Talks by André-Marie Tremblay, Les Houches, June 2006

Comments and complete reference on documents included mostly in cond-mat version in this pdf file :

Review on which talks were based (including results from TPSC-CPT-CDMFT-VCPT):

A.-M. S. Tremblay, B. Kyung, and D. Sénéchal, Low Temperature Physics (Fiz. Nizk. Temp.) **32**, 561 (2006).

Two-Particle Self-Consistent Approach

Formal derivation of the Two-Particle Self-Consistent Approach :

S. Allen, A.-M. Tremblay and Y. M. Vilk, in Theoretical Methods for Strongly Correlated Electrons, edited by D. Sénéchal, C. Bourbonnais and A.-M. Tremblay, CRM
Series in Mathematical Physics, Springer (2003).

See section 5 and Appendix D of following paper for simple scaling arguments concerning the pseudogap

Y. Vilk and A.-M. Tremblay, J. Phys I (France) **7**, 1309 (1997).

Self-Energy Functional Theory

M. Potthoff, Eur. Phys. J. B (France) **32**, 429 (2003).

M. Potthoff, “Non-perturbative construction of the Luttinger-Ward functional”, cond-mat/0406671.

Michael Potthoff, ”Dynamical Variational Principles for Strongly Correlated Electron Systems”, cond-mat/0503715

arXiv:cond-mat/0511334 v2 6 Apr 2006

Pseudogap and high-temperature superconductivity from weak to strong coupling. Towards quantitative theory.

A.-M.S. Tremblay, B. Kyung, D. Sénéchal
Département de physique and RQMP, Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada
(Dated: November 2005)

This is a short review of the theoretical work on the two-dimensional Hubbard model performed in Sherbrooke in the last few years. It is written on the occasion of the twentieth anniversary of the discovery of high-temperature superconductivity. We discuss several approaches, how they were benchmarked and how they agree sufficiently with each other that we can trust that the results are accurate solutions of the Hubbard model. Then comparisons are made with experiment. We show that the Hubbard model does exhibit d-wave superconductivity and antiferromagnetism essentially where they are observed for both hole and electron-doped cuprates. We also show that the pseudogap phenomenon comes out of these calculations. In the case of electron-doped high temperature superconductors, comparisons with angle-resolved photoemission experiments are nearly quantitative. The value of the pseudogap temperature observed for these compounds in recent photoemission experiments has been predicted by theory before it was observed experimentally. Additional experimental confirmation would be useful. The theoretical methods that are surveyed include mostly the Two-Particle Self-Consistent Approach, Variational Cluster Perturbation Theory (or variational cluster approximation), and Cellular Dynamical Mean-Field Theory.

Contents

I. INTRODUCTION

In the first days of the discovery of high-temperature superconductivity, Anderson[1] suggested that the two-dimensional Hubbard model held the key to the phenomenon. Despite its apparent simplicity, the two-dimensional Hubbard model is a formidable challenge for theorists. The dimension is not low enough that an exact solution is available, as in one dimension. The dimension is not high enough that some mean-field theory, like Dynamical Mean Field Theory[2,3] (DMFT), valid in infinite dimension, can come to the rescue. In two dimensions, both quantum and thermal fluctuations are important. In addition, as we shall see, it turns out that the real materials are in a situation where both potential and kinetic energy are comparable. We cannot begin with the wave picture (kinetic energy dominated, or so-called "weak coupling") and do perturbation theory, and we cannot begin from the particle picture (potential energy dominated, or so-called "strong coupling") and do perturbation theory. In fact, even if one starts from the wave picture, perturbation theory is not trivial in two dimensions, as we shall see. Variational approaches on

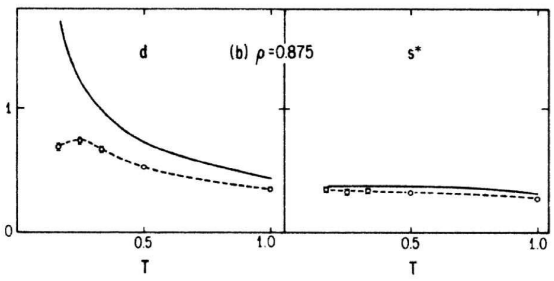

FIG. 1: The $d_{x^2-y^2}$ and extended s-wave susceptibilities obtained from QMC simulations for $U = 4t$ and a 4×4 lattice. The solid lines are the non-interacting results. From Ref. 5. The low temperature downturn of d-wave however seems to come from a mistreatment of the sign problem (D.J. Scalapino private communication).

the ground state have been proposed,[4] but even if they capture key aspects of the ground state, they say little about one-particle excitations.

Even before the discovery of high-temperature superconductivity, it was suggested that antiferromagnetic fluctuations present in the Hubbard model could lead to d-wave superconductivity,[6–8] a sort of generalization of the Kohn-Luttinger mechanism[9] analogous to the superfluidity mediated by ferromagnetic spin fluctuations in ^{3}He.[10] Nevertheless, early Quantum Monte Carlo (QMC) simulations[5] gave rather discouraging results, as illustrated in Fig. 1. In QMC, low temperatures are inaccessible because of the sign problem. At accessible temperatures, the d-wave pair susceptibility is smaller than the non-interacting one, instead of diverging. Since the observed phenomenon appears at temperatures that are about ten times smaller than what is accessible with QMC, the problem was left open. Detailed analysis of the irreducible vertex[45] deduced from QMC did suggest the importance of d-wave pairing, but other numerical work[12] concluded that long-range d-wave order is absent, despite the fact that slave-boson approaches[13,14] and many subsequent work suggested otherwise. The situation on the numerical side is changing since more recent variational,[4] Dynamical Cluster Approximation[15] and exact diagonalization[16] results now point towards the existence of d-wave superconductivity in the Hubbard model. Even more recently, new numerical approaches are making an even more convincing case.[17–19]

After twenty years, we should be as quantitative as possible. How should we proceed to investigate a model without a small parameter? We will try to follow this path: (1) Identify important physical principles and laws to constrain non-perturbative approximation schemes, starting from both weak (kinetic energy dominated) and strong (potential energy dominated) coupling. (2) Benchmark the various approaches as much as possible against exact (or numerically accurate) results. (3) Check that weak and strong coupling approaches agree at intermediate coupling. (4) Compare with experiment.

In brief, we are trying to answer the question, "Is the Hubbard model rich enough to contain the essential physics of the cuprates, both hole and electron doped?" The answer is made possible by new theoretical approaches, increased computing power, and the reassurance that theoretical approaches, numerical and analytical, give consistent results at intermediate coupling even if the starting points are very different.

This paper is a review of the work we have done in Sherbrooke on this subject. In the short space provided, this review will not cover all of our work. Needless to say, we will be unfair to the work of many other groups, even though we will try to refer to the work of others that is directly relevant to ours. We do not wish to make priority claims and we apologize to the authors that may feel unfairly treated.

Section II will introduce the methodology: First a method that is valid at weak to intermediate coupling, the Two-Particle Self-Consistent approach (TPSC), and then various quantum cluster methods that are better at strong coupling, namely Cluster Perturbation Theory (CPT), the Variational Cluster Approximation (VCA) also known as Variational Cluster Perturbation Theory (VCPT), and Cellular Dynamical Mean Field Theory (CDMFT) with a brief mention of the Dynamical Cluster Approximation (DCA). In all cases, we will mention the main comparisons with exact or numerically accurate results that have been used to benchmark the approaches. In Sect. III we give some of the results, mostly on the pseudogap and the phase diagram of high-temperature superconductors. More importantly perhaps, we show the consistency of the results obtained by both weak- and strong-coupling approaches when they are used at intermediate coupling. Finally, we compare with experiment in section IV.

II. METHODOLOGY

We consider the Hubbard model

$$H = -\sum_{i,j,\sigma} t_{ij} c^\dagger_{i\sigma} c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \qquad (1)$$

where $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) are creation and annihilation operators for electrons of spin σ, $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$ is the density of spin σ electrons, $t_{ij} = t^*_{ji}$ is the hopping amplitude, and U is the on-site Coulomb repulsion. In general, we write t, t', t'' respectively for the first-, second- and third-nearest neighbor hopping amplitudes.

In the following subsections, we first discuss how to approach the problem from the weak coupling perspective and then from the strong coupling point of view. The approaches that we will use in the end are non-perturbative, but in general they are more accurate either at weak or strong coupling.

A. Weak coupling approach

Even at weak coupling, the Hubbard model presents difficulties specific to two dimensions. The time-honored Random Phase Approximation (RPA) has the advantage of satisfying conservation laws, but it violates the Pauli principle and the Mermin-Wagner-Hohenberg-Coleman (or Mermin-Wagner, for short) theorem. This theorem states that a continuous symmetry cannot be broken at finite temperature in two dimensions. RPA gives a finite-temperature phase transition. The Pauli principle means, in particular, that $\langle n_{i\uparrow} n_{i\uparrow}\rangle = \langle n_{i\uparrow}\rangle$ in a model with only one orbital per site. This is violated by RPA since it can be satisfied only if all possible exchanges of electron lines are allowed (more on this in the following section). Since the square of the density at a given site is given by $\langle (n_{i\uparrow}+n_{i\downarrow})^2\rangle = 2\langle n_\uparrow n_\uparrow\rangle + 2\langle n_\uparrow n_\downarrow\rangle$, violating the Pauli condition $\langle n_{i\uparrow} n_{i\uparrow}\rangle = \langle n_{i\uparrow}\rangle$ will in general lead to large errors in double occupancy, a key quantity in the Hubbard model since it is proportional to the potential energy. Another popular approach is the Moriya[20] self-consistent spin-fluctuation approach[21] that uses a Hubbard-Stratonovich transformation and a $\langle\phi^4\rangle \sim \phi^2\langle\phi^2\rangle$ factorization. This satisfies the Mermin-Wagner theorem but, unfortunately, violates the Pauli principle and introduces an unknown mode-coupling constant as well as an unknown renormalized U in the second-order term. The conserving approximation known as Fluctuation Exchange (FLEX) approximation[22] is an Eliashberg-type theory that is conserving but violates the Pauli principle, assumes a Migdal theorem and does not reproduce the pseudogap phenomenon observed in QMC. More detailed criticism of this and other approaches may be found in Refs. 23,24. Finally, the renormalization group[25–29] has the great advantage of being an unbiased method to look for instabilities towards various ordered phases. However, it is quite difficult to implement in two dimensions because of the proliferation of coupling constants, and, to our knowledge, no one has yet implemented a two-loop calculation without introducing additional approximations.[30,31] Such a two-loop calculation is necessary to observe the pseudogap phenomenon.

1. *Two-Particle Self-consistent approach (TPSC)*

The TPSC approach, originally proposed by Vilk, Tremblay and collaborators,[33,37] aims at capturing nonperturbative effects. It does not use perturbation theory or, if you want, it drops diagrammatic expansions. Instead, it is based on imposing constraints and sum rules: the theory should satisfy (a) the spin and charge conservation laws (b) the Pauli principle in the form $\langle n_{i\uparrow} n_{i\uparrow}\rangle = \langle n_{i\uparrow}\rangle$ (c) the local-moment and the local-density sum rules. Without any further explicit constraint, we find that the theory satisfies the Mermin-Wagner theorem, that it satisfies consistency between one- and two-particle quantities in the sense that $\frac{1}{2}\mathrm{Tr}(\Sigma G) = U\langle n_\uparrow n_\downarrow\rangle$ and finally that the theory contains the physics of Kanamori-Brückner screening (in other words, scattering between electrons and holes includes T-matrix quantum fluctuation effects beyond the Born approximation).

Several derivations of our approach have been given,[32,33] including a quite formal one[34] based on the functional derivative Baym-Kadanoff approach.[35] Here we only give an outline[36] of the approach with a more phenomenological outlook. We proceed in two steps. In the first step (in our earlier work sometimes called zeroth step), the self-energy is obtained by a Hartree-Fock-type factorization of the four-point function with the *additional constraint* that the factorization is exact when all space-time coordinates coincide.[157] Functional differentiation, as in the Baym-Kadanoff approach[35], then leads to a momentum- and frequency-independent irreducible particle-hole vertex for the spin channel that satisfies[37] $U_{sp} = U\langle n_\uparrow n_\downarrow\rangle/(\langle n_\uparrow\rangle\langle n_\downarrow\rangle)$. The local moment sum rule and the Pauli principle in the form $\langle n_\sigma^2\rangle = \langle n_\sigma\rangle$ then determine double occupancy and U_{sp}. The irreducible vertex for the charge channel is too complicated to be computed exactly, so it is assumed to be constant and its value is found from the Pauli principle and the local charge fluctuation sum rule. To be more specific, let us use the notation, $q = (\mathbf{q}, iq_n)$ and $k = (\mathbf{k}, ik_n)$ with iq_n and ik_n respectively bosonic and fermionic Matsubara frequencies. We work in units where k_B, $\hbar$, and lattice spacing are all unity. The spin and charge susceptibilities now take the form

$$\chi_{sp}^{-1}(q) = \chi_0(q)^{-1} - \frac{1}{2}U_{sp} \tag{2}$$

and

$$\chi_{ch}^{-1}(q) = \chi_0(q)^{-1} + \frac{1}{2}U_{ch} \tag{3}$$

with χ_0 computed with the Green function $G_\sigma^{(1)}$ that contains the self-energy whose functional differentiation gave the vertices. This self-energy is constant, corresponding to the Hartree-Fock-type factorization.[158] The susceptibilities thus satisfy conservation laws[35]. One enforces the Pauli principle $\langle n_\sigma^2\rangle = \langle n_\sigma\rangle$ implicit in the following two sum rules,

$$\frac{T}{N}\sum_q \chi_{sp}(q) = \left\langle (n_\uparrow - n_\downarrow)^2\right\rangle = n - 2\langle n_\uparrow n_\downarrow\rangle \tag{4}$$

$$\frac{T}{N}\sum_q \chi_{ch}(q) = \left\langle (n_\uparrow + n_\downarrow)^2\right\rangle - n^2 = n + 2\langle n_\uparrow n_\downarrow\rangle - n^2$$

where n is the density. The above equations, in addition to[37]

$$U_{sp} = \frac{U\langle n_\uparrow n_\downarrow\rangle}{\langle n_\uparrow\rangle\langle n_\downarrow\rangle}, \tag{5}$$

suffice to determine the constant vertices U_{sp} and U_{ch}.

Once the two-particle quantities have been found as above, the next step of the approach of Ref. 23, consists in improving the approximation for the single-particle self-energy by starting from an exact expression where the high-frequency Hartree-Fock behavior is explicitly factored out. One then substitutes in the exact expression the irreducible low-frequency vertices U_{sp} and U_{ch} as well as $G^{(1)}_\sigma(k+q)$ and $\chi_{sp}(q), \chi_{ch}(q)$ computed above. The exact form for the self-energy expression can however be obtained either in the longitudinal or in the transverse channel. To satisfy crossing symmetry of the fully reducible vertex appearing in the general expression and to preserve consistency between one- and two-particle quantities, one averages the two possibilities to obtain[36]

$$\Sigma^{(2)}_\sigma(k) = U n_{-\sigma} + \frac{U}{8}\frac{T}{N}\sum_q \left[3U_{sp}\chi_{sp}(q) + U_{ch}\chi_{ch}(q)\right] G^{(1)}_\sigma(k+q). \quad (6)$$

The resulting self-energy $\Sigma^{(2)}_\sigma(k)$ on the left hand-side is at the next level of approximation so it differs from the self-energy entering the right-hand side. One can verify that the longitudinal spin fluctuations contribute an amount $U\langle n_\uparrow n_\downarrow\rangle/4$ to the consistency condition[24] $\frac{1}{2}\mathrm{Tr}(\Sigma^{(2)}G^{(1)}) = U\langle n_\uparrow n_\downarrow\rangle$ and that each of the two transverse spin components as well as the charge fluctuations also each contribute $U\langle n_\uparrow n_\downarrow\rangle/4$. In addition, one verifies numerically that the exact sum rule[23] $-\int d\omega' \,\mathrm{Im}[\Sigma_\sigma(\mathbf{k},\omega')]/\pi = U^2 n_{-\sigma}(1-n_{-\sigma})$ determining the high-frequency behavior is satisfied to a high degree of accuracy.

The theory also has a consistency check. Indeed, the exact expression for consistency between one- and two-particle quantities should be written with $G^{(2)}$ given by $(G^{-1})^{(2)} = (G^{-1})^{(0)} - \Sigma^{(2)}$ instead of with $G^{(1)}$. In other words $\frac{1}{2}\mathrm{Tr}(\Sigma^{(2)}G^{(2)}) = U\langle n_\uparrow n_\downarrow\rangle$ should be satisfied instead of $\frac{1}{2}\mathrm{Tr}(\Sigma^{(2)}G^{(1)}) = U\langle n_\uparrow n_\downarrow\rangle$, which is exactly satisfied here. We find through QMC benchmarks that when the left- and right-hand side of the last equation differ only by a few percent, then the theory is accurate.

To obtain the thermodynamics, one finds the entropy by integrating $1/T$ times the specific heat $(\partial E/\partial T)$ so that we know $F = E - TS$. There are other ways to obtain the thermodynamics and one looks for consistency between these.[38] We will not discuss thermodynamic aspects in the present review.

At weak coupling in the repulsive model the particle-hole channel is the one that is influenced directly. Correlations in crossed channels, such as pairing susceptibilities, are induced indirectly and are harder to evaluate. This simply reflects the fact the simplest Hartree-Fock factorization of the Hubbard model does not lead to a d-wave order parameter (even though Hartree-Fock factorization of its strong-coupling version does). The $d_{x^2-y^2}$-wave susceptibility is defined by $\chi_d = \int_0^\beta d\tau \langle T_\tau \Delta(\tau)\Delta^\dagger\rangle$ with the d-wave order parameter equal to $\Delta^\dagger = \sum_i \sum_\gamma g(\gamma) c^\dagger_{i\uparrow} c^\dagger_{i+\gamma\downarrow}$ the sum over γ being over nearest-neighbors, with $g(\gamma) = \pm 1/2$ depending on whether γ is a neighbor along the $\hat{\mathbf{x}}$ or the $\hat{\mathbf{y}}$ axis. Briefly speaking,[39,40] to extend TPSC to compute pairing susceptibility, we begin from the Schwinger-Martin-Kadanoff-Baym formalism with both diagonal[23,34] and off-diagonal[41] source fields. The self-energy is expressed in terms of spin and charge fluctuations and the irreducible vertex entering the Bethe-Salpeter equation for the pairing susceptibility is obtained from functional differentiation. The final expression for the d-wave susceptibility is,

$$\chi_d(\mathbf{q}=0, iq_n=0) = \frac{T}{N}\sum_k \left(g_d^2(\mathbf{k}) G^{(2)}_\uparrow(-k) G^{(2)}_\downarrow(k)\right) - \frac{U}{4}\left(\frac{T}{N}\right)^2 \sum_{k,k'} g_d(\mathbf{k}) G^{(2)}_\uparrow(-k) G^{(2)}_\downarrow(k) \times \left(\frac{3}{1-\frac{1}{2}U_{sp}\chi_0(k'-k)} + \frac{1}{1+\frac{1}{2}U_{ch}\chi_0(k'-k)}\right) G^{(1)}_\uparrow(-k') G^{(1)}_\downarrow(k') g_d(\mathbf{k}'). \quad (7)$$

In the above expression, $g_d(\mathbf{k})$ is the form factor for the gap symmetry, while k and k' stand for both wave-vector and fermionic Matsubara frequencies on a square-lattice with N sites at temperature T. The spin and charge susceptibilities take the form $\chi^{-1}_{sp}(q) = \chi_0(q)^{-1} - \frac{1}{2}U_{sp}$ and $\chi^{-1}_{ch}(q) = \chi_0(q)^{-1} + \frac{1}{2}U_{ch}$ with χ_0 computed with the Green function $G^{(1)}_\sigma$ that contains the self-energy whose functional differentiation gave the spin and charge vertices. The values of U_{sp}, U_{ch} and $\langle n_\uparrow n_\downarrow\rangle$ are obtained as described above. In the pseudogap regime, one cannot use $U_{sp} = U\langle n_\uparrow n_\downarrow\rangle/(\langle n_\uparrow\rangle\langle n_\downarrow\rangle)$. Instead,[23] one uses the local-moment sum rule with the zero temperature value of $\langle n_\uparrow n_\downarrow\rangle$ obtained by the method of Ref. 42 that agrees very well with QMC calculations at all values of U. Also, $G^{(2)}_\sigma$ contains self-energy effects coming from spin and charge fluctuations, as described above.[34,36]

The same principles and methodology can be applied for the attractive Hubbard model.[39,41,43] In that case, the dominant channel is the s-wave pairing channel. Correlations in the crossed channel, namely the spin and charge susceptibilities, can also be obtained *mutatis mutandi* along the lines of the previous paragraph.

