vanish:

$$\text{SC (i): } \frac{1}{N} \sum_{\mathbf{k}} \tilde{G}_{\mathbf{k}v} = 0. \quad (15)$$

In the following, we refer to this condition as the dual self-consistency condition. Condition (ii) is familiar from DMFT and reads

$$\text{SC (ii): } g_v = \frac{1}{N} \sum_{\mathbf{k}} G_{\mathbf{k}v}. \quad (16)$$

We refer to it as the lattice condition. When no dual self-energy diagrams are taken into account, this condition is seen to be equivalent to requiring the local part of the bare dual Green function to be zero, $(1/N) \sum_{\mathbf{k}} \tilde{G}_{\mathbf{k}v}^0 = 0$, by virtue of Eqs. (6) and (7). This shows that a theory of noninteracting dual fermions is equivalent to DMFT [9]. As soon as dual self-energy diagrams are taken into account however, this no longer holds and the resulting hybridization will be different from the DMFT. The physical lattice Green function then contains a momentum dependent self-energy, Eq. (13).

The two conditions are not equivalent in general, except in case of noninteracting dual fermions, where they both correspond to DMFT. We discuss the physical content of these conditions in the results section below.

In practice, we enforce the self-consistency condition through an iterative update of the hybridization function, similar to DMFT. For example, condition (i) is enforced by the update rule

$$\text{SC (i): } \Delta_v^{\text{new}} = \Delta_v^{\text{old}} + \xi [g_v^{-1} \tilde{G}_v^{\text{loc}} (G_v^{\text{loc}})^{-1}], \quad (17)$$

where $\xi \in (0, 1]$ is a mixing parameter. Convergence is reached when $\tilde{G}_v^{\text{loc}} = 0$. The other Green's functions only act as scaling factors. They are chosen such that for $\tilde{G} = \tilde{G}^0$, the angular bracket evaluates to $g_v^{-1} - (G_v^{\text{loc, DMFT}})^{-1}$. Condition (ii) is enforced through the update

$$\text{SC (ii): } \Delta_v^{\text{new}} = \Delta_v^{\text{old}} + \xi [g_v^{-1} - (G_v^{\text{loc}})^{-1}]. \quad (18)$$

FIG. 2. Visualization of the computational scheme.

In summary, the computational scheme is as follows:

0. Generate an initial guess for the hybridization function Δ (e.g., from DMFT).
1. For a given Δ compute g , Σ_{imp} , and γ .
2. Evaluate the diagrams, Eq. (12), to compute $\tilde{\Sigma}$.
3. Compute an updated $\tilde{G}$ from the dual Dyson equation, $\tilde{G}^{-1} = \tilde{G}_0^{-1} - \tilde{\Sigma}$.
4. Repeat steps 2 and 3 until convergence (inner loop).
5. From a converged $\tilde{G}$ compute an update of the hybridization function Δ according to (17).
6. Repeat steps 1–5 until convergence (outer loop).

This procedure is illustrated in Fig. 2. If steps 2 to 4 are omitted (inner loop), the outer loop is equivalent to DMFT. In this case γ does not have to be calculated. The computational effort for a dual fermion iteration is dominated by the measurement of the vertex function.

IV. RESULTS

The following results are obtained for a lattice of size 64×64 sites as a compromise between fine momentum resolution and consumption of computational and memory resources. For the solution of the impurity model, we use the implementation of the segment picture variant of the continuous-time hybridization expansion algorithm [23] of Ref. [24] with improved estimators for the self-energy and vertex function [25]. Here we employ the self-consistency condition (i). We compare the two alternative conditions in Sec. V.

A. Mott transition

We plot the phase diagram in the interaction-temperature plane of the two-dimensional Hubbard model at half-filling in Fig. 3. The transition line $U_{c,1}$ (red in color) delimits the region of stability of the insulating phase, while $U_{c,2}$ delimits the stability region of the metallic phase. We have determined the phase boundaries from the hysteresis of $\beta G(\beta/2)$, which approximates the density of states at the Fermi level. The transition lines encompass a coexistence region of metallic and insulating solutions. They merge at the critical point $U_c \sim 6.64$ at a temperature $T \sim 0.11$. Notably, their slope is positive. Above the critical point, we find a crossover region between the metallic and insulating states. The approximate

FIG. 3. Phase diagram of the two-dimensional Hubbard model obtained with dual fermion (DF) calculations. The critical point occurs at $U_c \sim 6.64$ and $T = 0.11$. Corresponding data for cellular DMFT (CDMFT) and DMFT taken from Ref. [3] are shown for comparison.

position of the crossover line has been determined from the condition of maximum slope of $\beta G(\beta/2)$ vs U .