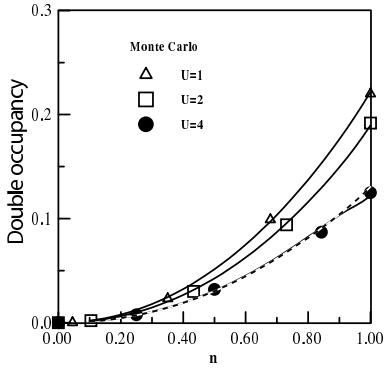

FIG. 2: Comparisons between the QMC simulations (symbols) and TPSC (solid lines) for the filling dependence of the double occupancy. The results are for $T = t/6$ as a function of filling and for various values of U expect for $U = 4t$ where the dashed line shows the results of our theory at the crossover temperature $T = T_X$. From Ref. 23.

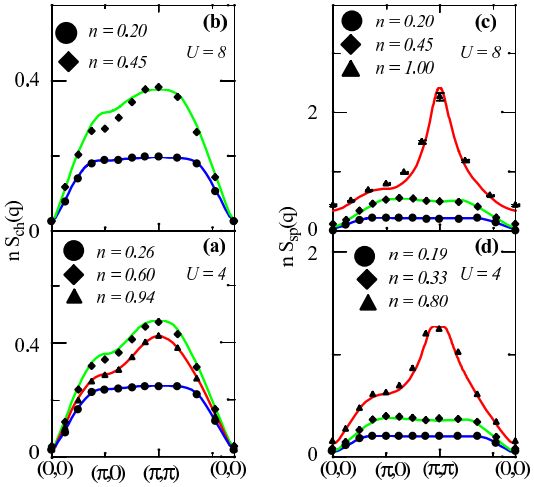

FIG. 3: Wave vector ($\mathbf{q}$) dependence of the spin and charge structure factors for different sets of parameters. Solid lines are from TPSC and symbols are our QMC data. Monte Carlo data for $n = 1$ and $U = 8t$ are for 6×6 clusters and $T = 0.5t$; all other data are for 8×8 clusters and $T = 0.2t$. Error bars are shown only when significant. From Ref. 37.

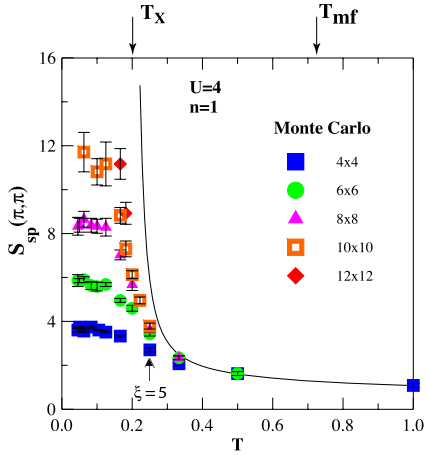

FIG. 4: Temperature dependence of $S_{sp}(\pi, \pi)$ at half-filling $n = 1$. The solid line is from TPSC and symbols are Monte Carlo data from Ref. 44. Taken from Ref. 37.

2. *Benchmarks for TPSC*

To test any non-perturbative approach, we need reliable benchmarks. Quantum Monte Carlo (QMC) simulations provide such benchmarks. The results of such numerical calculations are unbiased and they can be obtained on much larger system sizes than any other simulation method. The statistical uncertainty can be made as small as required. The drawback of QMC is that the sign problem renders calculations impossible at temperatures low enough to reach those that are relevant for d-wave superconductivity. Nevertheless, QMC can be performed in regimes that are non-trivial enough to allow us to eliminate some theories on the grounds that they give qualitatively incorrect results. Comparisons with QMC allow us to estimate the accuracy of the theory. An approach like TPSC can then be extended to regimes where QMC is unavailable with the confidence provided by agreement between both approaches in regimes where both can be performed.

In order to be concise, details are left to figure captions. Let us first focus on quantities related to spin and charge fluctuations. The symbols on the figures refer to QMC results while the solid lines come from TPSC calculations. Fig. 2 shows double occupancy, a quantity that plays a very important role in the Hubbard model in general and in TPSC in particular. That quantity is shown as a function of filling for various values of U at inverse temperature $\beta = 6$. Fig. 3 displays the spin and charge structure factors in a regime where size effects are not important. Clearly the results are non-perturbative given the large difference between the spin and charge structure factors, which are plotted here in units where they are equal at $U = 0$. In Fig. 4 we exhibit the static structure factor at half-filling as a function of temperature. Below the crossover temperature T_X, there is an important size dependence in the QMC results. The TPSC calculation, represented by a solid line, is done for an infinite system. We see that the mean-field finite transition temperature T_{MF} is replaced by a crossover temperature T_X at which the correlations enter an exponential growth regime. One can show analytically[23,37] that the correlation length becomes infinite only at zero temperature, thus satisfying the Mermin-Wagner theorem. The QMC results approach the TPSC results as the system size grows. Nevertheless, TPSC is in the $N = \infty$ universality class[46] contrary to the Hubbard model for which $N = 3$, so one expects quantitative differences to increase as the correlation length becomes larger. It is important to note that T_X does not coincide with the mean-field transition temperature T_{MF}. This is because of Kanamori-Brueckner screening[37,47] that manifests it-

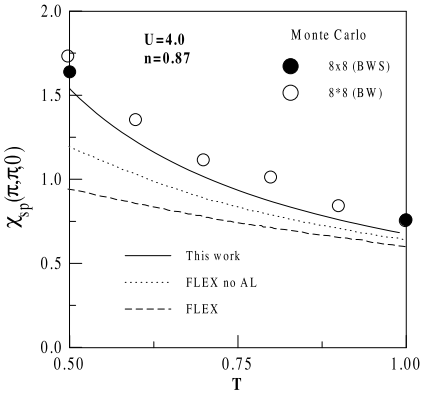

FIG. 5: Comparisons between Monte Carlo simulations (BW), FLEX calculations and TPSC for the spin susceptibility at $Q = (\pi, \pi)$ as a function of temperature at zero Matsubara frequency. The filled circles (BWS) are from Ref. 45. Taken from Ref. 23.

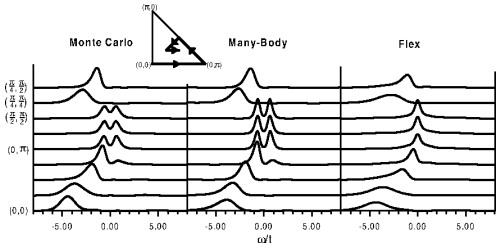

FIG. 6: Single particle spectral weight $A(\mathbf{k}, \omega)$ for $U = 4t$, $\beta = 5/t$, $n = 1$, and all independent wave vectors $\mathbf{k}$ of an 8×8 lattice. Results obtained from Maximum Entropy inversion of QMC data on the left panel and many-body TPSC calculations with Eq.(6) on the middle panel and with FLEX on the right panel. From Ref. 36.

self in the difference between U_{sp} and the bare U. Below T_X, the main contribution to the static spin structure factor in Fig. 4 comes from the zero-Matsubara frequency component of the spin susceptibility. This is the so-called renormalized classical regime where the characteristic spin fluctuation frequency ω_{sp} is much less than temperature. Even at temperatures higher than that, TPSC agrees with QMC calculation much better than other methods, as shown in Fig. 5.

Below the crossover temperature to the renormalized classical regime, a pseudogap develops in the single-particle spectral weight. This is illustrated in Fig. 6.[36] Eliashberg-type approaches such as FLEX do not show the pseudogap present in QMC. The size dependence of the results is also quite close in TPSC and in QMC, as shown in Fig. 7.

The d-wave susceptibility[40] shown in Fig. 8 again clearly demonstrates the agreement between TPSC and QMC. In particular, the dome shape dependence of the QMC results is reproduced to within a few percent. We will see in Sec. III how one understands the dome shape and the fact that the d-wave susceptibility of the interacting system is smaller than that of the non-interacting

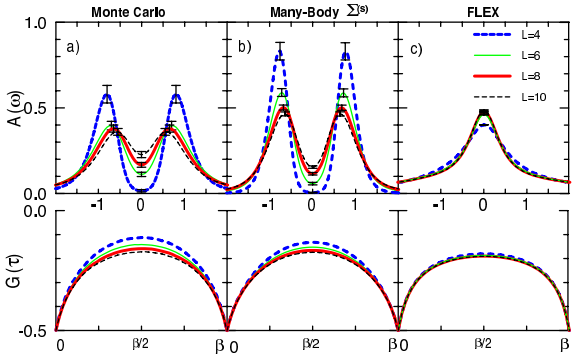

FIG. 7: Size dependent results for various types of calculations for $U = 4t$, $\beta = 5/t$, $n = 1$, $\mathbf{k} = (0, \pi)$, $L = 4, 6, 8, 10$. Upper panels show $A(\mathbf{k}, \omega)$ extracted from Maximum Entropy on $G(\tau)$ shown on the corresponding lower panels. (a) QMC. (b) TPSC using Eq. (6). (c) FLEX. From Ref. 36.

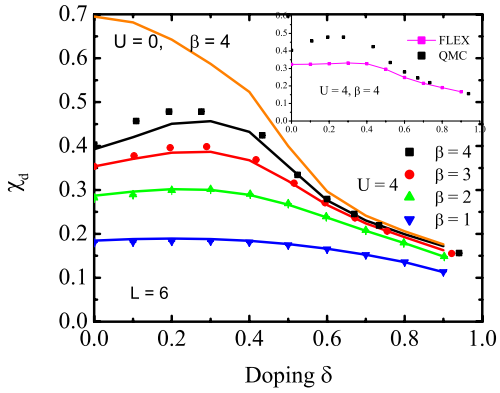

FIG. 8: Comparisons between the $d_{x^2-y^2}$ susceptibility obtained from QMC simulations (symbols) and from the TPSC approach (lines) in the two-dimensional Hubbard model. Both calculations are for $U = 4t$, a 6×6 lattice. QMC error bars are smaller than the symbols. Analytical results are joined by solid lines. The size dependence of the results is small at these temperatures. The $U = 0$ case is also shown at $\beta = 4/t$ as the upper line. The inset compares QMC and FLEX at $U = 4, \beta = 4/t$. From Ref. 40.

one in this temperature range.

To conclude this section, we quickly mention a few other results obtained with TPSC. Fig. 9 contrasts the crossover phase diagram obtained for the Hubbard model at the van Hove filling[48] with the results of a renormalization group calculation.[28] The difference occurring in the ferromagnetic region is discussed in detail in Ref. 48. Finally, we point out various comparisons for the attractive Hubbard model. Fig. 10 shows the static s-wave pairing susceptibility, Fig. 11 the chemical potential and the occupation number, and finally Fig. 12 the local density of states and the single-particle spectral weight at a given wave vector.

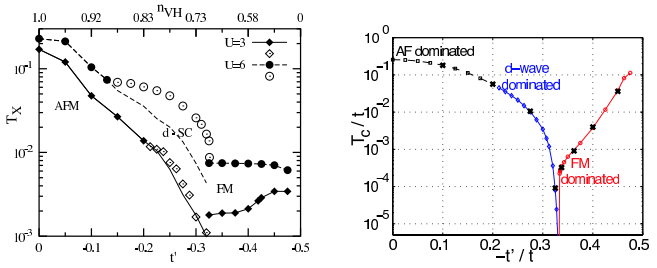

FIG. 9: The crossover diagram as a function of next-nearest-neighbor hopping t' from TPSC (left) and from a temperature cutoff renormalization group technique from Ref. 28 (right). The corresponding Van Hove filling is indicated on the upper horizontal axis. Crossover lines for magnetic instabilities near the antiferromagnetic and ferromagnetic wave vectors are represented by filled symbols while open symbols indicate instability towards $d_{x^2-y^2}$-wave superconducting. The solid and dashed lines below the empty symbols show, respectively for $U = 3t$ and $U = 6t$, where the antiferromagnetic crossover temperature would have been in the absence of the superconducting instability. The largest system size used for this calculation is 2048×2048. From Ref. 48.

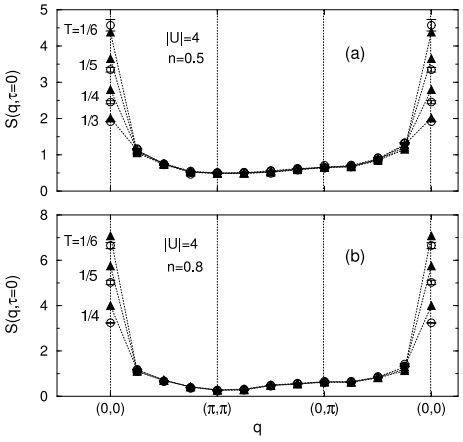

FIG. 10: TPSC s-wave paring structure factor $S(\mathbf{q}, \tau = 0)$ (filled triangles) and QMC $S(\mathbf{q}, \tau = 0)$ (open circles) for $U = -4t$ and various temperatures (a) at $n = 0.5$ and (b) at $n = 0.8$ on a 8×8 lattice. The dashed lines are to guide the eye. From Ref. 43.

B. Strong-coupling approaches: Quantum clusters

DMFT[3,51] has been extremely successful in helping us understand the Mott transition, the key physical phenomenon that manifests itself at strong coupling. However, in high dimension where this theory becomes exact, spatial fluctuations associated with incipient order do not manifest themselves in the self-energy. In low dimension, this is not the case. The self-energy has strong momentum dependence, as clearly shown experimentally in the high-temperature superconductors, and theoretically in the TPSC approach, a subject we shall discuss again below. It is thus necessary to go beyond DMFT by studying clusters instead of a single Anderson impu-

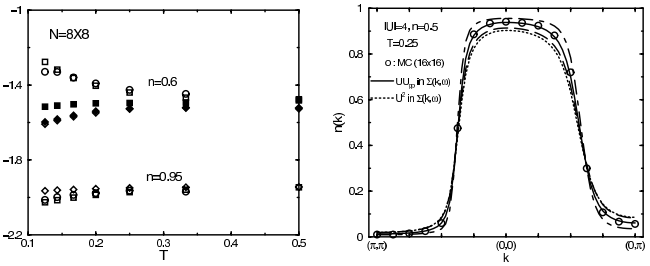

FIG. 11: Left: chemical potential shifts $\mu^{(1)} - \mu_0$ (open diamonds) and $\mu^{(2)} - \mu_0$ (open squares) with the results of QMC calculations (open circles) for $U = -4t$. Right: The momentum dependent occupation number $n(\mathbf{k})$. Circles: QMC calculations from Ref. 49. The solid curve: TPSC. The dashed curve obtained by replacing U_{pp} by U in the self-energy with all the rest unchanged. The long-dash line is the result of a self-consistent T-matrix calculation, and the dot-dash line the result of second-order perturbation theory. From Ref. 43.

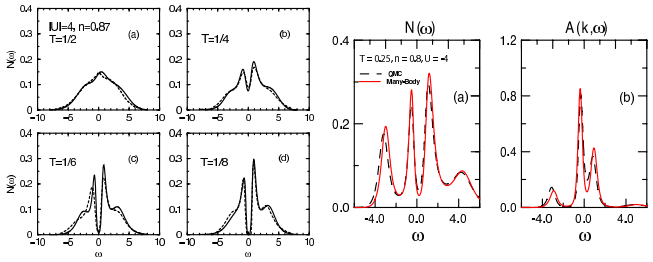

FIG. 12: Comparisons of local density of states and single-particle spectral weight from TPSC (solid lines) and QMC (dashed lines) on a 8×8 lattice. QMC data for the density of states taken from Ref. 50. Figures from Ref. 43.

rity as done in DMFT. The simplest cluster approach that includes strong-coupling effects and momentum dependence is Cluster Perturbation Theory (CPT).[52,53] In this approach, an infinite set of disconnected clusters are solved exactly and then connected to each other using strong-coupling perturbation theory. Although the resulting theory turns out to give the exact result in the $U = 0$ case, its derivation clearly shows that one expects reliable results mostly at strong coupling. This approach does not include the self-consistent effects contained in DMFT. Self-consistency or clusters was suggested in Ref. 2,54 and a causal approach was first implemented within DCA,[55] where a momentum-space cluster is connected to a self-consistent momentum-space medium. In our opinion, the best framework to understand all other cluster methods is the Self-Energy Functional approach of Potthoff.[56,57] The form of the lattice Green function obtained in this approach is the same as that obtained in CPT, clearly exhibiting that such an approach is better at strong-coupling, even though results often extrapolate correctly to weak coupling. Amongst the special cases of this approach, the Variational Cluster Approach (VCA), or Variational Cluster Perturbation Theory (VCPT)[57] is the one closest to the original approach. In a variant,

Cellular Dynamical Mean Field Theory[58] (CDMFT), a cluster is embedded in a self-consistent medium instead of a single Anderson impurity as in DMFT (even though the latter approach is accurate in many realistic cases, especially in three dimensions). The strong-coupling aspects of CDMFT come out clearly in Refs. 59,60. A detailed review of quantum cluster methods has appeared in Ref. 61.

1. Cluster perturbation theory

Even though CPT does not have the self-consistency present in DMFT type approaches, at fixed computing resources it allows for the best momentum resolution. This is particularly important for the ARPES pseudogap in electron-doped cuprates that has quite a detailed momentum space structure, and for d-wave superconducting correlations where the zero temperature pair correlation length may extend well beyond near-neighbor sites. CPT was developed by Gros[52] and Sénéchal[53] independently. This approach can be viewed as the first term of a systematic expansion around strong coupling.[62] Let us write the hopping matrix elements in the form

$$t_{\mu\nu}^{mn} = t_{\mu\nu}^{(c)}\delta_{mn} + V_{\mu\nu}^{mn} \tag{8}$$

where m and n label the different clusters, and μ, ν label the sites within a cluster. Then $t_{\mu\nu}^{(c)}$ labels all the hopping matrix elements within a cluster and the above equation defines $V_{\mu\nu}^{mn}$.

We pause to introduce the notation that will be used throughout for quantum cluster methods. We follow the review article Ref. 61. In reciprocal space, any wave vector $\mathbf{k}$ in the Brillouin zone may be written as $\mathbf{k} = \tilde{\mathbf{k}} + \mathbf{K}$ where both $\mathbf{k}$ and $\tilde{\mathbf{k}}$ are continuous in the infinite size limit, except that $\tilde{\mathbf{k}}$ is defined only in the reduced Brillouin zone that corresponds to the superlattice. On the other hand, $\mathbf{K}$ is discrete and denotes reciprocal lattice vectors of the superlattice. By analogy, any position $\mathbf{r}$ in position space can be written as $\tilde{\mathbf{r}} + \mathbf{R}$ where $\mathbf{R}$ is for positions within clusters while $\tilde{\mathbf{r}}$ labels the origins of the clusters, an infinite number of them. Hence, Fourier's theorem allows one to define functions of $\mathbf{k}$, $\tilde{\mathbf{k}}$ or $\mathbf{K}$ that contain the same information as functions of, respectively, $\mathbf{r}$, $\tilde{\mathbf{r}}$ or $\mathbf{R}$. Also, we have $\mathbf{K}\cdot\tilde{\mathbf{r}} = 2\pi n$ where n is an integer. Sites within a cluster are labelled by greek letters so that the position of site μ within a cluster is $\mathbf{R}_\mu$, while clusters are labelled by Latin letters so that the origin of cluster m is at $\tilde{\mathbf{r}}_m$.