DMFT and CDMFT results taken from Ref. [3] are shown for comparison. In agreement with these methods, the Mott transition remains first order in $DF^{(2)}$. In accordance with CDMFT, the critical interaction is significantly reduced compared to DMFT. Furthermore, both methods exhibit a positive slope of the transition lines which is reversed compared to DMFT. This is a consequence of the effects of spatial correlations and has previously been explained through an entropy argument [3]. In DMFT at $T = 0$, the insulating solution represents a lattice of independent spin $s = 1/2$ magnetic moments with a large residual entropy of $\log 2$ per site, while the entropy of the metal is lower. As temperature is increased, the insulator is preferred over the metal. In the presence of spatial correlations, the residual entropy is greatly reduced and the situation is essentially opposite: the metal is preferred at elevated temperatures. As a result the slope of the transition lines is reversed.

The entropy also determines the slope of the crossover lines at high temperature. While the slope is positive immediately above the critical point, it becomes negative at very high temperature. In this regime, spatial correlations are thermally destroyed and the high-temperature insulating phase is favored because of its high entropy. The negative slope is therefore common among all three approaches.

The critical U_c in $DF^{(2)}$ is substantially reduced compared to the DMFT value $U_c = 9.35$ [3] and in good agreement with 2×2 CDMFT, which has the lowest value ($U_c = 6.05$) [3] and even closer to the 16-site DCA value $U_c = 6.53$ [26]. It is likely that the four-site CDMFT cluster overestimates short-range correlations due to absence of correlations on intermediate length scales and therefore underestimates the critical U . Indeed the plaquette singlet ground state has the highest probability among the cluster eigenstates [3]. In the 16-site DCA calculation we expect the nearest-neighbor correlations to be reduced, as the cluster size extends well

beyond the plaquette. As a result the critical U_c is higher compared to CDMFT. In the $DF^{(2)}$, correlations are included in principle up to the extension of the lattice and hence well beyond the DCA cluster. However, due to the truncation of the diagrammatic series, correlations are only treated approximately at any length scale. Since the dual Green's function decays fast in real space, correlations are effectively relatively short ranged. This may be the reason why the dual fermion result is even somewhat higher than in DCA. Nevertheless, the two values are remarkably close.

In Fig. 4 we plot $-\beta G(\tau = \beta/2)$ as a function of U (G denotes the local lattice Green's function) as an indicator of the local density of states $A(\omega = 0)$ at the Fermi level and to distinguish metallic and insulating phases. The phase diagram in Fig. 3 has been determined from such curves. At high temperatures we find a crossover from the metallic to the insulating state as the interaction is increased. As the temperature is lowered, the curves become very steep at the crossover and a coexistence region develops below the critical point (lower panels). The hysteresis is also visible in other observables, such as the impurity double occupancy (not shown).

In the same figure, we have included the values of $-\beta G(\beta/2)$ from individual dual fermion outer loop iterations, marked as gray dots. This provides information on convergence. The uppermost dot for a given U is the result from the first dual fermion iteration. It varies with U because we start from a converged DMFT hybridization at the respective U as the initial guess. In the lower panels we additionally plot the individual values from iterations initialized with the same insulating seed. In both cases, the final value is usually approached monotonously, although very close to the solution oscillations may occur. Points which are denser on a vertical scale indicate a decelerated convergence towards the final solution. One can clearly see this effect in the crossover regime at lower temperature (upper right panel), or at the edges of the hysteresis curves which delineate the coexistence region (lower panels). The critical slowing down is also observed in DMFT. Overall the convergence properties are similar to DMFT.