Returning to CPT, the Green function for the whole system is given by

$$\left[\hat{G}^{-1}(\tilde{\mathbf{k}}, z)\right]_{\mu\nu} = \left[\hat{G}^{(c)-1}(z) - \hat{V}(\tilde{\mathbf{k}})\right]_{\mu\nu} \tag{9}$$

where hats denote matrices in cluster site indices and z is the complex frequency. At this level of approximation, the CPT Green function has the same structure as in the Hubbard I approximation except that it pertains to a cluster instead of a single site. Since $\hat{G}^{(c)-1}(z) = z + \mu - \hat{t}^{(c)} - \hat{\Sigma}^{(c)}$ and $\hat{G}^{(0)-1}(\tilde{\mathbf{k}}, z) = z + \mu - \hat{t}^{(c)} - \hat{V}(\tilde{\mathbf{k}})$, the Green function (9) may also be written as

$$\hat{G}^{-1}(\tilde{\mathbf{k}}, z) = \hat{G}^{(0)-1}(\tilde{\mathbf{k}}, z) - \hat{\Sigma}^{(c)}(z). \tag{10}$$

This form allows a different physical interpretation of the approach. In the above expression, the self-energy of the lattice is approximated by the self-energy of the cluster. The latter in real space spans only the size of the cluster.

We still need an expression to extend the above result to the lattice in a translationally invariant way. This is done by defining the following residual Fourier transform:

$$G_{\mathrm{CPT}}(\mathbf{k}, z) = \frac{1}{N_c}\sum_{\mu,\nu}^{N_c} e^{i\mathbf{k}\cdot(\mathbf{R}_\mu - \mathbf{R}_\nu)} G_{\mu\nu}(\tilde{\mathbf{k}}, z). \tag{11}$$

Notice that $G_{\mu\nu}(\tilde{\mathbf{k}}, z)$ may be replaced by $G_{\mu\nu}(\mathbf{k}, z)$ in the above equation since $\hat{V}(\tilde{\mathbf{k}} + \mathbf{K}) = \hat{V}(\tilde{\mathbf{k}})$.

2. Self-energy functional approach

The self-energy functional approach, devised by Potthoff[57] allows one to consider various cluster schemes from a unified point of view. It begins with $\Omega_{\mathbf{t}}[G]$, a functional of the Green function

$$\Omega_{\mathbf{t}}[G] = \Phi[G] - \mathrm{Tr}((G_{0\mathbf{t}}^{-1} - G^{-1})G) + \mathrm{Tr}\ln(-G). \tag{12}$$

The Luttinger Ward functional $\Phi[G]$ entering this equation is the sum of connected vacuum skeleton diagrams. A diagram-free definition of this functional is also given in Ref. 63. For our purposes, what is important is that (1) The functional derivative of $\Phi[G]$ is the self-energy

$$\frac{\delta\Phi[G]}{\delta G} = \Sigma \tag{13}$$

and (2) it is a universal functional of G in the following sense: whatever the form of the one-body Hamiltonian, it depends only on the interaction and, functionnally, it has the same dependence on G. The dependence of the functional $\Omega_{\mathbf{t}}[G]$ on the one-body part of the Hamiltonian is denoted by the subscript $\mathbf{t}$ and it comes only through $G_{0\mathbf{t}}^{-1}$ appearing on the right-hand side of Eq. (12).

The functional $\Omega_{\mathbf{t}}[G]$ has the important property that it is stationary when G takes the value prescribed by Dyson's equation. Indeed, given the last two equations, the Euler equation takes the form

$$\frac{\delta\Omega_{\mathbf{t}}[G]}{\delta G} = \Sigma - G_{0\mathbf{t}}^{-1} + G^{-1} = 0. \tag{14}$$

This is a dynamic variational principle since it involves the frequency appearing in the Green function, in other words excited states are involved in the variation. At this stationary point, and only there, $\Omega_{\mathbf{t}}[G]$ is equal to the

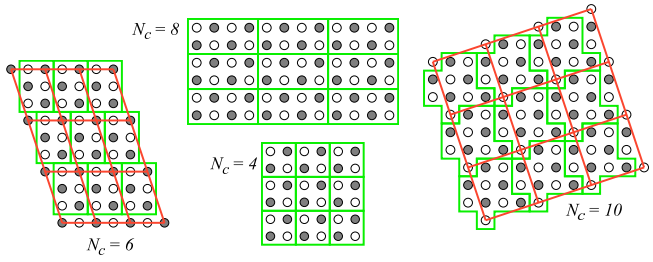

FIG. 13: Various tilings used in quantum cluster approaches. In these examples the grey and white sites are inequivalent since an antiferromagnetic order is possible.

grand potential. Contrary to Ritz's variational principle, this last equation does not tell us whether $\Omega_{\mathbf{t}}[G]$ is a minimum or a maximum or a saddle point there.

There are various ways to use the stationarity property that we described above. The most common one, is to approximate $\Phi[G]$ by a finite set of diagrams. This is how one obtains the Hartree-Fock, the FLEX approximation[22] or other so-called thermodynamically consistent theories. This is what Potthoff calls a type II approximation strategy.[64] A type I approximation simplifies the Euler equation itself. In a type III approximation, one uses the exact form of $\Phi[G]$ but only on a limited domain of trial Green functions.

Following Potthoff, we adopt the type III approximation on a functional of the self-energy instead of on a functional of the Green function. Suppose we can locally invert Eq. (13) for the self-energy to write G as a functional of Σ. We can use this result to write,

$$\Omega_{\mathbf{t}}[\Sigma] = F[\Sigma] - \mathrm{Tr}\ln(-G_{0\mathbf{t}}^{-1} + \Sigma). \tag{15}$$

where we defined

$$F[\Sigma] = \Phi[G] - \mathrm{Tr}(\Sigma G). \tag{16}$$

and where it is implicit that $G = G[\Sigma]$ is now a functional of Σ. $F[\Sigma]$, along with the expression (13) for the derivative of the Luttinger-Ward functional, define the Legendre transform of the Luttinger-Ward functional. It is easy to verify that

$$\frac{\delta F[\Sigma]}{\delta \Sigma} = \frac{\delta \Phi[G]}{\delta G}\frac{\delta G[\Sigma]}{\delta \Sigma} - \Sigma\frac{\delta G[\Sigma]}{\delta \Sigma} - G = -G \tag{17}$$

hence, $\Omega_{\mathbf{t}}[\Sigma]$ is stationary with respect to Σ when Dyson's equation is satisfied

$$\frac{\delta \Omega_{\mathbf{t}}[\Sigma]}{\delta \Sigma} = -G + (G_{0\mathbf{t}}^{-1} - \Sigma)^{-1} = 0. \tag{18}$$

To perform a type III approximation on $F[\Sigma]$, we take advantage that it is universal, i.e., that it depends only on the interaction part of the Hamiltonian and not on the one-body part. This follows from the universal character of its Legendre transform $\Phi[G]$. We thus evaluate $F[\Sigma]$ exactly for a Hamiltonian H' that shares the same interaction part as the Hubbard Hamiltonian, but that is exactly solvable. This Hamiltonian H' is taken as a cluster decomposition of the original problem, i.e., we tile the infinite lattice into identical, disconnected clusters that can be solved exactly. Examples of such tilings are given in Fig. 13. Denoting the corresponding quantities with a prime, we obtain,

$$\Omega_{\mathbf{t}'}[\Sigma'] = F[\Sigma'] - \mathrm{Tr}\ln(-G_{0\mathbf{t}'}^{-1} + \Sigma'). \tag{19}$$

from which we can extract $F[\Sigma']$. It follows that

$$\Omega_{\mathbf{t}}[\Sigma'] = \Omega_{\mathbf{t}'}[\Sigma'] + \mathrm{Tr}\ln(-G_{0\mathbf{t}'}^{-1} + \Sigma') - \mathrm{Tr}\ln(-G_{0\mathbf{t}}^{-1} + \Sigma'). \tag{20}$$

The type III approximation comes from the fact that the self-energy Σ' is restricted to the exact self-energy of the cluster problem H', so that variational parameters appear in the definition of the one-body part of H'.

In practice, we look for values of the cluster one-body parameters $\mathbf{t}'$ such that $\delta\Omega_{\mathbf{t}}[\Sigma']/\delta\mathbf{t}' = 0$. It is useful for what follows to write the latter equation formally, although we do not use it in actual calculations. Given that $\Omega_{\mathbf{t}'}[\Sigma']$ is the actual grand potential evaluated for the cluster, $\partial\Omega_{\mathbf{t}'}[\Sigma']/\partial\mathbf{t}'$ is canceled by the explicit $\mathbf{t}'$ dependence of $\mathrm{Tr}\ln(-G_{0\mathbf{t}'}^{-1} + \Sigma')$ and we are left with

$$\begin{aligned} 0 &= \frac{\delta\Omega_{\mathbf{t}}[\Sigma']}{\delta\Sigma'}\frac{\delta\Sigma'}{\delta\mathbf{t}'} \\ &= -\mathrm{Tr}\left[\left(\frac{1}{G_{0\mathbf{t}'}^{-1} - \Sigma'} - \frac{1}{G_{0\mathbf{t}}^{-1} - \Sigma'}\right)\frac{\delta\Sigma'}{\delta\mathbf{t}'}\right]. \end{aligned} \tag{21}$$

Given that the clusters corresponding to $\mathbf{t}'$ are disconnected and that translation symmetry holds on the superlattice of clusters, each of which contains N_c sites, the last equation may be written

$$\begin{aligned} &\sum_{\omega_n}\sum_{\mu\nu}\left[\frac{N}{N_c}\left(\frac{1}{G_{0\mathbf{t}'}^{-1} - \Sigma'(i\omega_n)}\right)_{\mu\nu}\right. \\ &\left. - \sum_{\tilde{\mathbf{k}}}\left(\frac{1}{G_{0\mathbf{t}}^{-1}(\tilde{\mathbf{k}}) - \Sigma'(i\omega_n)}\right)_{\mu\nu}\right]\frac{\delta\Sigma'_{\nu\mu}(i\omega_n)}{\delta\mathbf{t}'} = 0. \end{aligned} \tag{22}$$

3. *Variational cluster perturbation theory, or variational cluster approximation*

In Variational Cluster Perturbation Theory (VCPT), more aptly named the Variational Cluster Approach (VCA), solutions to the Euler equations (22) are found by looking for numerical minima (or more generally, saddle-points) of the functional. Typically, the VCA cluster Hamiltonian H' will have the same form as H except that there is no hopping between clusters and that long-range order is allowed by adding some Weiss fields, for instance like in Eq. (37) below. The hopping terms and chemical potential within H' may also be treated like additional variational parameters. In contrast with Mean-Field theory, these Weiss fields are not mean fields, in

the sense that they do not coincide with the corresponding order parameters. The interaction part of H (or H') is not factorized in any way and short-range correlations are treated exactly. In fact, the Hamiltonian H is not altered in any way; the Weiss fields are introduced to let the variational principle act on a space of self-energies that includes the possibility of specific long-range orders, without imposing those orders. Indeed, the more naturally an order arises in the system, the smaller the Weiss field needs to be, and one observes that the strength of the Weiss field at the stationary point of the self-energy functional generally decreases with increasing cluster size, as it should since in the thermodynamic limit no Weiss field should be necessary to establish order.

4. Cellular dynamical mean-field theory

The Cellular dynamical mean-field theory (CDMFT) is obtained by including in the cluster Hamiltonian H' a bath of uncorrelated electrons that somehow must mimic the effect on the cluster of the rest of the lattice. Explicitly, H' takes the form

$$H' = -\sum_{\mu,\nu,\sigma} t'_{\mu\nu} c^\dagger_{\mu\sigma} c_{\nu\sigma} + U \sum_{\mu} n_{\mu\uparrow} n_{\mu\downarrow} + \sum_{\mu,\alpha,\sigma} V_{\mu\alpha} (c^\dagger_{\mu\sigma} a_{\alpha\sigma} + \text{H.c.}) + \sum_{\alpha} \epsilon_\alpha a^\dagger_{\alpha\sigma} a_{\alpha\sigma} \quad (23)$$

where $a_{\alpha\sigma}$ annihilates an electron of spin σ on a bath orbital labelled α. The bath is characterized by the energy of each orbital (ϵ_α) and the bath-cluster hybridization matrix $V_{\mu\alpha}$. This representation of the environment through an Anderson impurity model was introduced in Ref. 65 in the context of DMFT (i.e., a single site). The effect of the bath on the electron Green function is encapsulated in the so-called hybridization function

$$\Gamma_{\mu\nu}(\omega) = \sum_{\alpha} \frac{V_{\mu\alpha} V^*_{\nu\alpha}}{\omega - \epsilon_\alpha} \quad (24)$$

which enters the Green function as

$$[G'^{-1}]_{\mu\nu} = \omega + \mu - t'_{\mu\nu} - \Gamma_{\mu\nu}(\omega) - \Sigma_{\mu\nu}(\omega). \quad (25)$$

Moreover, the CDMFT does not look for a strict solution of the Euler equation (22), but tries instead to set each of the terms between brackets to zero separately. Since the Euler equation (22) can be seen as a scalar product, CDMFT requires that the modulus of one of the vectors vanish to make the scalar product vanish. From a heuristic point of view, it is as if each component of the Green function in the cluster were equal to the corresponding component deduced from the lattice Green function. This clearly reduces to single site DMFT when there is only one lattice site.

When the bath is discretized, i.e., is made of a finite number of bath "orbitals", the left-hand side of Eq. (22) cannot vanish separately for each frequency, since the number of degrees of freedom in the bath is insufficient. Instead, one adopts the following self-consistent scheme: (1) one starts with a guess value of the bath parameters $(V_{\mu\alpha}, \epsilon_\alpha)$ and solves the cluster Hamiltonian H' numerically. (2) One then calculates the combination

$$\hat{\mathcal{G}}_0^{-1} = \left[\sum_{\tilde{\mathbf{k}}} \frac{1}{\hat{G}^{-1}_{0\mathbf{t}}(\tilde{\mathbf{k}}) - \hat{\Sigma}'(i\omega_n)} \right]^{-1} + \hat{\Sigma}'(i\omega_n) \quad (26)$$

and (3) minimizes the following canonically invariant distance function:

$$d = \sum_{n,\mu,\nu} \left| \left(i\omega_n + \mu - \hat{t}' - \hat{\Gamma}(i\omega_n) - \hat{\mathcal{G}}_0^{-1} \right)_{\mu\nu} \right|^2 \quad (27)$$

over the set of bath parameters (changing the bath parameters at this step does not require a new solution of the Hamiltonian H', but merely a recalculation of the hybridization function $\hat{\Gamma}$). The bath parameters obtained from this minimization are then put back into step (1) and the procedure is iterated until convergence.

In practice, the distance function (27) can take various forms, for instance by adding a frequency-dependent weight in order to emphasize low-frequency properties[17,59,66] or by using a sharp frequency cutoff.[67] These weighting factors can be considered as rough approximations for the missing factor $\delta\Sigma'_{\nu\mu}(i\omega_n)/\delta\mathbf{t}'$ in the Euler equation (22). The frequencies are summed over on a discrete, regular grid along the imaginary axis, defined by some fictitious inverse temperature β, typically of the order of 20 or 40 (in units of t^{-1}). Even when the total number of cluster plus bath sites in CDMFT equals the number of sites in a VCA calculation, CDMFT is much faster than the VCA since the minimization of a grand potential functional requires many exact diagonalizations of the cluster Hamiltonian H'.

The final lattice Green function from which one computes observable quantities may be obtained by periodizing the self-energy, as in Ref. 58 or in the CPT manner described above in Eq. (11). We prefer the last approach because it corresponds to the Green function needed to obtain the density from $\partial\Omega/\partial\mu = -\text{Tr}(G)$ and also because periodization of the self-energy gives additional unphysical states in the Mott gap[68] (see also Ref. 60).

5. The Dynamical cluster approximation

The DCA[55] cannot be formulated within the self-energy functional approach.[159] It is based on the idea of discretizing irreducible quantities, such as the energy, in reciprocal space. It is believed to converge faster for $\mathbf{q} = \mathbf{0}$ quantities whereas CDMFT converges exponentially fast for local quantities.[69–71]

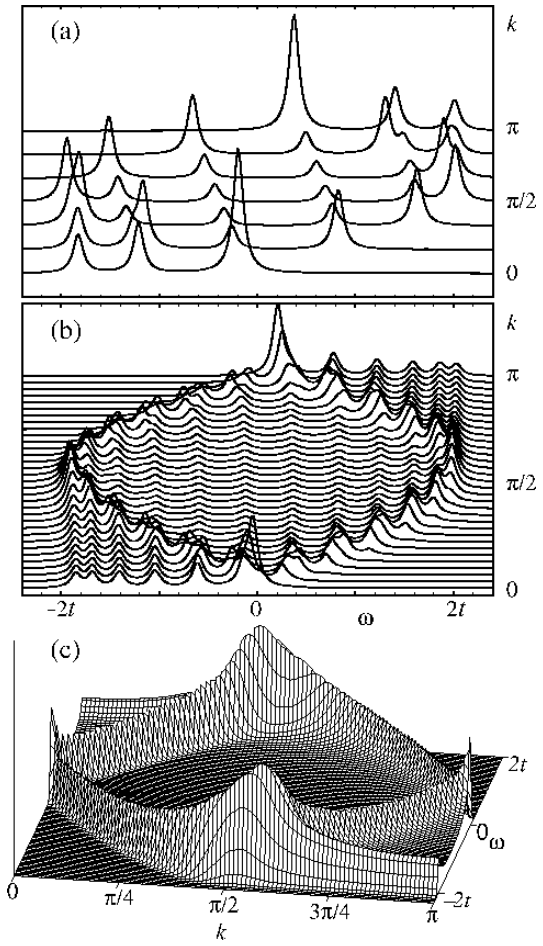

FIG. 14: The spectral function of the $U \to \infty$ limit of the one- dimensional Hubbard model, as calculated from (a) an exact diagonalization of the Hubbard model with $U/t = 100$ on a periodic 12- site cluster; (b) the same, but with CPT, on a 12-site cluster with open boundary conditions; (c) the exact solution, taken from Ref. 72; beware: the axes are oriented differently. In (a) and (b) a finite width η has been given to peaks that would otherwise be Dirac δ-functions.

6. *Benchmarks for quantum cluster approaches*

Since DMFT becomes exact in infinite dimension, the most difficult challenge for cluster extensions of this approach is in one dimension. In addition, exact results to compare with exist only in one dimension so it is mostly in $d = 1$ that cluster methods have been checked. In $d = 2$ there have also been a few comparisons with QMC as we shall discuss.

CPT has been checked[68] for example by comparing with exact results[72] for the spectral function at $U \to \infty$ in $d = 1$ as shown in Fig. 14. Fig. 15 shows the chemical potential as a function of density for various values of U. Fig. 16 shows the convergence rates for the total energy and for the double occupancy in the $d = 1$ half-filled model. Clearly, there is a dramatic improvement compared with exact diagonalizations.

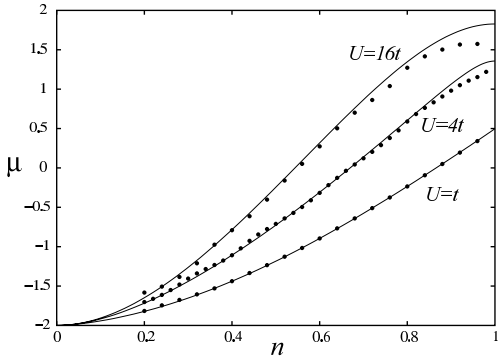

FIG. 15: Chemical potential as a function of density in the one-dimensional Hubbard model, as calculated by CPT (from Ref. 62). The exact, Bethe-Ansatz result is shown as a solid line.

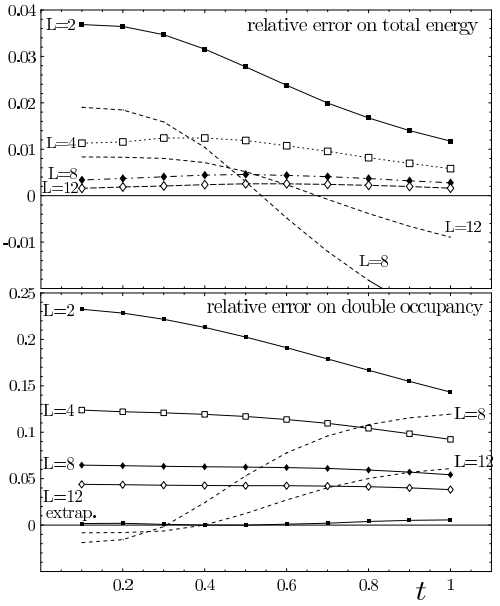

FIG. 16: Top: Comparison (expressed in relative difference) between the ground-state energy density of the half-filled, one-dimensional Hubbard model calculated from the exact, Bethe-Ansatz result. The results are displayed as a function of the hopping t, for $U = 2t$ and various cluster sizes L (connected symbols). For comparison, the exact diagonalization values of finite clusters with periodic boundary conditions are also shown (dashed lines) for $L = 8$ and $L = 12$. Bottom: Same for the double occupancy. An extrapolation of the results to infinite cluster size ($L \to \infty$) using a quadratic fit in terms of $1/L$ is also shown, and is accurate to within 0.5%. Taken from Ref. 62.

The main weakness of CPT is that it cannot take into account tendency towards long-range order. This is remedied by VCPT, as shown in Fig. 17 where CPT, VCPT are both compared with QMC as a benchmark. Despite this agreement, we should stress that long wave length fluctuations are clearly absent from cluster approaches. Hence, the antiferromagnetic order parameter at half-filling, for example, does not contain the effect of

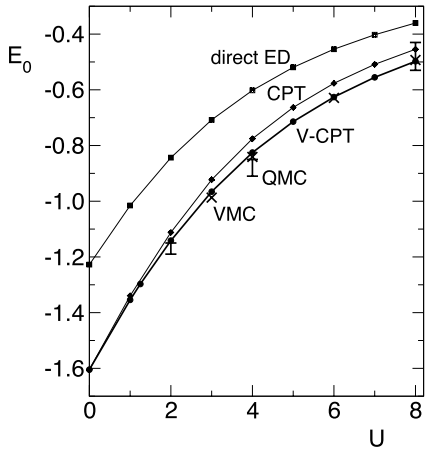

FIG. 17: Ground state energy of the half-filled, two-dimensional Hubbard model ($t = 1$) as a function of U, as obtained from various methods: Exact diagonalization (ED), CPT and VCPT on a 10-site cluster, quantum Monte Carlo (QMC) and variational Monte Carlo (VMC). Taken from Ref. 73.

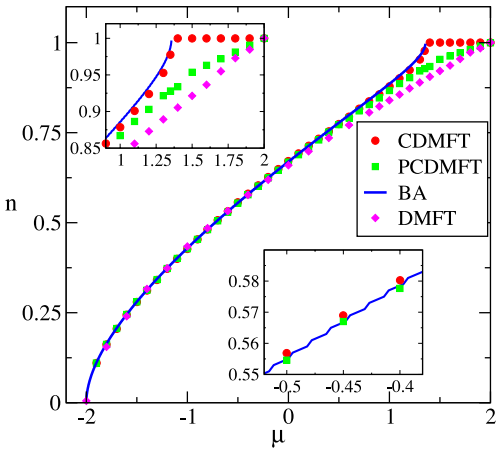

FIG. 18: CDMFT calculation on a 2×2 cluster with 8 bath sites of the density as a function of the chemical potential in the one-dimensional Hubbard model for $U = 4t$, as compared with the exact solution, DMFT and other approximation schemes. Taken from Ref. 74.

zero-point long wave length transverse spin fluctuations. This is discussed for example in the context of Fig. 9 of Ref. 73.

CDMFT corrects the difficulties of CPT near half-filling by reproducing the infinite compressibility predicted by the Bethe ansatz in one dimension as shown in Fig. 18.[74] Detailed comparisons between the local and near-neighbor Green functions[66,74] have been performed. One should note that these results, obtained from exact diagonalization, also need the definition of a distance function (See Eq. (27) above) that helps find the best bath parametrization to satisfy the self-consistency condition. This measure forces one to define calculational parameters such as a frequency cutoff and an fictitious temperature defining the Matsubara frequencies to sum over. The final results are not completely insensitive to the choice of fictitious temperature or frequency-weighing scheme but are usually considered reliable and consistent with each other when β lies between 20 and 40. The cut-off procedures have been discussed in Ref. 67.

The relative merits of DCA and CDMFT have been discussed for example in Refs 69–71,75,76. Briefly speaking, convergence seems faster in DCA for long wave length quantities but CDMFT is faster (exponentially) for local quantities.

III. RESULTS AND CONCORDANCE BETWEEN DIFFERENT METHODS

In this section, we outline the main results we obtained concerning the pseudogap and d-wave superconductivity in the two-dimensional Hubbard model. Quantum cluster approaches are better at strong coupling while TPSC is best at weak coupling. Nevertheless, all these methods are non-perturbative, the intermediate coupling regime presenting the physically most interesting case. But it is also the regime where we have the least control over the approximations. As we will show, it is quite satisfying that, at intermediate coupling, weak-coupling and strong-coupling approaches give results that are nearly in quantitative agreement with each other. This gives us great confidence into the validity of the results. As an example of concordance, consider the fact that to obtain spectral weight near $(\pi/2, \pi/2)$ at optimal doping in the electron-doped systems, U has to be roughly $6t$. For larger U, ($U = 8t$ in CPT) that weight, present in the experiments, disappears. Smaller values of U ($U = 4t$ in CPT) do not lead to a pseudogap. Other examples of concordance include the value of the superconducting transition temperature T_c obtained with DCA and with TPSC as well as the temperature dependence of souble occupancy obtained with the same two methods.

A. Weak and strong-coupling pseudogap

To understand the pseudogap it is most interesting to consider both hole and electron-doped cuprates at once. This means that we have to include particle-hole symmetry breaking hoppings, t' and t''. We will see in the present section that it is possible to obtain a pseudogap at strong coupling without a large correlation length in the particle-hole or in the particle-particle channels. By contrast, at weak coupling one does need a long-correlation length and low dimension. So there appears to be theoretically two different mechanisms for the pseudogap.

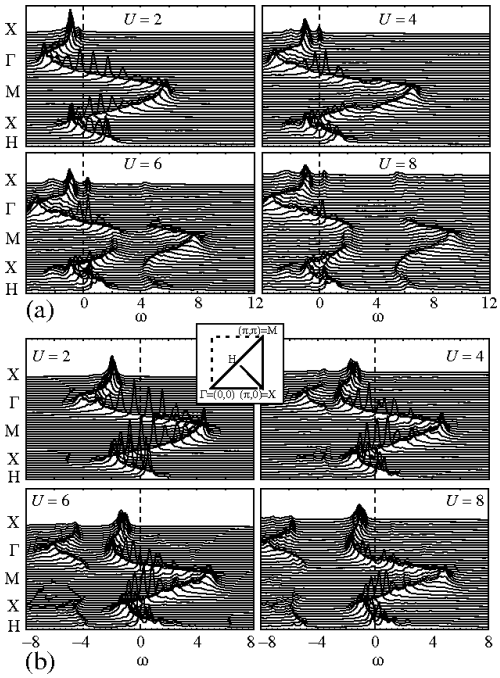

FIG. 19: Single particle spectral weight, as a function of energy ω in units of t, for wave vectors along the high-symmetry directions shown in the inset. (a) CPT calculations on a 3×4 cluster with ten electrons (17% hole doped). (b) the same as (a), with 14 electrons (17% electron doped). In all cases $t' = -0.3t$ and $t'' = 0.2t$. A Lorentzian broadening $\eta = 0.12t$ is used to reveal the otherwise delta peaks. From Ref. 77.

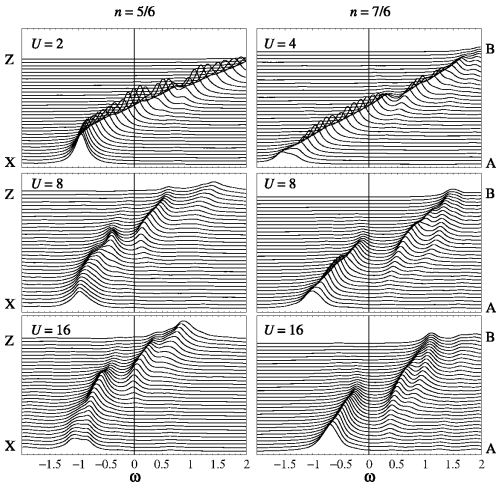

FIG. 20: Onset of the pseudogap as a function of U corresponding to Fig. 19, taken from Ref. 77. Hole-doped case on the left, electron-doped case on the right panel

1. CPT

The top panel in Fig. 19 presents the single-particle spectral weight, $A(\mathbf{k}, \omega)$ or imaginary part of the single-particle Green function, for the model with $t' = -0.3t$, $t'' = 0.2t$ in the hole-doped case, for about 17% doping.[77]

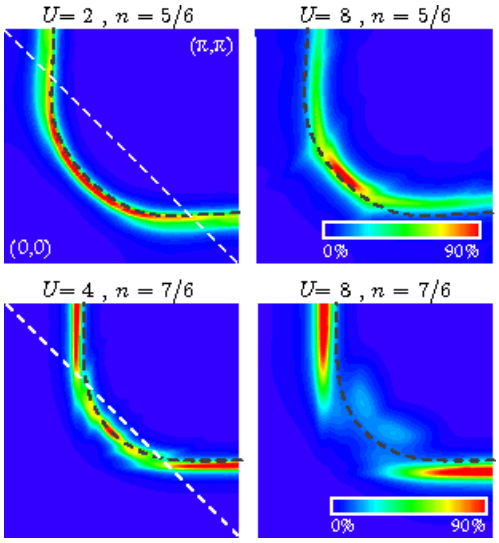

FIG. 21: MDC from CPT in the t-t' $= -0.3t$, $t'' = 0.2t$ Hubbard model, taken from Ref. 77.

Each curve is for a different wave vector (on a trajectory shown in the inset) and is plotted as a function of frequency in units of t. This kind of plot is known as Energy Dispersion Curves (EDC). It is important to point out that the theoretical results are obtained by broadening a set of delta function, so that the energy resolution is $\eta = 0.12t$ corresponding roughly to the experimental resolution we will compare with in the next section. At small $U = 2t$ on the top panel of Fig. 19, one recovers a Fermi liquid. At large U, say $U = 8t$, the Mott gap at positive energy is a prominent feature. The pseudogap is a different feature located around the Fermi energy. To see it better, we present on the left-hand panel of Fig. 20 a blow-up in the vicinity of the Fermi surface crossing occurring near $(\pi, 0)$. Clearly, there is a minimum in $A(\mathbf{k}, \omega)$ at the Fermi-surface crossing when U is large enough instead of a maximum like in Fermi liquid theory.

It is also possible to plot $A(\mathbf{k}, \omega)$ at fixed frequency for various momenta. They are so-called Momentum Dispersion Curves (MDC). In Fig. 21 we take the Fermi energy $\omega = 0$, and we plot the magnitude of the single-particle spectral weight in the first quadrant of the Brillouin zone using red for high-intensity and blue for low intensity. The figure shows that, in the hole-doped case (top panel), weight near $(\pi/2, \pi/2)$ survives while it tends to disappear near $(\pi, 0)$ and $(0, \pi)$. That pseudogap phenomenon is due not only to large U but also to the fact that the line that can be drawn between the points $(\pi, 0)$ and $(0, \pi)$ crosses the Fermi surface. When there is no such crossing, one recovers a Fermi surface (not shown here). The $(\pi, 0)$ to $(0, \pi)$ line has a double role. It is the antiferromagnetic zone boundary, as well as the line that indicates where umklapp processes become possible, i.e., the line where we can scatter a pair of particles on one side of the Fermi surface to the other side with loss or

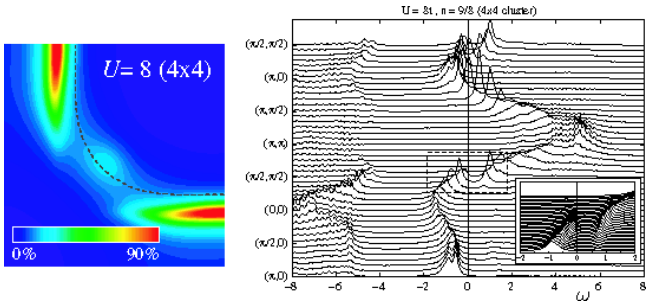

FIG. 22: Right: The corresponding EDC in the t-t'-t'' Hubbard model, calculated on a 16-site cluster in CPT, at $n = 9/8$. Inset: the pseudogap. Left: The corresponding Fermi energy momentum distribution curve.

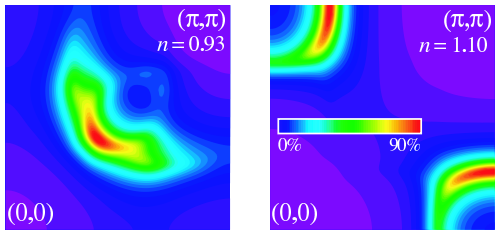

FIG. 23: EDC in the t-t'-t'' Hubbard model, with $t' = -0.3t$ and $t'' = 0.2t$, calculated on a 8-site cluster for $U = 8t$ in VCPT. d-wave superconductivity is present in the holde-doped case (left) and both antiferromagnetism and d-wave superconductivity in the electron-doped case. The resolution is not large enough in the latter case to see the superconducting gap. The Lorentzian broadening is $0.2t$. From Ref. 19.

gain of a reciprocal lattice vector. Large scattering rates explain the disappearance of the Fermi surface.[77] We also note that the size of the pseudogap in CPT, defined as the distance between the two peaks, does not scale like $J = 4t^2/U$ at large coupling. It seems to be very weakly U dependent, its size being related to t instead. This result is corroborated by CDMFT.[67]

The EDC for the electron-doped case is shown on the bottom panels of Fig. 19 near optimal doping again. This time, the Mott gap appears below the Fermi surface so that the lower Hubbard band becomes accessible to experiment. The EDC in Fig. 22 shows very well both the Mott gap and the pseudogap. Details of that pseudogap can be seen both in the inset of Fig. 22 or on the right-hand panel of Fig. 20. While in the hole-doped case the MDC appeared to evolve continuously as we increase U (top panel of Fig. 21), in the electron-doped case (bottom panel) the weight initially present near $(\pi/2, \pi/2)$ at $U = 4t$ disappears by the time we reach $U = 8t$.

In Fig. 23 we show, with the same resolution as CPT, the MDC for VCPT.[19] In this case the effect of long-range order is included and visible but, at this resolution, the results are not too different from those obtained from CPT in Fig. 21.

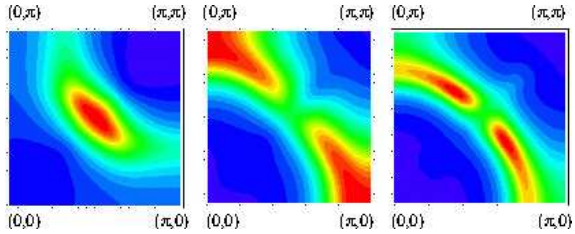

FIG. 24: MDC in the t-t', $U = 8t$ Hubbard model, calculated on a 4-site cluster in CDMFT. Energy resolution, $\eta = 0.1t$ (left and middle). Left: Hole-doped dSC ($t' = -0.3t, n = 0.96$), Middle: Electron-doped dSC ($t' = 0.3t, n = 0.93$), Right: Same as middle with $\eta = 0.02t$. Note the particle-hole transformation in the electron-doped case. From Ref. 17.

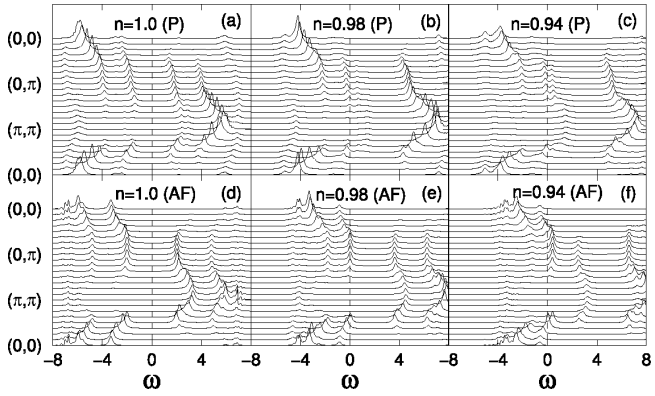

FIG. 25: EDC in the t-$t' = 0$, $U = 8t$ Hubbard model, calculated on a 4-site cluster in CDMFT. Top: normal (paramagnetic) state for various densities. Bottom: same for the antiferromagnetic state. From Ref. 67.

2. *CDMFT and DCA*

CDMFT[17] gives MDC that, at comparable resolution, $\eta = 0.1t$, are again compatible with CPT and with VCPT. The middle panel in Fig. 24 is for the electron-doped case but with a particle-hole transformation so that $t' = +0.3t$ and $\mathbf{k} \to \mathbf{k} + (\pi, \pi)$. Since there is a non-zero d-wave order parameter in this calculation, improving the resolution to $\eta = 0.02t$ reveals the d-wave gap, as seen in the right most figure.

It has been argued for a while in DCA that there is a mechanism whereby short-range correlations at strong coupling can be the source of the pseudogap phenomenon.[78] To illustrate this mechanism in CDMFT, we take the case $t' = t'' = 0$ and compare in Fig. 25 the EDC for $U = 8t$ without long-range order (top panels) and with long-range antiferromagnetic order (bottom panels).[67] The four bands appearing in Figs 25a and 25d are in agreement with what has been shown[73,79,80] with CPT, VCPT and QMC in Fig. 26. Evidently there are additional symmetries in the antiferromagnetic case. The bands that are most affected by the long-range order are those that are closest to the Fermi energy, hence they reflect spin correlations, while the bands far from

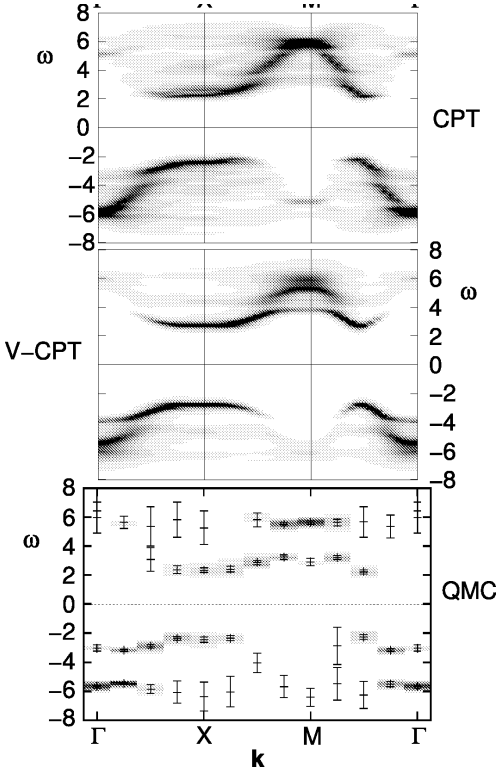

FIG. 26: EDC in the Hubbard model, $U = 8t$, $t' = 0$ calculated in CPT, VCPT and QMC. From Ref. 73.