B. Momentum space differentiation

Because of the momentum dependence of the $DF^{(2)}$ self-energy one can expect some differentiation introduced through the diagrammatic corrections, which is absent in DMFT. In Fig. 5 we plot $-\beta G_{\mathbf{k}}(\beta/2)$ for two high-symmetry points in the Brillouin zone, which correspond to the nodal $(\pi/2, \pi/2)$ and antinodal $(\pi, 0)$ directions. A clear momentum space differentiation occurs sufficiently far from the transition. The value of the spectral function at the antinode is smaller than at the node, in accordance with the fact that quasiparticles are more strongly renormalized in this direction [4–6, 19, 27]. As the transition is approached however, the differentiation diminishes and vanishes at the transition point.

This behavior is contrary to the DCA, where the transition manifests itself as orbital selective in momentum space: the self-energy at the antinode undergoes the transition before the one at the node. To compare our results with those of the DCA,

FIG. 4. Estimator for the local density of states $-\beta G(\beta/2) \approx \pi A(0)$ as a function of U for different temperatures. Results marked in red were obtained by starting the iterations with a metallic seed, while those marked by in blue were started from an insulating solution. Gray dots mark results from individual iterations and illustrate the convergence (for explanation see text).

we coarse-grain the Green's function according to the DCA patches.

In DCA, the self-energy is a piecewise constant function in momentum space and is expressed in terms of basis functions $\phi_{\mathbf{K}}(\mathbf{k})$ which are equal to one if $\mathbf{k}$ is contained in the patch centered around $\mathbf{K}$ and zero otherwise. Here we choose the eight-site patch geometry, because the nodal and antinodal directions can be well distinguished and sufficient data is available for comparison. In Fig. 6 we show the mapping of the 64×64 individual k points of the DF calculation to the corresponding DCA patches. Each patch contains the same number of k points. The $N_p = 8$ patches can further be grouped by symmetry as indicated by color. The patches of type B contain the nodal and the type C patches the antinodal points. Since the Brillouin zone is tiled completely by all patches, they have the property that

$$\sum_{\mathbf{K}=1}^{N_p} \phi_{\mathbf{K}}(\mathbf{k}) = 1. \quad (19)$$

FIG. 5. Momentum differentiation in the dual fermion approach. The estimator $-\beta G_{\mathbf{k}}(\beta/2)$ for the Fermi level spectral function is plotted for two k points on the Fermi surface at $T = 0.08$ as a function of U . The estimator for the local DOS is shown for comparison.

The *coarse-grained* Green's function is defined as

$$\bar{G}_{\mathbf{K}}(i\nu) = \frac{N_p}{N} \sum_{\mathbf{k}} \phi_{\mathbf{K}}(\mathbf{k}) G_{\mathbf{k}}(i\nu), \quad (20)$$

so that the local Green's function is obtained as a momentum integral over the patch momenta,

$$G(i\nu) = \frac{1}{N_p} \sum_{\mathbf{K}} \bar{G}_{\mathbf{K}}(i\nu) = \frac{1}{N} \sum_{\mathbf{k}} G_{\mathbf{k}}(i\nu). \quad (21)$$

We plot $-\beta G_{\mathbf{K}}(\beta/2)$ obtained from the coarse-grained Green's function on the four different patches in Fig. 7. We

FIG. 6. Mapping of individual k points in the dual fermion calculation to the corresponding DCA patches in an eight-site patch geometry. The number of points corresponds to the native momentum resolution used in the dual fermion calculations (64×64 in this case). The noninteracting Fermi surfaces for $t'/t = 0$ (square) and $t'/t = -0.3$ are also shown. Patches B and C contain the Fermi surface. Patch B covers the nodal and patch C the antinodal direction.

FIG. 7. Estimator for the density of states integrated over the momentum patches of Fig. 6 as a function of U/t approaching the Mott transition in $DF^{(2)}$ for fixed temperature $T/t = 1/12$ and $t' = 0$. The plot shows $-\beta G_K(\beta/2)$ rescaled by the number n_α of equivalent patches, where $n_A = 1$, $n_B = 4$, $n_C = 2$, and $n_D = 1$. The Fermi surface is entirely contained in patches B and C, so that the contribution on patches A and D is negligibly small.

observe that sufficiently below the transition, the result in sector B (around the nodal point) is smaller than that in sector C, which contains the antinodal point [28]. This seems contrary to the previous result of Fig. 5. Close to the transition however, the sector C value shifts below the sector B value as indicated by the arrow. It remains smaller all the way to the transition. This is in qualitative agreement with the DCA results of Fig. 8 of Ref. [6]. What we do not observe, however, is a gap opening in sector C before it happens in sector B. In second-order DF, the gap appears to open in both sectors simultaneously. The difference between the two sectors is also much smaller. Taken together with the results of Sec. IV A, we find that while our method and cluster methods agree qualitatively concerning the strong reduction of the critical interaction and the inversion of the slope of the transition lines, the momentum differentiation is much smaller than predicted by the DCA.

In Fig. 8 we show results similar to the ones in Fig. 7, albeit for the doping driven transition for different values of the next-nearest-neighbor hopping t' . Similarly to the case before we find that the momentum space differentiation vanishes at the transition.

C. Higher-order diagrams

The results presented so far were all been obtained using the $DF^{(2)}$ approximation, i.e., using (12) for the self-energy. It is possible to take into account higher-order diagrams in the dual theory. Two common techniques for this are the ladder approximation [13] and diagrammatic Monte Carlo techniques [11, 12]. Here we compare $DF^{(2)}$ with higher-order DF approaches: $DF^{(3)}$ and $DF^{(4)}$. That is, we add the third- and fourth-order ladder diagrams to the expression for the dual self-energy in Eq. (12).

Figure 9 shows the physical self-energy at $U/t = 6.2$, $T/t = 0.1$. In $DF^{(2)}$, these parameters are close to the transition and the convergence of the outer loop is relatively slow, cf. Fig. 4. Figure 9 shows that the inclusion of higher-order

FIG. 8. Quantity $-\beta G_K(\beta/2)$ as in Fig. 7 but for the doping driven transition as a function of $\tilde{\mu} = \mu - U/2$ for $U/t = 7$, $T/t = 0.08$ and different values of the next-nearest-neighbor hopping t' on the patches A and B. The color coding corresponds to the one in Fig. 6.

diagrams leads to more momentum differentiation compared to $DF^{(2)}$. However, we also observe that the self-energy in $DF^{(4)}$ actually has an unphysical positive imaginary part.

These results indicate that the DF series behaves as an asymptotic series, in which the lowest-order approximation yields reasonable results in accordance with other methods, whereas higher-order approximations of a divergent series may become unphysical. Consistent with Fig. 10 of Ref. [12], these parameters are deep inside the region inside the regime of poor convergence of the series sampled using a diagrammatic Monte Carlo approach. This should be kept in mind if the series is outside of its convergence radius. Convergence of the series can be easily tested by evaluating the leading eigenvalue of the matrix $(-T/N) \sum_{\nu' \mathbf{k}'} \gamma_{\nu \nu'} \omega \tilde{G}_{\mathbf{k} \nu'} \tilde{G}_{\mathbf{k} + \mathbf{q} \nu' + \omega}$. The divergence is closely related to the perfect nesting of the Fermi surface and indicative of a (spurious) second-order transition to the antiferromagnetic state because of the mean-field starting point. Away from half-filling or on frustrated lattices, this problem is less of an issue. On the triangular lattice, the second-order approximation gives results which are very close to 16-site DCA results [29]. We emphasize that the infinite

FIG. 9. Comparison of the self-energy at $U/t = 6.2$ and $T/t = 0.1$ between $DF^{(2)}$ with higher-order dual approximations. The solid figures correspond to $\mathbf{k} = (\pi, \pi)$, the empty symbols to $\mathbf{k} = (\pi, 0)$.

FIG. 10. Comparison of impurity and local lattice Green's functions in DMFT and DF with the dual self-consistency condition (i), at $\beta t = 12$. In contrast to DMFT, in DF the impurity model Green's function does not equal the local lattice Green's function.

ladder approximation does not suffer from this shortcoming. In the latter, convergence of an initially diverging series can be ensured through a self-consistent renormalization of the dual Greens functions [18]. It however describes different physics, as it includes long-range correlations on all length scales.

V. SELF-CONSISTENCY CONDITIONS

Previous DF results were exclusively based on the dual self-consistency condition (i), Eq. (15). In this section we compare the dual and lattice self-consistency conditions.