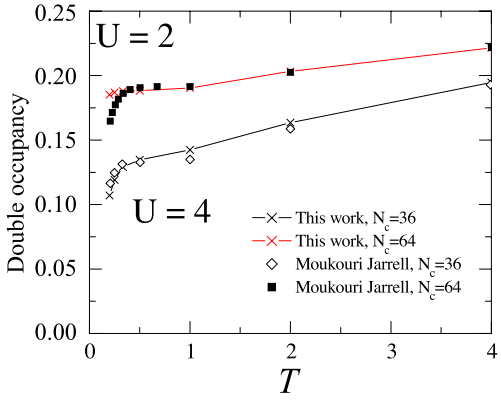

FIG. 27: Double occupancy $\langle n_\uparrow n_\downarrow \rangle$, in the two-dimensional Hubbard model for $n = 1$, as calculated from TPSC (lines with x) and from DCA (symbols) from Ref. 81. Taken from Ref. 84.

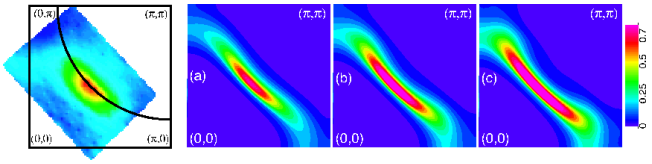

FIG. 28: MDC at the Fermi energy for the two-dimensional Hubbard model for $U = 6.25$, $t' = -0.175t, t'' = 0.05t$ at various hole dopings, as obtained from TPSC. The far left from Ref. 85 is the Fermi surface plot for 10% hole-doped $Ca_{2-x}Na_xCuO_2Cl_2$.

the Fermi energy seem less sensitive to the presence of long-range order. These far away bands are what is left from the atomic limit where we have two dispersionless bands. As we dope, the chemical potential moves into the lower band closest to the Fermi energy. When there is no long-range order (Figs 25b and 25c) the lower band closest to the Fermi energy moves very close to it, at the same time as the upper band closest to the Fermi energy looses weight, part of it moving closer to the Fermi energy. These two bands leave a pseudogap at the Fermi energy[82,83], although we cannot exclude that increasing the resolution would reveal a Fermi liquid at a very small energy scale. In the case when there is long-range antiferromagnetic order, (Figs 25e and 25f) the upper band is less affected while the chemical potential moves in the lower band closest to the Fermi energy but without creating a pseudogap, as if we were doping an itinerant antiferromagnet. It seems that forcing the spin correlations to remain short range leads to the pseudogap phenomenon in this case. When a pseudogap appears, it is created again by very large scattering rates.[67]

3. *TPSC (including analytical results)*

In Hartree-Fock theory, double occupancy is given by $n^2/4$ and is independent of temperature. The correct result does depend on temperature. One can observe in Fig. 27 the concordance between the results for the temperature-dependent double occupancy obtained with DCA and with TPSC[84] for the $t' = t'' = 0$ model. We have also done extensive comparisons between straight QMC calculations and TPSC.[38] The downturn at low temperature has been confirmed by the QMC calculations. It comes from the opening of the pseudogap due to antiferromagnetic fluctuations, as we will describe below. The concomitant increase in the local moment corresponds to the decrease in double-occupancy. There seems to be a disagreement at low temperature between TPSC and DCA at $U = 2t$. In fact TPSC is closer to the direct QMC calculation. Since we expect quantum cluster methods in general and DCA in particular to be less accurate at weak coupling, this is not too worrisome. At $U = 4t$ the density of states obtained with TPSC and with DCA at various temperatures are very close to each other.[84] We stress that as we go to temperatures well below the pseudogap, TPSC becomes less and less accurate, generally overemphasizing the downfall in double occupancy.

We will come back to more details on the predictions of TPSC for the pseudogap, but to illustrate the concordance with quantum cluster results shown in the previous subsection, we show in Fig. 28 MDC obtained at the Fermi energy in the hole doped case for $t' = -0.175t$, $t'' = 0.05t$. Again there is quasi-particle weight near

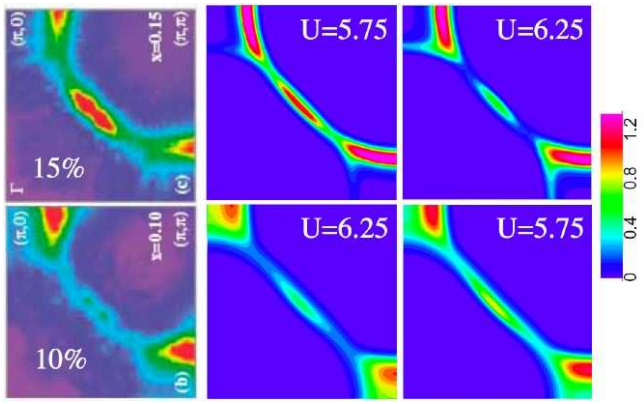

FIG. 29: MDC at the Fermi energy in the electron-doped case with $t' = -0.175t, t'' = 0.05t$ and two different U's, $U = 6.25t$ and $U = 5.75t$ obtained from TPSC. The first column is the corresponding experimental plots at 10% and 15% doping in Ref. 86. From Refs. 40,87.

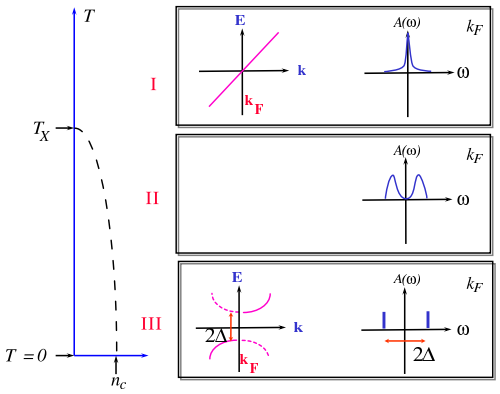

FIG. 30: Cartoon explanation of the pseudogap in the weak-coupling limit. Below the dashed crossover line to the renormalized classical regime, when the antiferromagnetic correlation length becomes larger than the thermal de Broglie wave length, there appears precursors of the zero-temperature Boboliubov quasiparticles for the long-range ordered antiferromagnet.

$(\pi/2, \pi/2)$ and a pseudogap near $(\pi, 0)$ and $(0, \pi)$. However, as we will discuss below, the antiferromagnetic correlation length necessary to obtain that pseudogap is too large compared with experiment. The electron-doped case is shown in Fig. 29 near optimal doping and for different values of U. As U increases, the weight near $(\pi/2, \pi/2)$ disappears. That is in concordance with the results of CPT shown in Fig. 21 where the weight at that location exists only for small U. That also agrees with slave-boson calculations[88] that found such weight for $U = 6t$ and it agrees also with one-loop calculations[89] starting from a Hartree-Fock antiferromagnetic state that did not find weight at that location for $U = 8t$. The simplest Hartree-Fock approach[21,90] yields weight near $(\pi/2, \pi/2)$ only for unreasonably small values of U.

A cartoon explanation of the pseudogap is given in Fig. 30. At high temperature we have a Fermi liquid, as illustrated in panel I. Now, suppose we start from a ground state with long-range order as in panel III, in other words at a filling between half-filling and n_c. In the Hartree-Fock approximation we have a gap and the fermion creation-annihilation operators now project on Bogoliubov-Valentin quasiparticles that have weight at both positive and negative energies. In two dimensions, the Mermin-Wagner theorem means that as soon as we raise the temperature above zero, long-range order disappears, but the antiferromagnetic correlation length ξ remains large so we obtain the situation illustrated in panel II, as long as ξ is much larger than the thermal de Broglie wave length $\xi_{th} \equiv v_F/(\pi T)$ in our usual units. At the crossover temperature T_X then the relative size of ξ and ξ_{th} changes and we recover the Fermi liquid. We now proceed to sketch analytically where these results come from starting from finite temperature. Details and more complete formulae may be found in Refs. 23,24,33,37. Note also that a study starting from zero temperature has also been performed in Ref. 91.

First we show how TPSC recovers the Mermin-Wagner theorem. Consider the self-consistency conditions given by the local moment sum rule Eq. (4) together with the expression for the spin-susceptibility Eq. (2) and U_{sp} in Eq. (5). First, it is clear that if the left-hand side of the local moment sum rule Eq. (4) wants to increase because of proximity to a phase transition, the right-hand side can do so only by decreasing $\langle n_\uparrow n_\downarrow \rangle$ which in turns decreases U_{sp} through Eq. (5) and moves the system away from the phase transition. This argument needs to be made more precise to include the effect of dimension. First, using the spectral representation one can show that every term of $\chi_{sp}(\mathbf{q}, iq_n)$ is positive. Near a phase transition, the zero Matsubara frequency component of the susceptibility begins to diverge. On can check from the real-time formalism that the zero-Matsubara frequency contribution dominates when the characteristic spin fluctuation frequency $\omega_{sp} \sim \xi^{-2}$ becomes less than temperature, the so-called renormalized-classical regime. We isolate this contribution on the left-hand side of the local moment sum rule and we move the contributions from the non-zero Matsubara frequencies, that are non-divergent, on the right-hand side. Then, converting the wave vector sum to an integral and expanding the denominator of the susceptibility around the wave vector where the instability would occurs to obtain an Ornstein-Zernicke form, the local moment sum rule Eq. (4) can be written in the form

$$T \int q^{d-1} dq \frac{1}{q^2 + \xi^{-2}} = C(T). \qquad (28)$$

The constant on the right-hand side contains only non-singular contributions and ξ^{-2} contains U_{sp} that we want to find. From the above equation, one finds immediately that in $d = 2$, $\xi \approx \exp(C(T)/T)$ so that the correlation length diverges only at $T = 0$. In three dimensions, isotropic or not, exponents correspond to those of the

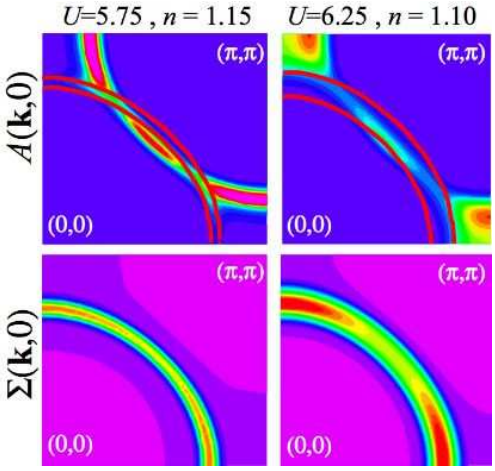

FIG. 31: MDC plots at the Fermi energy (upper) and corresponding scattering rates (lower) obtained from TPSC. The red lines on the upper panel indicate the region where the scattering rate in the corresponding lower panels is large.

$N = \infty$ universality class.[46]

To see how the pseudogap opens up in the single-particle spectral weight, consider the expression (6) for the self-energy. Normally one has to do the sum over bosonic Matsubara frequencies first, but the zero Matsubara frequency contribution has the correct asymptotic behavior in fermionic frequencies ik_n so that one can once more isolate on the right-hand side the zero Matsubara frequency contribution. This is confirmed by the real-time formalism[23] (See also Eq. (36) below). In the renormalized classical regime then, we have

$$\Sigma(\mathbf{k}_F, ik_n) \propto T \int q^{d-1} dq \frac{1}{q^2 + \xi^{-2}} \frac{1}{ik_n - \varepsilon_{\mathbf{k}_F+\mathbf{Q}+\mathbf{q}}} \quad (29)$$

where $\mathbf{Q}$ is the wave vector of the instability. Hence, when $\varepsilon_{\mathbf{k}_F+\mathbf{Q}} = 0$, in other words at hot spots, we find after analytical continuation and dimensional analysis that

$$\mathrm{Im}\,\Sigma^R(\mathbf{k}_F, 0) \propto -\pi T \int d^{d-1} q_\perp dq_{||} \frac{1}{q_\perp^2 + q_{||}^2 + \xi^{-2}} \delta(v'_F q_{||}) \quad (30)$$

$$\propto \frac{\pi T}{v'_F} \xi^{3-d}. \quad (31)$$

Clearly, in $d = 4$, $\mathrm{Im}\,\Sigma^R(\mathbf{k}_F, 0)$ vanishes as temperature decreases, $d = 3$ is the marginal dimension and in $d = 2$ we have that $\mathrm{Im}\,\Sigma^R(\mathbf{k}_F, 0) \propto \xi/\xi_{th}$ that diverges at zero temperature. In a Fermi liquid that quantity vanishes at zero temperature. A diverging $\mathrm{Im}\,\Sigma^R(\mathbf{k}_F, 0)$ corresponds to a vanishingly small $A(\mathbf{k}_F, \omega = 0)$ as we can see from

$$A(\mathbf{k}, \omega) = \frac{-2\,\mathrm{Im}\,\Sigma^R(\mathbf{k}_F, \omega)}{(\omega - \varepsilon_{\mathbf{k}} - \mathrm{Re}\,\Sigma^R(\mathbf{k}_F, \omega))^2 + \mathrm{Im}\,\Sigma^R(\mathbf{k}_F, \omega)^2}. \quad (32)$$

To see graphically this relationship between the location of the pseudogap and large scattering rates at the Fermi surface, we draw in Fig. 31 both the Fermi surface MDC and, in the lower panels, the corresponding plots for $\mathrm{Im}\,\Sigma^R(\mathbf{k}, 0)$. Note that at stronger U the scattering rate is large over a broader region, leading to a depletion of $A(\mathbf{k},\omega)$ over a broader range of $\mathbf{k}$ values.

An argument for the splitting in two peaks seen in Figs. 6 and 30 is as follows. Consider the singular renormalized contribution coming from the spin fluctuations in Eq. (29) at frequencies $\omega \gg v_F \xi^{-1}$. Taking into account that contributions to the integral come mostly from a region $q \leq \xi^{-1}$, this expression leads to

$$\mathrm{Re}\,\Sigma^R(\mathbf{k}_F, \omega) = \left(T \int q^{d-1} dq \frac{1}{q^2 + \xi^{-2}} \right) \frac{1}{ik_n - \varepsilon_{\mathbf{k}_F+\mathbf{Q}}} \equiv \frac{\Delta^2}{\omega - \varepsilon_{\mathbf{k}_F+\mathbf{Q}}} \quad (33)$$

which, when substituted in the expression for the spectral weight (32) leads to large contributions when

$$\omega - \varepsilon_{\mathbf{k}} - \frac{\Delta^2}{\omega - \varepsilon_{\mathbf{k}_F+\mathbf{Q}}} = 0 \quad (34)$$

or, equivalently,

$$\omega = \frac{(\varepsilon_{\mathbf{k}} + \varepsilon_{\mathbf{k}_F+\mathbf{Q}}) \pm \sqrt{(\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}_F+\mathbf{Q}})^2 + 4\Delta^2}}{2}, \quad (35)$$

which corresponds to the position of the hot spots in Fig. 29 for example.

Note that analogous arguments hold for any fluctuation that becomes soft,[23] including superconducting ones.[41,43] The wave vector $\mathbf{Q}$ would be different in each case.

4. *Weak- and strong-coupling pseudogaps*

The CPT results of Figs. 19 and 22 clearly show that the pseudogap is different from the Mott gap. At finite doping, the Mott gap remains a local phenomenon, in the sense that there is a region in frequency space that is not tied to $\omega = 0$ where *for all wave vectors* there are no states. The peudogap by contrast is tied to $\omega = 0$ and occurs in regions nearly connected by (π, π), whether we are talking about hole- or about electron-doped cuprates. That the phenomenon is caused by short-range correlations can be seen in CPT from the fact that the pseudogap is independent of cluster shape and size (most of the results were presented for 3×4 clusters and we did not go below size 2×2). The antiferromagnetic correlations and any other two-particle correlations do not extend beyond the size of the lattice in CPT. Hence, the pseudogap phenomenon cannot be caused by antiferromagnetic long-range order since no such order exists in CPT. This is also vividly illustrated by the CDMFT results in Fig. 25 that contrast the case with and without antiferromagnetic long-range order. The CDMFT results also suggest that the pseudogap appears in the bands

that are most affected by antiferromagnetic correlations hence it seems natural to associate it with short-range spin correlations. The value of t' has an effect, but it mostly through the fact that it has a strong influence on the relative location of the antiferromagnetic zone boundary and the Fermi surface, a crucial factor determining where the pseudogap is. All of this as well as many results obtained earlier by DCA[78] suggest that there is a strong coupling mechanism that leads to a pseudogap in the presence of only short-range two-body correlations. However, the range cannot be zero. Only the Mott gap appears at zero range, thus the pseudogap is absent in single-site DMFT.

In the presence of a pseudogap at strong coupling ($U > 8t$), wave vector is not, so to speak, such a bad quantum number in certain directions. In other words the wave description is better in those directions. In other directions that are "pseudogapped", it is as if the localized, or particle description was better. This competition between wave and particle behavior is inherent to the Hubbard model. At the Fermi surface in low dimension, it seems that this competition is resolved by dividing (it is a crossover, not a real division) the Fermi surface in different sections.

There is also a weak-coupling mechanism for the pseudogap. This has been discussed at length just in the previous section on TPSC. Another way to rephrase the calculations of the previous section is in the real frequency formalism. There one finds[23] that

$$\Sigma''^{R}(\mathbf{k}_F,\omega) \propto \int \frac{d^{d-1}q_\perp}{(2\pi)^{d-1}} \int \frac{d\omega'}{\pi}[n(\omega')+f(\omega+\omega')]\chi''_{sp}(\mathbf{q};\omega') \tag{36}$$

so that if the characteristic spin fluctuation frequency in $\chi''_{sp}(\mathbf{q};\omega')$ is much larger than temperature, then $[n(\omega')+f(\omega+\omega')]$ can be considered to act like a window of size ω or T and $\chi''_{sp}(\mathbf{q};\omega')$ can be replaced by a function of $\mathbf{q}$ times ω' which immediately leads to the Fermi liquid result $[\omega^2+(\pi T)^2]$. In the opposite limit where the characteristic spin fluctuation frequency in $\chi''_{sp}(\mathbf{q};\omega')$ is much less than temperature, then it acts as a window narrower than temperature and $[n(\omega')+f(\omega+\omega')]$ can be approximated by the low frequency limit of the Bose factor, namely T/ω'. Using the thermodynamic sum rule, that immediately leads to the result discussed before in Eq.(31), $\mathrm{Im}\,\Sigma(\mathbf{k}_F,0) \propto (\pi T/v'_F)\xi^{3-d}$. This mechanism for the pseudogap needs long correlation lengths. In CPT, this manifests itself by the fact that the apparent pseudogap in Fig. 21 at $U = 4t$ is in fact mostly a depression in spectral weight that depends on system size and shape. In addition, in contrast to the short-range strong-coupling mechanism, at weak coupling the pseudogap is more closely associated with the intersection of the antiferromagnetic zone boundary with the Fermi surface.

Which mechanism is important for the cuprates will be discussed below in the section on comparisons with experiments.

B. d-wave superconductivity

The existence of d-wave superconductivity at weak coupling in the Hubbard model mediated by the exchange of antiferromagnetic fluctuations[92,93] had been proposed even before the discovery of high-temperature superconductivity.[6–8] At strong-coupling, early papers[13,14] also proposed that the superconductivity would be d-wave. The issue became extremely controversial, and even recently papers have been published[12] that suggest that there is no d-wave superconductivity in the Hubbard model. That problem could have been solved very long ago through QMC calculations if it had been possible to do them at low enough temperature. Unfortunately, the sign problem hindered these simulations, and the high temperature results[5,44,95,96] were not encouraging: the d-wave susceptibility was smaller than for the non-interacting case. Since that time, numerical results from variational QMC,[4,97] exact diagonalization[16] and other numerical approaches[98] for example, suggest that there is indeed d-wave superconductivity in the Hubbard model.