The outer self-consistency modifies the auxiliary quantum impurity model, with the idea that it captures a much of the correlation effects as possible. The importance of this outer self-consistency was already illustrated in Fig. 4. As discussed in Sec. III, it is possible to contemplate alternative self-consistency conditions, as was also done in other DMFT-based methods [30,31].

A. Self-consistency at half-filling

We compare the effect of the two DF self-consistency conditions based on the impurity and local lattice Green's function in DMFT and DF. Figure 10 shows the result for the condition (i), Eq. (15). In all four panels the DMFT impurity Green's function is equal to the local lattice DMFT Green's function $g^{\text{DMFT}} = G_{\text{loc}}^{\text{DMFT}}$, as required by DMFT self-consistency. In DF however, the impurity Green's function (g^{DF}) and the local lattice Green's function ($G_{\text{loc}}^{\text{DF}}$) are different. For small U (top left panel) the DF impurity Green's function g^{DF} remains essentially unchanged during the outer loop self-consistency. However, both show a visible correction compared to the DMFT result, which corresponds to a

FIG. 11. Comparison of impurity and local lattice Green's functions in DMFT and DF using two different self-consistency conditions, Eqs. (15) and (16). This corresponds to the top left and bottom right panels in Fig. 10, with the additional brown diamonds denoting the results of self-consistency condition (ii), i.e., Eq. (16). Note that g^{DF} is equal to $G_{\text{loc}}^{\text{DF}}$ when using self-consistency condition (ii), so that the former is not shown in the figure.

suppression of spectral weight at the Fermi level. For larger values of U closer to the transition this correction increases and also the impurity Green's function in DF differs substantially from DMFT.

The lower right panel further shows that the outer loop iterations are essential to capture the Mott transition: the result from the first DF iteration is based on a metallic environment where the hybridization is equal to its DMFT value. $G_{\text{loc}}^{\text{DF}}$ corresponds to a diagrammatic expansion around DMFT and yields a metallic solution. The self-consistent solution $G_{\text{loc}}^{\text{DF}} \text{ sc}$ however is insulating, as is the impurity model.

An explanation may be that the Mott transition is nonperturbative and hence cannot be reached by expanding around a metallic state. On the other hand, condition (i) corresponds to the summation of an infinite partial series: all diagrams with a local loop are absorbed into the impurity model. Furthermore, the perturbation series contains nonlocal contributions to the self-energy, which give rise to local contributions to the dual Green's function. Nonlocal correlations therefore effectively alter the local environment (the hybridization function).

The interpretation of condition (ii) in terms of dual diagrams is not obvious. Figure 11 compares results for the two conditions. For small U (left panel) well below the transition, the impurity model remains essentially the same as in DMFT using either self-consistency condition. However, for large U (right panel), the results of the two self-consistency conditions are notably different. The dual self-consistency condition (i) results in an insulating solution, while the lattice condition (ii) actually mediates between the DMFT solution and the first dual iteration and does not capture the transition to the insulating state. We have also performed calculations using the lattice self-consistency condition at larger interaction strengths, above the DMFT MIT, and did not find an insulating state. However, above the DMFT MIT, the simulations with

this self-consistency condition were hard to stabilize due to issues that will be discussed below.

B. Self-consistency away from half-filling

Additional understanding of the role of the self-consistency condition can be gained by moving away from half-filling. The motivation for the traditional DF self-consistency condition (15) is that it ensures that Hartree-like diagrams (the first diagram in Fig. 1) vanish [9]. Looking at an asymptotic expansion of $\Sigma_{\mathbf{k}\nu}$ in terms of ν , this means that there are no nonlocal contributions to the constant and $(i\nu)^{-1}$ terms and the leading dual contributions occur at order $(i\nu)^{-2}$. This is of practical importance since the dual contribution requires knowledge of the vertex $\gamma_{\nu\nu'}/\omega$, and the fast decay means that the vertex only needs to be determined for a limited number of fermionic frequencies. Apart from this practical computational issue, the vanishing Hartree diagram also has important theoretical considerations.

A point of attention for diagrammatic extensions based on impurity models is that the final local observables of the theory are generally not identical to those of the impurity model [12,31–33]. In particular, away from half-filling the density can be inconsistent between the impurity model and the lattice theory [34]. As explained above, using self-consistency condition (15) ensures that the Hartree diagram vanishes so that

$$\lim_{\nu \rightarrow \infty} \text{Re } \Sigma_{\mathbf{k}\nu} \stackrel{\text{SC (i)}}{=} \lim_{\nu \rightarrow \infty} \text{Re } \Sigma_{\text{imp}} = \frac{U}{2} \langle n \rangle_{\text{imp}}. \quad (22)$$

In this sense, the self-energy with nonlocal corrections is consistent with the density of the impurity model. The same holds for the $(i\nu)^{-1}$ term. On the other hand, there is a mismatch between the single-particle Green's function of the lattice and that of the auxiliary impurity model,

$$\text{Tr } G \stackrel{\text{SC (i)}}{\neq} \text{Tr } g = \langle n \rangle_{\text{imp}}. \quad (23)$$

Combining Eqs. (22) and (23) shows that the dual approximation does not satisfy this property of the exact solution

$$\frac{U}{2} \text{Tr } G \stackrel{\text{SC (i)}}{\neq} \lim_{\nu \rightarrow \infty} \text{Re } \Sigma_{\mathbf{k}\nu}. \quad (24)$$

The alternative self-consistency condition (16) enforces $G^{\text{loc}} = g$ and thus also

$$\text{Tr } G \stackrel{\text{SC (ii)}}{=} \text{Tr } g = \langle n \rangle_{\text{imp}}. \quad (25)$$

This comes at a cost, since there is now a finite dual contribution to the asymptotics of the self-energy and

$$\lim_{\nu \rightarrow \infty} \text{Re } \Sigma_{\mathbf{k}\nu} \stackrel{\text{SC (ii)}}{\neq} \lim_{\nu \rightarrow \infty} \text{Re } \Sigma_{\text{imp}} = \frac{U}{2} \langle n \rangle_{\text{imp}}, \quad (26)$$

which again implies

$$\frac{U}{2} \text{Tr } G \stackrel{\text{SC (ii)}}{\neq} \lim_{\nu \rightarrow \infty} \text{Re } \Sigma_{\mathbf{k}\nu}. \quad (27)$$

We should stress that these arguments hold when only diagrams containing two-particle vertices are used. Diagrams with higher-order vertices do contribute to the local asymptotics. The dual transformation is an exact transformation, so that the exact solution of the dual action (all diagrams with

FIG. 12. The hybridization function at $U/t = 4$, $\mu/t = 1$, $\beta t = 6$, in DMFT and in self-consistent DF using self-consistency conditions (15) and (16). The cutoff for dual self-energy corrections is at $\nu_{\text{max}} = (2 \cdot 40 + 1)\pi T \approx 42$.

all vertices taken into account) results in the exact solution of the original model and in particular satisfies all analytical properties of the exact solution.

From a practical point of view, (26) is problematic for the frequency cutoff in the dual calculations. It essentially means that any finite frequency cutoff ν_{max} in the calculation of $\tilde{\Sigma}$ always leads to a *finite* discontinuity in $\text{Re } \Sigma$ at ν_{max} . This differs from the traditional self-consistency condition, where the magnitude of the discontinuity in Σ decays as $(i\nu)^{-2}$. To illustrate this, Fig. 12 shows the self-consistently determined hybridization function away from half-filling. In the results using (16) (brown diamonds), the real part has a discontinuity at the frequency corresponding to the cutoff used for dual diagrams, even though this cutoff happens at a rather large energy, approximately 5 times the bandwidth. Thus for calculations away from half-filling based on the self-consistency condition (16) is impractical unless measures are taken to ensure the proper frequency asymptotics of the dual diagrams.

VI. CONCLUSIONS

We have investigated the second-order dual fermion approximation and how it captures the Mott transition in the two-dimensional Hubbard model: The transition is first order and the critical interaction is strongly reduced compared to (single-site) DMFT. The results are in close agreement with results obtained from cluster approaches. The approximation also captures the inversion of the slope of the transition lines with respect to DMFT. We interpret this as an effect of the dynamical short-range nearest-neighbor antiferromagnetic correlations which are incorporated through the diagrams and which reduce the entropy of the paramagnetic metallic state.