In the first subsection, we show that VCPT leads to a zero-temperature phase diagram for both hole and electron-doped systems that does show the basic features of the cuprate phase diagram, namely an antiferromagnetic phase and a d-wave superconducting phase in doping ranges that are quite close to experiment[19] (The following section will treat in more detail comparisons with experiment). The results are consistent with CDMFT.[17] The fall in the d-wave superconducting order parameter near half-filling is associated with the Mott phenomenon. The next subsection stresses the instability towards d-wave superconductivity as seen from the normal state and mostly at weak coupling. We show that TPSC can reproduce available QMC results and that its extrapolation to lower temperature shows d-wave superconductivity in the Hubbard model. The transition temperature found at optimal doping[40] for $U = 4t$ agrees with that found by DCA,[18] a result that could be fortuitous. But again the concordance between theoretical results obtained at intermediate coupling with methods that are best at opposite ends of the range of coupling strengths is encouraging.

1. *Zero-temperature phase diagram*

In VCPT, one adds to the cluster Hamiltonian the terms[19]

$$H'_M = M\sum_\mu e^{i\mathbf{Q}\cdot\mathbf{R}_\mu}(n_{\mu\uparrow}-n_{\mu\downarrow}) \tag{37}$$

$$H'_D = D\sum_{\mu\nu} g_{\mu\nu}(c_{\mu\uparrow}c_{\nu\downarrow}+\mathrm{H.c.}) \tag{38}$$

with M and D are respectively antiferromagnetic and d-wave superconducting Weiss fields that are determined

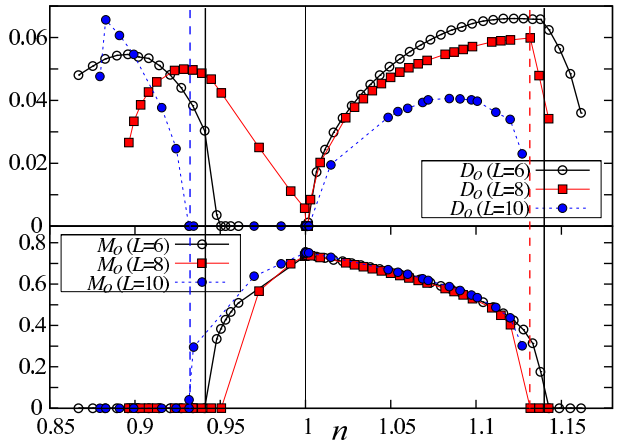

FIG. 32: Antiferromagnetic (bottom) and d-wave (top) order parameters for $U = 8t$, $t' = -0.3t$ $t'' = 0.2t$ as a function of the electron density (n) for 2×3, 2×4 and 10-site clusters, calculated in VCPT. Vertical lines indicate the first doping where only d-wave order is non-vanishing. From Ref. 19.

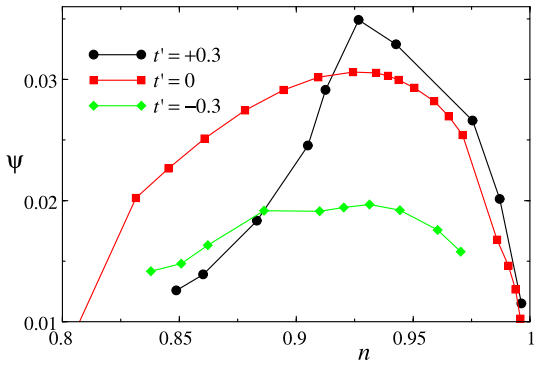

FIG. 33: d-wave order parameter as a function of n for various values of t', calculated in CDMFT on a 2×2 cluster for $U = 8t$. The positive t' case corresponds to the electron-doped case when a particle-hole transformation is performed. From Ref. 17.

self-consistently and $g_{\mu\nu}$ equal to ± 1 on near-neighbor sites following the d-wave pattern. We recall that the cluster Hamiltonian should be understood in a variational sense. Fig. 32 summarizes, for various cluster sizes, the results for the d-wave order parameter D_0 and for the antiferromagnetic order parameter M_0 for a fixed value of $U = 8t$ and the usual hopping parameters $t' = -0.3t$ and $t'' = 0.2t$. The results for antiferromagnetism are quite robust and extend over ranges of dopings that correspond quite closely to those observed experimentally. Despite the fact that the results for d-wave superconductivity still show some size dependence, it is clear that superconductivity alone without coexistence extends over a much broader range of dopings on the hole-doped than on the electron-doped side as observed experimentally. VCPT calculations on smaller system sizes[99] but that include, for thermodynamic consistency, the cluster chemical potential as a variational parameter show superconductivity

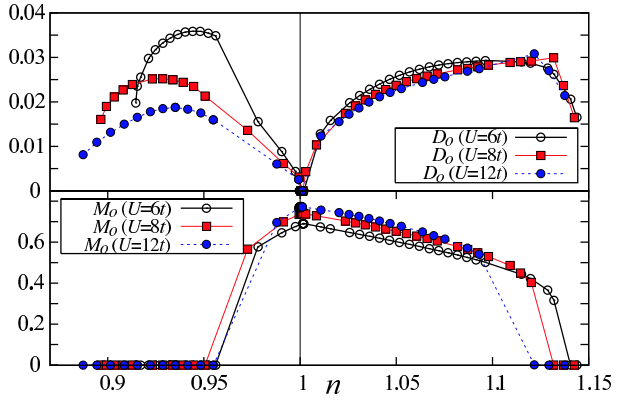

FIG. 34: Antiferromagnetic (bottom) and d-wave (top) order parameters as a function of the electron density (n) for $t' = -0.3t$ $t'' = 0.2t$ and various values of U on a 8-site cluster, calculated in VCPT. From Ref. 19.

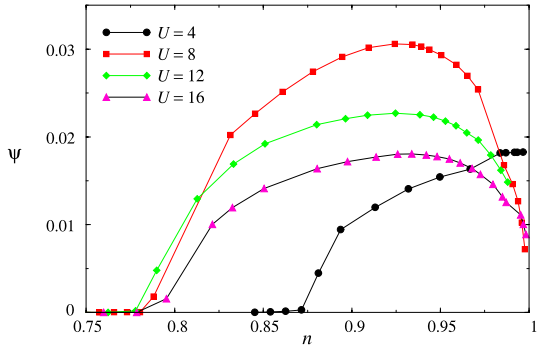

FIG. 35: d-wave order parameter as a function of n for various values of U, and $t' = t'' = 0$ calculated in CDMFT on a 4-site cluster. From Ref. 17.

that extends over a much broader range of dopings. Also, for small 2×2 clusters, VCPT has stronger order parameter on the electron than on the hole-doped side, contrary to the results for the largest system sizes in Fig. 32. This is also what is found in CDMFT as can be seen in Fig. 33. It is quite likely that the zero-temperature Cooper pair size is larger than two sites, so we consider the results for 2×2 systems only for their qualitative value.

Concerning the question of coexistence with antiferromagnetism, one can see that it is quite robust on the electron-doped side whereas on the hole-doped side, it is size dependent. That suggests that one should also look at inhomogeneous solutions on the hole-doped side since stripes are generally found experimentally near the regions where antiferromagnetism and superconductivity meet.

Fig. 34 shows clearly that at strong coupling the size of the order parameter seems to scale with J, in other words it decreases with increasing U. This is also found in CDMFT,[17] as shown in Fig. 35 for $t' = t'' = 0$.

If we keep the antiferromagnetic order parameter to zero, one can check with both VCPT and CDMFT (Fig. 33) that the d-wave superconducting order parame-

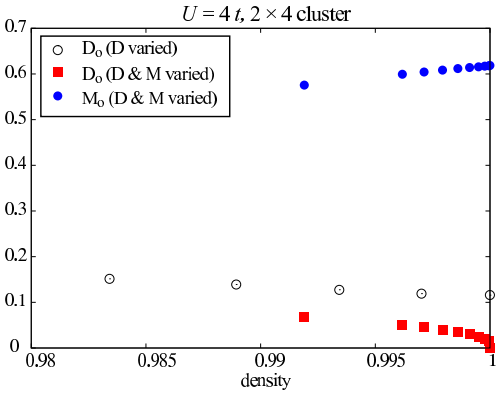

FIG. 36: VCPT calculations for $U = 4t, t' = t'' = 0$ near half-filling on 2×4 lattice. Contrary to the strong coupling case, the d-wave order parameter D_0 survives all the way to half-filling at weak coupling, unless we also allow for antiferromagnetism.

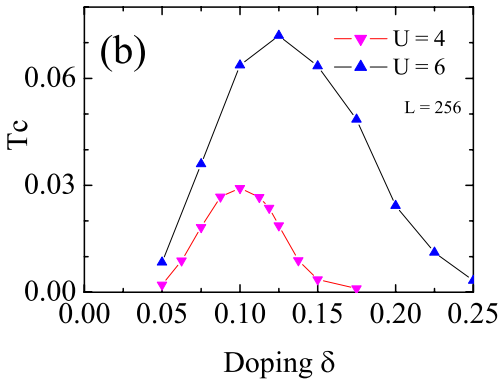

FIG. 37: T_c as a function of doping, $\delta = 1 - n$, for $t' = t'' = 0$ calculated in TPSC using the Thouless criterion. From Ref. 40.

ter goes to zero at half-filling. This is clearly due to Mott localization. Indeed, at smaller $U = 4t$ for example, the order parameter does not vanish at half-filling if we do not allow for long-range antiferromagnetic order, as illustrated in Fig. 35 for CDMFT.[17] The same result was found in VCPT, as shown in Fig.36.[19] Restoring long-range antiferromagnetic order does however make the d-wave order parameter vanish at half-filling.

There are thus two ways to make d-wave superconductivity disappear at half-filling, either through long antiferromagnetic correlation lengths[94] or through the Mott phenomenon. In the real systems, that are Mott insulators and also antiferromagnets at half-filling, both effects can contribute.

2. *Instability of the normal phase*

In the introduction to this section, we alluded to QMC calculations for the d-wave susceptibility.[5,44,95,96] Recent results[40] for that quantity as a function of doping for various temperatures and for $U = 4t$, $t' = t'' = 0$ are shown by symbols in Fig. 8. For lower temperatures, the sign problem prevents the calculation near half-filling. Yet, the lowest temperature is low enough that a dome shape begins to appear. Nevertheless, comparison with the non-interacting case, shown by the top continuous line, leads one to believe that interactions only suppress d-wave superconductivity. We can easily understand why this is so. As we already know, the TPSC results obtained from Eq. (7) are very close to the QMC calculations, as shown by the solid lines in Fig. 8. In the temperature range of interest, the main contribution comes from the first term in Eq. (7). That term represents the contribution to the susceptibility that comes from dressed quasiparticles that do not interact with each other. Since dressed quasiparticles have a lifetime, a pair breaking effect, it is normal that this contribution to the interacting susceptibility leads to a smaller contribution than in the non-interacting case. At the lowest temperature, $\beta = 4/t$, the vertex contribution represented by the second term in Eq. (7) accounts for about 20% of the total. It goes in the direction of increasing the susceptibility. As we decrease the temperature further in TPSC, that term becomes comparable with the first one. Since the vertex in Eq. (7) accounts for the exchange of a single spin wave, equality with the first term signals the divergence of the series, as in $1/(1-x) \sim 1 + x$. The divergence of that series represents physically the instability of the normal phase to a d-wave superconducting phase. This is analogous to the Thouless criterion and hence it gives an upper bound to T_c. In other words, Berezinskii-Kosterlitz-Thouless physics is not included.

Fig. 37 shows the TPSC transition temperature for $U = 4t$ and for $U = 6t$. As we move towards half-filling, located to the left of the diagram, starting from large dopings, T_c goes up because of the increase in antiferromagnetic fluctuations. Eventually, T_c turns around and decreases because of the opening of the pseudogap. Physically, when the pseudogap opens up, weight is removed from the Fermi level, the density of states becomes very small, and pairing cannot occur any more. In the FLEX approximation[22,100] that does not exhibit a pseudogap,[101] that downturn, observed already in QMC at high temperature, does not occur. We have observed that in cases where $t' \neq 0$ so that the pseudogap opens only in a limited region around hot spots, the downturn can become less pronounced.

The case $n = 0.9 = 1 - \delta$ that corresponds to optimal doping for $U = 4t$ in Fig. 37 has been studied by DCA. In an extensive and systematic study of the size dependence, Maier *et al.*[18] established the existence of a d-wave instability at a temperature that coincides to within a few percent with the result in Fig. 37. Since very few vortices can fit within even the largest cluster sizes studied in Ref. 18, it is quite likely that the T_c that they find does not include Berezinskii-Kosterlitz-Thouless effects, just like ours. Despite the fact that, again, the

concordance between weak and strong coupling methods at intermediate coupling comforts us, the uncertainties in the results found with TPSC and DCA force us to also allow for a fortuitous coincidence.

IV. COMPARISONS WITH EXPERIMENT

The reduction of the real problem of high-temperature superconducting materials to a one-band Hubbard is a non-trivial one. It has been discussed already in the early days of high T_c superconductivity. The notion of a Zhang-Rice singlet[102] emerged for hole doped systems. The mapping to a one-band model has been discussed in many references,[103,104] and we do not wish to discuss this point further here. In fact it is far from obvious that this mapping is possible. It is known that about 0.5 eV below the Fermi surface, that mapping fails in hole-doped systems.[104] Nevertheless, the one-band Hubbard model is in itself a hard enough problem for us. So it is satisfying to see that, in the end, it gives a picture that agrees with experiment in a quite detailed manner for the ARPES spectrum near the Fermi surface, for the phase diagram as well as for neutron scattering in cases where it can be calculated.

Although we will not come back on this point at all, we briefly mention that fitting the spin wave spectrum[105] for all energies and wave vectors at half filling in La_2CuO_4 gives values of U, t, t', t'' that are close to those used in the rest of this paper.[106–109] It is in this context that ring exchange terms are usually discussed.

A. ARPES spectrum, an overview

ARPES experiments have played a central role in the field of high-temperature superconductivity. We cannot expect to be able to present the vast experimental literature on the subject. We refer the reader to a very exhaustive review[110] and to some less complete but recent ones.[111,112] The main facts about ARPES have been summarized in Ref. 110. We comment on their main points one by one, using italics for our paraphrase of the reported experimental observations.

(i) The importance of Mott Physics and the renormalization of the bandwidth from t to J for the undoped parent compounds. This renormalization was clear already in early QMC calculations.[80,113] We already discussed the presence of four peaks. The one nearest to the Fermi surface at negative energies is the one referred to by experimentalists when they refer to this renormalization. This band has a dispersion of order J (not shown on Fig. 26, but see Ref. 113). This result also agrees with VCPT as shown in Fig. 26 and CDMFT (Fig. 25a). As shown in Figs. 25a and 25d, whether the state is ordered or not the band width is similar. Analytical strong-coupling expansions[114,115] and exact diagonalizations also find the same result. To find detailed agreement with experiment, one needs to include t' and t''.[116] The evolution of the position of chemical potential for extremely small dopings as discussed in Ref. 117 is not reproduced by the strong-coupling calculations, although the result on chemical potential is somewhat material dependent.[118]

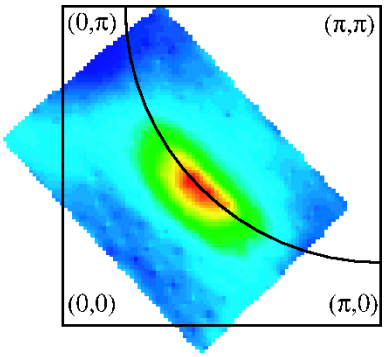

FIG. 38: MDC at the Fermi energy for 10% hole-doped $Ca_{2-x}Na_xCuO_2Cl_2$ from Ref. 85.

(ii) In the overdoped case, the Fermi surface is well defined. Although we have not shown any figures concerning this point, VCPT and CDMFT show the same result.

(iii) The evolution with doping of the electronic structure has been mapped. It has shown the importance of antiferromagnetic correlations in the p-type underdoped cuprates and especially in the n-type ones in which the hot-spot physics is still observed at optimal doping. We will come back on the latter point for electron-doped cuprates in the following subsection. The strong-coupling results obtained with VCPT and CDMFT have a resolution of order $0.1t$, which translates into about 30 meV. This is not enough to accurately measure the Fermi velocity, which was found to be doping independent in LSCO.[119] However, this suffices to compare with MDC curves obtained experimentally by integrating over an energy range of about 60 meV, as shown in Fig. 38 obtained in Ref. 85 on Calcium oxyclorate $Ca_{2-x}Na_xCuO_2Cl_2$, a 10% hole-doped high temperature superconductor. The similarities between that figure and the CPT (Fig. 21), VCPT (Fig. 23) and CDMFT (Fig. 24) results is striking. The agreement is better when no antiferromagnetic long-range order is assumed, as in the CPT case. The flattening of the band structure near $(\pi, 0)$ observed experimentally, can also be seen in CPT by comparing the top and middle EDC's taken at small and large U respectively on the left panel of Fig. 20. This flattening is associated with the pseudogap phenomenon. Recall that the theoretical results were obtained with $t' = -0.3t$ and $t'' = 0.2t$. This in turn implies an electron-hole asymmetry that is observed experimentally. We come back to this in the following subsection.

(iv) The overall d-wave symmetry of the superconducting gap has been observed for both hole and electron doping, supporting the universality of the pairing nature in

the cuprates. In the next to next subsection, we discuss the phase diagram for competing antiferromagnetism and d-wave superconductivity and show striking similarities with the observations.

(v) A normal-state pseudogap has been observed to open up at a temperature $T^* > T_c$ in the underdoped regime with a d-wave form similar to the one of the superconducting gap. That statement is correct only in the hole-doped compounds. In electron-doped systems the pseudogap has a form that is not of d-wave shape. If T_c comes from a universal pairing mechanism, a universal mechanism may also be behind the pseudogap. As we have already discussed however, there are quantitative differences between strong and weak coupling mechanisms for both T_c and the pseudogap. For electron-doped systems, we made quantitative predictions for the value of T^* that have later been confirmed experimentally. All this is discussed further below. To date, in cluster methods the pseudogap temperature has been studied only with DCA.[78]

(vi) A coherent quasiparticle peak below T_c has been observed near $(\pi, 0)$ whose spectral weight scales with the doping level x in the underdoped regime. We expect that it is a general result that long-range order will restore quasiparticle like excitations in strongly correlated systems because gaps remove scattering channels near the Fermi level. This is clearly illustrated by comparing the upper and lower panels in Fig. 25 that contrast the same spectra with and without antiferromagnetic long-range order. We have not performed the analysis of our results yet that could tell us whether the spectral weight of the quasiparticle scales with x in the hole-underdoped regime. Our resolution may not be good enough to see the quasiparticle peak. Sharpening of the quasiparticle excitations in the superconducting state has however been observed in DCA.[120]

(vii) The presence of an energy scale of about $40 - 80$ meV *in the quasiparticle dynamics manifests itself through a sharp dispersion renormalization and drop in the scattering rate observed at those energies at different momenta.* In hole-doped systems there is a kink in the nodal direction that is already seen above T_c while in the antinodal direction it appears only below T_c. The energy scales and doping dependences of these two kinks are also different.[121] The energy resolution in VCPT and CDMFT is not sufficient to distinguish these subtleties. In electron-doped cuprates experiments[122] suggest that there is no observable kink feature, in agreement with the results presented in the following subsection.