While the strong reduction of the critical interaction and the inversion of the slope of the transition lines are in qualitative agreement with cluster DMFT results, we find a much weaker sector selective momentum differentiation. Contrary to DCA, it disappears at the Mott transition. It would be interesting to see whether the sector-selective character of the DCA solution is increased or decreased when augmented with DF corrections in a cluster dual fermion formulation [35–37].

The second-order approximation is a relatively simple approximation which allows one to study nontrivial effects of

short-range dynamical spatial correlations in correlated systems beyond DMFT. Including truly long-ranged correlations that respect the Mermin-Wagner theorem requires at least a ladder approximation.

We found that the dual perturbation series behaves like an asymptotic series. Higher-order diagrams yield unphysical results at low temperature, so that the approximation should be used with care outside its radius of convergence. This is in line with the observation that the DF diagrammatic series may converge to wrong results at low temperature [12]. We note that the ladder approximation, on the contrary, yields sensible results [12,13,15].

We further reported results based on the alternative self-consistency condition akin to the one used in DMFT. This condition does not capture the Mott transition. It also changes the asymptotics of the physical fermion self-energy, making it

less relevant in practice. This is consistent with the recent findings of Ref. [38]. The conventionally used self-consistency condition, on the other hand, corresponds to the elimination of an infinite partial series from the diagrammatic expansion, which may be crucial to obtain an insulating solution based on the metallic DMFT starting point.

ACKNOWLEDGMENTS

The simulations employed a modified version of an open source implementation of the hybridization expansion quantum impurity solver [24], based on the ALPS libraries [39]. The authors would like to thank Tin Ribic for useful discussion. E.G.C.P.v.L. and M.I.K. acknowledge support from the European Research Council Advanced Grant No. 338957 FEMTO/NANO.