B. The pseudogap in electron-doped cuprates

The ARPES spectrum of electron-doped cuprates is strikingly different from that of their hole-doped counterpart. The Fermi energy MDC's for the first quadrant of the Brillouin zone[86] are shown at the top of Fig. 39 for three different dopings. There is a very clear evolution

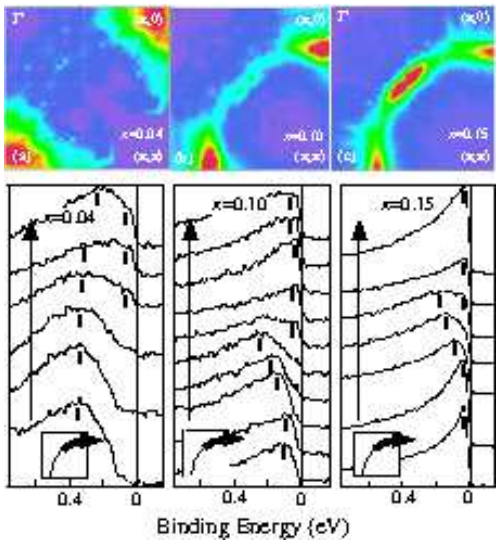

FIG. 39: Doping dependence of the MDC from experiments on NCCO with the corresponding EDC. From Ref. 86

with doping. At the lowest dopings, there is no weight near $(\pi/2, \pi/2)$, contrary to the hole-doped case shown in Fig. 38. For all dopings there is weight near $(\pi, 0)$ instead of the pseudogap that appeared there in the hole-doped case. The EDC's, also shown on the bottom of Fig. 39, are drawn for a trajectory in the Brillouin zone that follows what would be the Fermi surface in the non-interacting case. Regions that are more green than red on the corresponding MDC's along that trajectory are referred to as hot spots. On the EDC's we clearly see that hot spots do not correspond to simply a decrease in the quasiparticle weight Z. They truly originate from a pseudogap, in other words from the fact that the maximum is pushed away from zero energy. Even though the measurements are done at low temperature $(T = 10 - 20K)$ the energy resolution of about 60 meV makes the superconducting gap invisible. What is observed at this resolution is the pseudogap.

The contrast between the location of the pseudogap in the hole and electron-doped compounds is clearly seen in Fig. 40 obtained from VCPT.[19] In that figure, the magnitude of the spectral weight is represented by the different colors as a function of frequency (in units of t) along different cuts of the Brillouin zone. In the bottom panel, for the hole-doped case, one observes the pseudogap near $(\pi, 0)$. In the top panel, for the electron-doped case, it is only by zooming (inset) on the region for the Fermi energy crossing near $(\pi, 0)$ that one sees the d-wave superconducting gap. At 10% electron-doping, the pseudogap near $(\pi/2, \pi/2)$ is apparent. In this case there is antiferromagnetic long-range order, but even if we use CPT that does not exhibit long-range order, there appears a pseudogap in that region.[160] The main difference between CPT and VCPT results is the bending back of the bands (for example around the symmetry axis at $(\pi/2, \pi/2)$) caused by halving of the size of the Brillouin

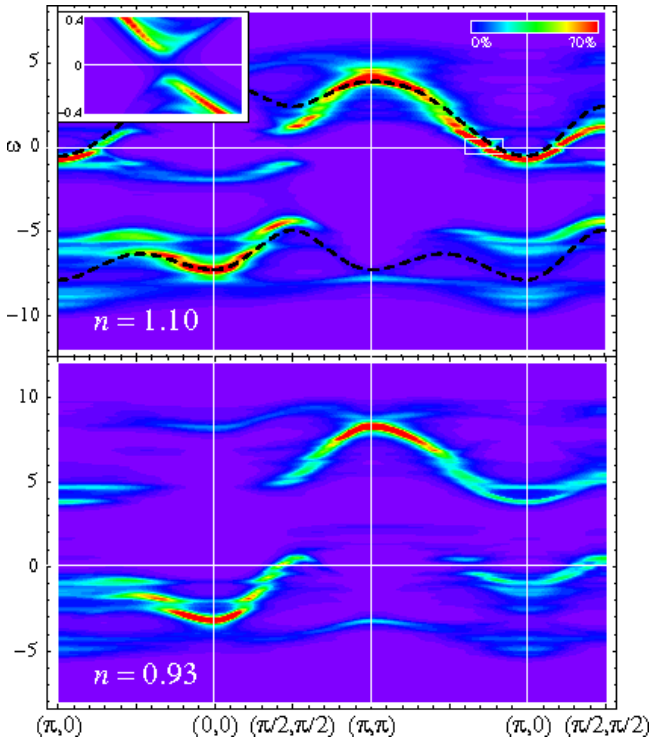

FIG. 40: Intensity plot of the spectral function as a function of ω in units of t and wave vector from VCPT for $U = 8t$ $t' = -0.3t, t'' = 0.2t$ and $n = 0.93$ at the bottom and $n = 1.10$ (electron-doped) at the top. The Lorentzian broadening is $0.12t$ in the main figure and $0.04t$ in the inset that displays the d-wave gap. Top panel is for the electron-doped case in the right-hand panel of Fig. 23, while bottom panel is for the hole-doped case on the left of Fig. 23. From Ref. 19.

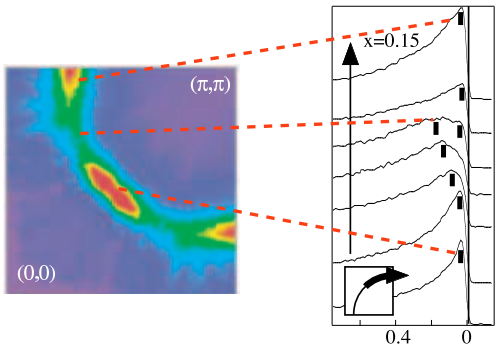

FIG. 41: Experimental Fermi surface plot (MDC at the Fermi energy) for NCCO (left) and corresponding energy distribution curves (right) for 15% electron-doping. From Ref. 86

zone in the antiferromagnetic case. Form factors[90] are such that the intensity is not symmetric even if the dispersion is. The faint band located at an energy about t below the Fermi energy near $(\pi/2, \pi/2)$ was also found in Ref. 89 by a one-loop spin-wave calculation around the Hartree-Fock antiferromagnetic ground state at $U = 8t$. Experimentalists[86] have suggested the existence of these states. The VCPT results go well beyond the spin-wave

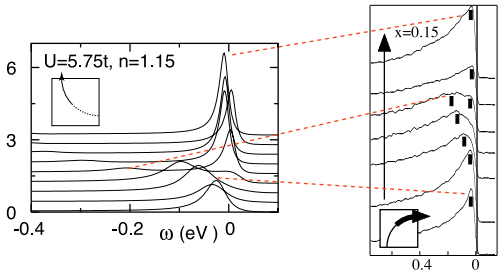

FIG. 42: EDC $A_<(\mathbf{k},\omega) \equiv A(\mathbf{k},\omega)f(\omega)$ along the Fermi surface calculated in TPSC (left) at optimal doping for $t' = -0.175t$, $t'' = 0.05t$, $t = 350$ meV and corresponding ARPES data on NCCO (right). From Ref. 87.

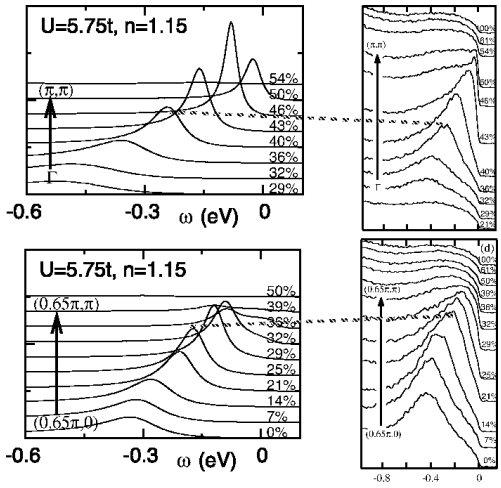

FIG. 43: EDC $A_<(\mathbf{k},\omega) \equiv A(\mathbf{k},\omega)f(\omega)$ along two other directions calculated for $t' = -0.175t$, $t'' = 0.05t$, $t = 350$ meV in TPSC (left column) and corresponding ARPES data on NCCO (right column). From Ref. 87.

calculation (dashed lines in Fig. 40) since one can also see numerous features in addition to remnants of the localized atomic levels around $+5t$ and $-10t$.

The optimally doped case is the real challenge for strong-coupling calculations. The spin-wave approach in Ref. 89 never shows the weight near $(\pi/2, \pi/2)$ that is seen in experiment (Fig. 41). Early mean-field calculations by Kusko et al.[90] suggest that this $(\pi/2, \pi/2)$ feature appears for $U = 3t$. This is very small compared with U of the order of the bandwidth $8t$ necessary to have a Mott insulator at half-filling. We already discussed in Sec. III A that both CPT and TPSC show that weight near $(\pi/2, \pi/2)$ appears for U not too large, say of order $6t$. This same result is also obtained in the Kotliar-Ruckenstein slave boson approach.[88]

Since TPSC is valid for a system of infinite size, we present detailed comparisons[87] with experiment[86] on $Nd_{1.85}Ce_{0.15}CuO_4$, an electron-doped cuprate. We take $t' = -0.175t$, $t'' = 0.05t$. Results obtained with $t' = -0.275$ are very close to those we present. With the values used in CPT, $t' = -0.3t$, $t' = 0.2t$, $U = 6t$, TPSC

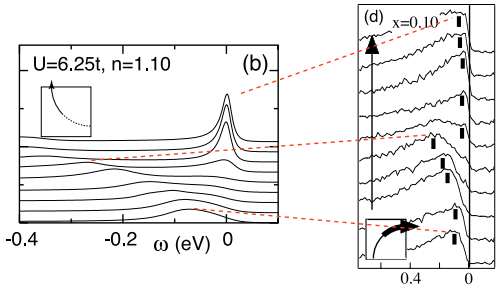

FIG. 44: EDC $A_<(\mathbf{k},\omega) \equiv A(\mathbf{k},\omega)f(\omega)$ along the Fermi surface shown in the insets for (b) $n = 1.10$, $U = 6.25t$. Lines are shifted by a constant for clarity. From Ref. 87.

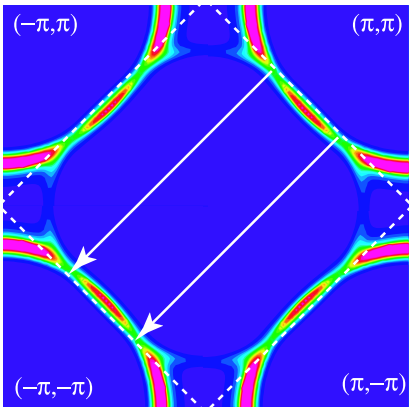

FIG. 45: Hot spots from quasi-static scatterings off antiferromagnetic fluctuations (renormalized classical regime).

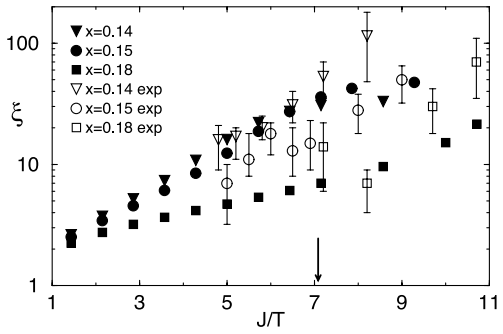

FIG. 46: Semi-log plot of the AFM correlation length (in units of the lattice constant) against inverse temperature (in units of $J = 125$ meV). Filled symbols denote calculated results and empty ones experimental data of Ref. 124 and Ref. 125 ($x = 0.15$). From Ref. 87.

does not lead to strong enough antiferromagnetic fluctuations to obtain non-trivial effects in the temperature range studied, $\beta = 20t$. We take $t = 350$ meV. Fig. 39 shows the correspondence between EDC and MDC. Comparisons with experimental EDC at wave vectors along the non-interacting Fermi surface appear in Fig. 42 for $U = 5.75t$ and 15% doping ($n = 1.15$). The dashed lines indicate the quite detailed agreement between theory and experiment. At the hot spot, (middle dashed line), the weight is pushed back about 0.2 eV and there is a very small peak left at the Fermi surface, as in the experiment. If U is not large enough the antiferromagnetic fluctuations are not strong enough to lead to a pseudogap. As in CPT (Fig. 21), if U is too large the $(\pi/2, \pi/2)$ weight disappears, as illustrated earlier in Fig. 29. In Fig. 43, cuts along the $(0, 0)$ to (π, π) and $(0.65\pi, 0)$ to $(0.65\pi, \pi)$ directions are compared with experiment. Again the peak positions and widths are very close, except for some experimental tails extending in the large binding energy direction. The theoretical results have similar asymmetry, but not as pronounced. Experimentally, the large binding energy tails ("the background") are the least reproducible features from sample to sample, especially for wave vectors near the Fermi surface".[161] The experimental renormalized Fermi velocities are 3.31×10^5 m/s and 3.09×10^5 m/s along the zone diagonal and along the $(\pi, 0)$-(π, π) direction, respectively. The corresponding renormalized Fermi velocities obtained by TPSC are 3.27×10^5 m/s and 2.49×10^5 m/s, respectively. The agreement is very good, particularly along the diagonal direction. The bare Fermi velocities are renormalized in TPSC by roughly a factor of two.[123]

As we move towards half-filling, we have to increase U slightly to find agreement with experiment, as discussed earlier in Fig. 29. Fig. 44 shows how well the EDC's agree for a Fermi surface cut at 10% doping ($n = 1.10$). The increase is expected physically from the fact that with fewer electrons the contribution to screening that comes from Thomas Fermi physics should not be as good. This is also consistent with the fact that a larger value of U is necessary to explain the Mott insulator at half-filling. It would also be possible to mimic the ARPES spectrum by keeping U fixed and changing the hopping parameters, but the changes would be of order 20%, which does not appear realistic.[87]

We have already explained that the physics behind the pseudogap in TPSC is scattering by nearly critical antiferromagnetic fluctuations. This is illustrated in Fig. 45. If this explanation is correct, the antiferromagnetic correlation length measured by neutron scattering should be quite large. The results of the measurement[124,125] and of the TPSC calculations are shown in Fig. 46. The agreement is again surprisingly good. As we move to smaller dopings $n = 1.1$ (not shown) the agreement becomes less good, but we do expect TPSC to deteriorate as U increases with underdoping. The arrow points to the temperature where EDC's shown earlier were calculated. Note however that the neutron measurements were done on samples that were not reduced, by contrast with the ARPES measurements mentioned earlier. We are expecting experiments on this subject.[162]. We should point out that the EDC's depend strongly on temperature and on the actual value of the antiferromagnetic correlation length only in the vicinity of the temperature where there is a crossover to the pseudogap regime. Decreasing the

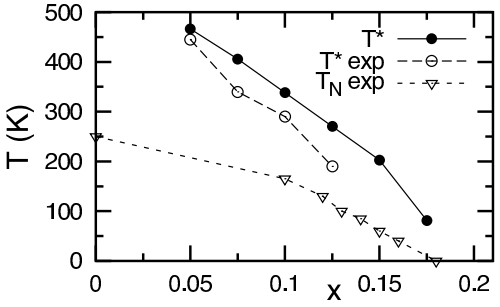

FIG. 47: Pseudogap temperature T^* (filled circles denote T^* calculated from TPSC, empty ones experimental data extracted from optical conductivity.[127]) Empty triangles are experimental Néel temperatures T_N. The samples are reduced.[124] From Ref. 87.

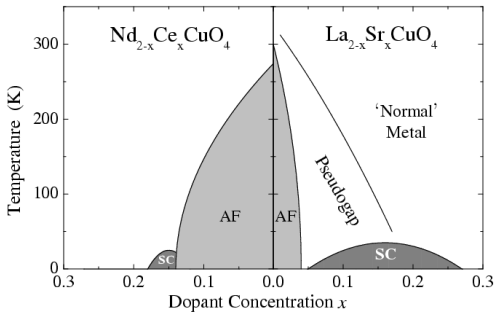

FIG. 48: The generic phase diagram of high-T_c superconductors, from Ref. 110. There should also be a pseudogap line on the electron-doped side. It was not well studied at the time of publication of that paper.

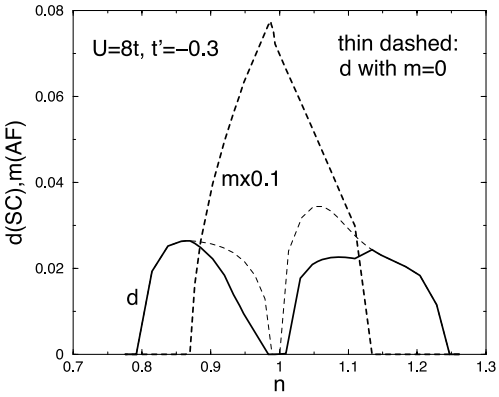

FIG. 49: Antiferromagnetic order parameter m (dashed) and d-wave (solid) order parameter obtained from CDMFT on a 2×2 cluster. The result obtained by forcing $m = 0$ is also shown as a thin dashed line.

temperature makes the $\omega = 0$ peaks near $(\pi, 0)$ sharper[87] as observed experimentally.[126]

The ARPES pseudogap temperature T^* has been predicted with TPSC.[87] The predictions are shown by the solid line in Fig. 47. The pseudogap temperature observed in optical experiments[127] is shown by the open circles. It differs from the ARPES result, especially as we move towards optimal doping. The size of the pseudogap observed in the optical experiments[127] ($10T^*$) is comparable to the ARPES pseudogap. The solid line in Fig. 47 contains several predictions. If we look at 13% doping ($n = 1.13$), the line predicts $T^* \sim 250K$. Experiments that were done without being aware of this prediction[128] have verified it. It would be most interesting to do neutron scattering experiments on the same samples to check whether the antiferromagnetic correlation length ξ and the thermal de Broglie wave length ξ_{th} are comparable at that temperature, as predicted by TPSC. Fig. 47 also predicts that the pseudogap induced by antiferromagnetic fluctuations will disappear at the quantum critical point where long-range antiferromagnetic order disappears, in other words it will coincide with the crossing of the experimentally observed Néel temperature (dashed line with triangles in Fig. 47) with the zero temperature axis (if that crossing is not masked by the superconducting transition). Recent optical conductivity experiments[129,130] confirm this prediction as well.

In TPSC, superconducting fluctuations can also lead to a pseudogap by an analogous mechanism.[23]

C. The phase diagram for high-temperature superconductors

The main features appearing in the phase diagram of high-temperature superconductors are the pseudogap phase, the antiferromagnetic phase and the d-wave superconducting phase. Fig. 48[110] shows the typical diagram with hole doping to the right and electron doping to the left. Zero on the horizontal axis corresponds to half-filling. There are other features on the phase diagram, in particular checkerboard patterns[131] or stripe phases[132] that appear in general close to the region where antiferromagnetism and superconductivity come close to each other. Before we try to understand these more detailed features, one should understand the most important phases. In the previous subsection we have discussed the pseudogap phase, in particular on the electron-doped side (not indicated on Fig. 48). A recent review of the pseudogap appears in Ref. 133. In the following, we discuss in turn the phase diagram and then the nature of the superconducting phase itself and its relation to the Mott phenomenon.

1. *Competition between antiferromagnetism and superconductivity*

We have already shown in Fig. 32 the prediction of VCPT for the zero-temperature phase diagram.[19] Here, we just point out how closely the position of the antiferromagnetic phase boundary, appearing in the lower panel, coincides with the experimental phase diagram in

Fig. 48. (Note that electron concentration increases from right to left on this experimental phase diagram). In particular, there is little size dependence to the position of this boundary, (6 to 10 sites) and in addition the dependence on the value of U is also weak, as can be seen from Fig. 34. Hence, the positions of the antiferromagnetic phase boundaries is a robust prediction of VCPT. The CDMFT result for a four site cluster in a bath is shown in Fig. 49 for $t' = -0.3t, t'' = 0$ and $U = 8t$. The agreement with experiment is not as good. Despite the useful presence of a bath in CDMFT, the cluster itself is of size 2×2, which is probably smaller than the Cooper pair size. We can obtain results closer to those of VCPT by increasing the variational space.

The d-wave superconducting order parameter on the top panel of Fig. 32 shows more size dependence than the antiferromagnetic order parameter. Nevertheless, there are some clear tendencies: (a) d-wave superconductivity can exist by itself, without antiferromagnetism. The vertical lines indicate the location of the end of the antiferromagnetic phase for the various system sizes to help this observation. (b) The range where d-wave-superconductivity exists without antiferromagnetism, is about three times larger on the hole than on the electron-doped side, as observed experimentally. (c) As system size increases, the maximum d-wave order parameter is larger on the hole than on the electron-doped side. (d) The tendency to have coexisting antiferromagnetism and d-wave superconductivity is rather strong on the electron-doped side of the phase diagram. This is observed experimentally[134] but only over a rather narrow region near optimal doping. Recent experiments[135] challenge this result, others[136,137] indicate that antiferromagnetism can be induced from the d-wave superconducting phases with very small fields. (e) On the hole-doped side, d-wave superconductivity and antiferromagnetism coexist for a very narrow range of dopings for system size $N_c = 6$, for a broad range extending to half-filling for $N_c = 8$ and not at all for $N_c = 10$. In other words, the tendency to coexistence is not even monotonic. We interpret this result as a reflection of the tendency to form stripes observed experimentally on the hole-doped side.[132,138,139] We cannot study systems large enough to allow for striped inhomogeneous states to check this statement.