-
- [1] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996).
 - [2] Y. Z. Zhang and M. Imada, *Phys. Rev. B* **76**, 045108 (2007).
 - [3] H. Park, K. Haule, and G. Kotliar, *Phys. Rev. Lett.* **101**, 186403 (2008).
 - [4] M. Ferrero, P. S. Cornaglia, L. De Leo, O. Parcollet, G. Kotliar, and A. Georges, *Phys. Rev. B* **80**, 064501 (2009).
 - [5] P. Werner, E. Gull, O. Parcollet, and A. J. Millis, *Phys. Rev. B* **80**, 045120 (2009).
 - [6] E. Gull, O. Parcollet, P. Werner, and A. J. Millis, *Phys. Rev. B* **80**, 245102 (2009).
 - [7] G. Rohringer, H. Hafermann, A. Toschi, A. A. Katanin, A. E. Antipov, M. I. Katsnelson, A. I. Lichtenstein, A. N. Rubtsov, and K. Held, *Rev. Mod. Phys.* **90**, 025003 (2018).
 - [8] A. Toschi, A. A. Katanin, and K. Held, *Phys. Rev. B* **75**, 045118 (2007).
 - [9] A. N. Rubtsov, M. I. Katsnelson, and A. I. Lichtenstein, *Phys. Rev. B* **77**, 033101 (2008).
 - [10] A. Valli, T. Schäfer, P. Thunström, G. Rohringer, S. Andergassen, G. Sangiovanni, K. Held, and A. Toschi, *Phys. Rev. B* **91**, 115115 (2015).
 - [11] S. Isakov, A. E. Antipov, and E. Gull, *Phys. Rev. B* **94**, 035102 (2016).
 - [12] J. Gukelberger, E. Kozik, and H. Hafermann, *Phys. Rev. B* **96**, 035152 (2017).
 - [13] H. Hafermann, G. Li, A. N. Rubtsov, M. I. Katsnelson, A. I. Lichtenstein, and H. Monien, *Phys. Rev. Lett.* **102**, 206401 (2009).
 - [14] T. Schäfer, F. Geles, D. Rost, G. Rohringer, E. Arrigoni, K. Held, N. Blümer, M. Aichhorn, and A. Toschi, *Phys. Rev. B* **91**, 125109 (2015).
 - [15] E. G. C. P. van Loon, H. Hafermann, and M. I. Katsnelson, *Phys. Rev. B* **97**, 085125 (2018).
 - [16] N. Takemori, A. Koga, and H. Hafermann, *J. Phys.: Conf. Ser.* **683**, 012040 (2016).
 - [17] N. Takemori, A. Koga, and H. Hafermann, *arXiv:1801.02441*.
 - [18] J. Otsuki, H. Hafermann, and A. I. Lichtenstein, *Phys. Rev. B* **90**, 235132 (2014).
 - [19] A. N. Rubtsov, M. I. Katsnelson, A. I. Lichtenstein, and A. Georges, *Phys. Rev. B* **79**, 045133 (2009).
 - [20] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung, Y. Deng, M. Ferrero, T. M. Henderson, C. A. Jiménez-Hoyos *et al.* (Simons Collaboration on the Many-Electron Problem), *Phys. Rev. X* **5**, 041041 (2015).
 - [21] W. Metzner and D. Vollhardt, *Phys. Rev. Lett.* **62**, 324 (1989).
 - [22] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, *Rev. Mod. Phys.* **77**, 1027 (2005).
 - [23] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, *Phys. Rev. Lett.* **97**, 076405 (2006).
 - [24] H. Hafermann, P. Werner, and E. Gull, *Comput. Phys. Commun.* **184**, 1280 (2013).
 - [25] H. Hafermann, K. R. Patton, and P. Werner, *Phys. Rev. B* **85**, 205106 (2012).
 - [26] P. Werner (private communication).
 - [27] Y. M. Vilk and A.-M. S. Tremblay, *J. Phys. I France* **7**, 1309 (1997).
 - [28] This can be understood from the non-interacting band structure. The antinodal point is a Van Hove singularity at the Fermi energy, i.e., $E_{\mathbf{k}} = 0$ and $\nabla_{\mathbf{k}} E_{\mathbf{k}} = 0$, so that there are many states close to the Fermi energy in sector C.
 - [29] H. Lee, G. Li, and H. Monien, *Phys. Rev. B* **78**, 205117 (2008).
 - [30] E. A. Stepanov, E. G. C. P. van Loon, A. A. Katanin, A. I. Lichtenstein, M. I. Katsnelson, and A. N. Rubtsov, *Phys. Rev. B* **93**, 045107 (2016).
 - [31] E. G. C. P. van Loon, F. Krien, H. Hafermann, E. A. Stepanov, A. I. Lichtenstein, and M. I. Katsnelson, *Phys. Rev. B* **93**, 155162 (2016).
 - [32] G. Rohringer and A. Toschi, *Phys. Rev. B* **94**, 125144 (2016).
 - [33] F. Krien, E. G. C. P. van Loon, H. Hafermann, J. Otsuki, M. I. Katsnelson, and A. I. Lichtenstein, *Phys. Rev. B* **96**, 075155 (2017).
 - [34] The case of half-filling in the square lattice Hubbard model is rather special. There is particle-hole symmetry, the vertex $\gamma_{v'v'\omega}$ is purely real and the Green's function $\tilde{G}_v$ is purely imaginary.

This combination ensures that there is no contribution to the real part of the self-energy from the Hartree diagram. This explains why the inconsistencies discussed below are less important at half-filling: The particle-hole symmetry makes the leading order inconsistency go away. Above, we mentioned that it was hard to converge calculations using the lattice self-consistency condition (16) for $U > U_{\text{MIT}}^{\text{DMFT}}$. The origin of this instability lies in the density which always deviates from half-filling due to Monte Carlo errors, in this regime these deviations are unstable within the self-consistent loop and led to larger and larger deviations of the density from half-filling.

- [35] H. Hafermann, S. Brener, A. Rubtsov, M. Katsnelson, and A. Lichtenstein, *JETP Lett.* **86**, 677 (2008).
- [36] S.-X. Yang, H. Fotsos, H. Hafermann, K.-M. Tam, J. Moreno, T. Pruschke, and M. Jarrell, *Phys. Rev. B* **84**, 155106 (2011).
- [37] S. Isakov, H. Terletska, and E. Gull, *Phys. Rev. B* **97**, 125114 (2018).
- [38] T. Ribic, P. Gunacker, and K. Held, *Phys. Rev. B* **98**, 125106 (2018).
- [39] B. Bauer, L. D. Carr, H. G. Evertz, A. Feiguin, J. Freire, S. Fuchs, L. Gamper, J. Gukelberger, E. Gull, S. Guertler *et al.*, *J. Stat. Mech.: Theory Exp.* (2011) P05001.