The more realistic two-band model has also been studied using DCA.[104] The results are shown on Fig. 50. Electron concentration increases from right to left. This phase diagram is very close to that obtained with the same method from the one-band Hubbard model[104] with $t' = -0.3t$, $t'' = 0$, $U = 8t$. The qualitative results agree with the other calculations and with experiment: antiferromagnetism extends over a narrower doping range for hole than for electron doping and d-wave superconductivity by itself exists over a broader range for the hole-doped case than for the electron-doped case. The actual ranges where antiferromagnetism and d-wave superconductivity exist are not in as good an agreement with experiment

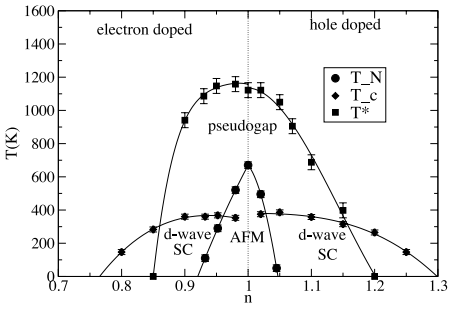

FIG. 50: Phase diagram obtained from DCA for $U = 8t$ for the two-band model. From Ref. 104

as in the VCPT case. However, as in CDMFT, the system sizes, 2×2, are very small. Overall then, quantum cluster methods, VCPT in particular, allow us to obtain from the Hubbard model the two main phases, antiferromagnetic and d-wave superconducting, essentially in the observed doping range of the zero-temperature phase diagram. At finite temperature, DCA and TPSC agree on the value of T_c for the particle-hole symmetric model at 10% doping and $U = 4t$. Recent studies of the irreducible vertex using DCA[140] also show that in the weak-coupling limit the particle-particle d-wave channel leads to an instability driven by antiferromagnetic fluctuations as temperature decreases, as found in TPSC.

To understand the effect of pressure on the phase diagram, note that U/t should decrease as pressure increases since the increase in the overlap between orbitals should lead mainly to an increase in t. Hence, as can be deduced from Fig. 37, applying pressure should lead to a decrease in the value of T_c at weak coupling, concomitant with the decrease in antiferromagnetic fluctuations that lead to pairing in the weak coupling case. This is indeed what pressure does experimentally in the case of electron-doped high-temperature superconductors,[141] reinforcing our argument that near optimal doping they are more weakly coupled. It is widely known on the other hand that pressure *increases* T_c in hole-doped systems. That is consistent with the strong-coupling result that we found in VCPT and CDMFT, namely that the maximum d-wave order parameter in that case scales with $J = 4t^2/U$, a quantity that increases with t and hence pressure. Whereas in the weak coupling case superconductivity is a secondary phenomenon that occurs after antiferromagnetic fluctuations have built up, in strong coupling they can be two distinct phenomena as can be seen from the phase diagram, even though they arise from the same microscopic exchange interaction represented by J.

2. Anomalous superconductivity near the Mott transition

Superconductivity in the underdoped regime is very much non-BCS. First of all, we notice in Fig. 35 obtained in CDMFT[17] that at strong coupling the d-wave superconducting order parameter vanishes as we move towards half-filling even in the absence of long-range antiferromagnetic order. In other words, the Mott phenomenon by itself suffices to destroy d-wave superconductivity. This conclusion is reinforced by the fact that at weak coupling ($U = 4t$) where there is no Mott localization, d-wave superconductivity survives at half-filling. In the presence of antiferromagnetic long-range order, that last statement would not be true, as confirmed by VCPT calculations in Fig. 36: at $U = 4t$ d-wave superconductivity survives at half-filling if we do not allow for antiferromagnetic long-range order but it disappears if we do. In BCS theory, the presence of an interaction J that leads to attraction in the d-wave channel would lead at $T = 0$ to d-wave superconductivity at all dopings including half-filling, unless we allow for competing long-range order. At strong coupling, no long-range order is necessary to destroy d-wave superconductivity.

Superconductivity at strong coupling[142,143] also differs from BCS in the origin of the condensation energy. Suppose we do BCS theory on the attractive Hubbard model. Then, as in the usual BCS model, kinetic energy is increased in the superconducting state because the Fermi surface is no-longer sharp. On the other hand, in the superconducting phase there is a gain in potential energy. The reverse is true at strong coupling. This result follows from DCA[120] and is in agreement with the kinetic energy drop in the superconducting state that has been estimated from the f-sum rule in optical conductivity experiments.[144–146] Photoemission data[147] had also suggested this kinetic energy drop in the superconducting state. A crossover from non-BCS-like to BCS behavior in the condensation mechanism as we go from underdoping to overdoping has also been seen recently experimentally.[144] We do not seem to have the resolution to find that crossover since the condensation energy becomes very small on the overdoped side. We expect that crossover from strong to weak coupling will also lead to a change from a kinetic-energy driven to a potential-energy driven pairing mechanism. This is confirmed by CDMFT calculations for the attractive Hubbard model.[148]

A third way in which superconductivity in the underdoped regime is non-BCS is that the drop in the order parameter as we go towards half-filling is accompanied by an increase in the gap as measured in the single-particle density of states. Fig. 6 of Ref. 149 summarizes the experimental evidence for the increase in the size of the gap. That increase, observed in the CDMFT calculation of the gap, is illustrated in Fig. 51.[17] That gap has essentially the same size as that observed in the normal pseudogap state[67].

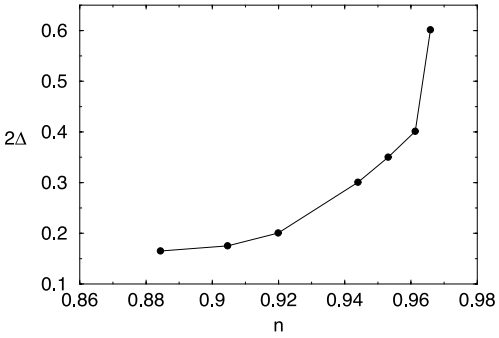

FIG. 51: The gap in the density of states of the dSC as a function of filling for $U = 8t$, $t' = -0.3t$ as calculated in CDMFT on a 2×2 cluster. From Ref. 17.

V. CONCLUSION, OPEN PROBLEMS

High-temperature superconductivity has forced both experimentalists and theorists to refine their tools and to develop new ones to solve the puzzles offered by this remarkable phenomenon. From a theoretical perspective, the original suggestion of Anderson[1] that the physics was in the one-band Hubbard model is being confirmed. In the absence of *ab initio* methods to tell us what is the correct starting point, such insight is essential. The non-perturbative nature of the phenomenon has however forced theorists to be extremely critical of each other's theories since none of them can pretend that a small parameter controls the accuracy of the approximations.

If theorists are to convince each other and experimentalists that a solution of the high-temperature superconductivity problem has been found, then the theories have to give quantitative results and to make predictions. Unlike most traditional problems in condensed matter physics however, the non-perturbative nature of the problem means that no simple mean-field like theory can be trusted, even if it seems to agree qualitatively with experiment. In fact several such theories have been proposed[13,14,150,151] not long after the experimental discovery of the phenomenon but they have not been accepted immediately. Theories have to be internally consistent, they have to agree with exact results whenever they are available, and then they can be compared with experiments. If there is a disagreement with experiment, the starting point (one-band Hubbard model) needs to be reconsidered. When approaches developed on the basis of weak-coupling ideas agree at intermediate coupling with approaches developed on the basis of strong-coupling ideas, then one gains confidence in the validity of the results. We have argued that such concordance is now found in a number of cases and that corresponding rather detailed quantitative agreement with experiment can be found. In a non-perturbative context it becomes essential to also cross check various approaches.

The main theoretical methods that we have discussed are those that we have developed or perfected or simply used in our group: The Two-Particle Self-

Consistent approach that is based on weak-coupling but non-perturbative ideas (no diagrams are involved), as well as heavily numerical approaches such as QMC and various quantum cluster methods, VCPT and CDMFT.

Based on our own work and that of many others, we think the following experimental facts about high-temperature superconductivity can be reproduced very accurately by calculations for the one-band Hubbard model with U in the intermediate coupling range ($U \sim 8t$) with $t \sim 350$ meV, and hopping parameters t' and t'' close to the values suggested by band structure calculations,[103] namely $t' = -0.3t$, $t'' = 0.2t$.

(i) In the one-band Hubbard model the main phases of the zero-temperature phase diagram, namely antiferromagnetic and d-wave superconducting, appear very near the observed ranges for both the hole- and electron-doped cases.

(ii) The normal state is unstable to a d-wave superconducting phase in a temperature range that has the correct order of magnitude. As usual the value of T_c is the most difficult quantity to evaluate since one must take into account Kosterlitz-Thouless physics as well as the effect of higher dimensions etc, so this level of agreement must be considered satisfying.

(iii) The ARPES MDC at the Fermi energy and the EDC near the Fermi energy are qualitatively well explained by cluster calculations for both hole- and electron-doped cases. These comparisons, made at a resolution of about 30 to 60 meV are not very sensitive to long-range order, although order does influence the results. One is mainly sensitive to the pseudogap, so this is the main phenomenon that comes out from the model. Energy resolution is not good enough to see a kink. More details about what aspects of ARPES are understood may be found in Sec. IV A.

(iv) In the case of electron-doped cuprates, the value of U near optimal doping seems to be in the range $U \sim 6t$, which means that it is accessible to studies with TPSC that have better resolution. In that case, the agreement with experiment is very accurate, even if there is room for improvement and a need for further experiments. In addition, the value of T^* for 13% doping has been predicted theoretically before it was observed experimentally, one of the very rare predictions in the field of high-temperature superconductors. All of this agreement with ARPES data is strong indication that $U \sim 6t$ is appropriate to describe electron-doped superconductors near optimal doping. Additional arguments come from the pressure dependence of the superconducting transition temperature T_c, which increases with t/U contrary to the strong-coupling result, and from simple ideas on Thomas-Fermi screening. The latter would predict that the screened interaction scales like $\partial\mu/\partial n$ and CPT results do lead to $\partial\mu/\partial n$ smaller on the electron- than on the hole-doped side.[77] In addition, the optical gap at half-filling is smaller in electron- than in hole-doped systems.

What is the physics? The physics of the antiferromagnetic phase at both weak and strong coupling is well understood and needs no further comment. For the pseudogap, we have argued that there seems to be two mechanisms, a weak coupling one that involves scattering off critical fluctuations and that is very well understood within TPSC, and a strong-coupling one where there is no need for large correlation lengths. There is no simple physical picture for the latter mechanism although the fact that it does not scale with J but with t seems to suggest forbidden hopping. The pseudogap is clearly different from the Mott gap. Whether there is a phase transition as a function of U that separates the weak and strong coupling regimes or whether there is only a crossover is an open question. The shape of the MDC's at the Fermi energy clearly show in any case that in some directions wave vector is not such a bad quantum number whereas in the pseudogap direction, a "localized" or "almost localized" particle-like picture would be appropriate. In fact the pseudogap occurs near the intersection with the antiferromagnetic zone boundary that turns out to also be the place where umklapp processes are possible. In other words, the presence of a lattice is extremely important for the appearance of the pseudogap. We have seen that with spherical Fermi surfaces the Fermi liquid survives even for large U. The dichotomy between the wave description inherent to the Fermi liquid and the particle (localized) description inherent to the Mott phenomenon seems to be resolved in the pseudogap phase by having certain directions where electrons are more wave-like and other directions where particle-like (gapped) behavior appears. The latter behavior appears near regions where the presence of the lattice is felt through umklapp processes.

It is clear that when weak-coupling-like ideas of quasiparticles scattering off each other and off collective excitations do not apply, a simple physical description becomes difficult. In fact, knowing the exact wave functions would give us the solution but we would not know how to understand "physically" the results.

This lack of simple physical images and the necessity to develop a new discourse is quite apparent for d-wave superconductivity. At weak coupling exchange of slow antiferromagnetic fluctuations is at the origin of the phenomenon, while at strong-coupling the fact that the maximum value of the d-wave order parameter scales with J tells us that this microscopic coupling is important, even though there is no apparent boson exchange. This is where mean-field like theories[13,14,150] or variational approaches[4] can help when they turn out to give results that are confirmed by more accurate and less biased methods.

There are many open problems, some of which are material dependent and hence may depend on interactions not included in the simplest Hubbard model. We have already mentioned the problem of the chemical potential shift in ARPES for very small dopings[117] that seems to be somewhat material dependent[118]. It would also be important to understand additional inhomogeneous phases that are observed in certain high-temperature supercon-

ductors. That is extremely challenging for quantum cluster methods and unlikely to be possible in the very near future, except for inhomogeneities of very short wave length. Also, we still need to improve concordance between the methods before we can make predictions that are quantitative at the few percent level for all physical quantities. Apart from DCA, there are no quantum cluster methods that have been developed yet to study two-particle response functions that are necessary to obtain results on the superfluid density and on transport in general. Transport studies are being completed in TPSC.[152] Such studies are crucial since they are needed to answer questions such as: (i) Why is it that for transport properties, such as optical conductivity, the number of carriers appears to scale with doping whereas in ARPES the surface of the Brillouin zone enclosed by the apparent Fermi surface appears to scale with the number of electrons? Is it because the weight of quasiparticles at the Fermi surface scales like the doping or because of vertex corrections or because of both? (ii) Can we explain a vanishing superfluid density as doping goes to zero[153] only through Mott physics or can competing order do the job.[163],[94]

After twenty years all the problems are not solved, but we think that we can say with confidence that the essential physics of the problem of high-temperature superconductivity is in the one-band Hubbard model. At least the pseudogap, the antiferromagnetic and the d-wave superconducting phases come out from the model. Refinements of that model may however be necessary as we understand more and more details of the material-specific experimental results.

Has a revolution been necessary to understand the basic physics of high-temperature superconductors? Certainly, it has been necessary to change our attitude towards methods of solution. We have seen that to study intermediate coupling, even starting from weak coupling, it has been necessary to drop diagrams and to rely instead on sum rules and other exact results to devise a non-perturbative approach. At strong coupling we had to accept that numerical methods are essential for progress and that we need to abandon some of the traditional physical explanations of the phenomena in terms of elementary excitations. Even though progress has been relatively slow, the pace is accelerating in the last few years and there is hope that in a few years the problem will be considered for the most part solved. The theoretical methods (numerical and analytical) that have been developed and that still need to be developed will likely remain in the tool box of the theoretical physicist and will probably be useful to understand and perhaps even design other yet undiscovered materials with interesting properties. The success will have been the result of the patient and focused effort of a large community of scientists fascinated by the remarkable phenomenon of high-temperature superconductivity.

Acknowledgments

The present work was supported by NSERC (Canada), FQRNT (Québec), CFI (Canada), CIAR, the Tier I Canada Research chair Program (A.-M.S.T.). We are grateful to our collaborators, G. Albinet, S. Allen, M. Boissonneault, C. Brillon, M. Capone, L. Chen, M. Civelli, A.-M. Daré, B. Davoudi, J.-Y. P. Delannoy, A. Gagné-Lebrun, M. J. P. Gingras, A. Georges, M. Guillot, V. Hankevych, P. C. W. Holdsworth, F. Jackson, S. Kancharla, G. Kotliar, J.-S. Landry, P.-L. Lavertu, F. Lemay, S. Lessard, M.-A. Marois, S. Pairault, D. Perez, M. Pioro-Ladrière, D. Plouffe, D. Poulin, L. Raymond, S. Roy, P. Sahebsara, H. Touchette, and especially Y.M. Vilk. We also acknowledge useful discussions with P. Fournier, M. Greven, I. Herbut, K. Shen and L. Taillefer and we are grateful to V. Hankevych and S. Kancharla for permission to include some of their unpublished figures in this paper.

APPENDIX A: LIST OF ACRONYMS

ARPES: Angle Resolved Photoemission Spectroscopy: Experiment from which one can extract $A(\mathbf{k},\omega)f(\omega)$.

CPT: Cluster Perturbation Theory: Cluster method based on strong coupling perturbation theory.[52,53,62]

CDMFT: Cellular Dynamical Mean Field Theory: A cluster generalization of DMFT that allows one to take into account both wave vector and frequency dependence of the self-energy.[58] It is best formulated in real space.

DCA: Dynamical Cluster approximation: A cluster generalization of DMFT that allows one to take into account both wave vector and frequency dependence of the self-energy based on coarse graining of the self-energy in reciprocal space.[55,154]

DMFT: Dynamical Mean Field Theory: This approach is exact in infinite dimension. It takes the frequency dependence of the self-energy into account and includes both the Mott and the Fermi liquid limits.[2,3]

EDC: Energy Dispersion Curves: A representation of $A\left(\mathbf{k},\omega\right) f\left(\omega\right)$ at fixed $\mathbf{k}$ as a function of ω.

FLEX: Fluctuation Exchange Approximation: A conserving many-body approach, similar in spirit to Eliashberg theory.[22]

MDC: Momentum Dispersion Curves: A representation of $A(\mathbf{k},\omega)f(\omega)$ at fixed ω as a function of $\mathbf{k}$.

QMC: Quantum Monte Carlo: Determinental approach[155]. This provides an essentially exact solution to the model for a given system size

and within statistical errors that can be made smaller by performing more measurements.

RPA: Random Phase Approximation.

TPSC: Two-Particle Self-Consistent Approach: Based on sum rules and other constraints, allows to treat the Hubbard model non-perturbatively in the weak to intermediate coupling limit.[23,34]

VCA: Variational Cluster Approach. Analogous to *CDMFT* but with a convergence criterion based on an extremum principle. In the applications quoted here, there is no bath.[56,64,73,156]

VCPT: Variational Cluster Perturbation Theory. In the present paper synonymous with *VCA*.

[1] P. Anderson, Science **235**, 1196 (1987).
[2] A. Georges, G. Kotliar, W. Krauth and M. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[3] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992); M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992).
[4] A. Paramekanti, M. Randeria and N. Trivedi, Phys. Rev. B **70**, 54504 (2004).
[5] J. Hirsch and H. Lin, Phys. Rev. B **37**, 5070 (1988).
[6] M. Beal-Monod, C. Bourbonnais and V. Emery, Phys. Rev. B **34**, 7716 (1986).
[7] K. Miyake, S. Schmitt-Rink and C. Varma, Phys. Rev. B **34**, 6554 (1986).
[8] D. Scalapino, E. Loh and J. Hirsch, Phys. Rev. B **34**, 8190 (1986).
[9] W. Kohn and J. M. Luttinger, Phys. Rev. Lett. **15**, 524 (1965).
[10] A. Leggett, Rev. Mod. Phys. **47**, 331 (1975).
[45] N. Bulut, D.J. Scalapino, S.R. White, Phys. Rev. B **47**, 14 599 (1993).
[12] S. Zhang, J. Carlson and J. E. Gubernatis, Phys. Rev. Lett. **78**, 4486 (1997).
[13] M. Inui, S. Doniach, P. Hirschfeld and A. Ruckenstein, Phys. Rev. B **37**, 2320 (1988).
[14] G. Kotliar and J. Liu, Phys. Rev. Lett. **61**, 1784 (1988).
[15] T. Maier, M. Jarrell, T. Pruschke and J. Keller, Phys. Rev. Lett. **85**, 1524 (2000).
[16] D. Poilblanc and D. Scalapino, Phys. Rev. B **66**, 052513 (2002).
[17] S. S. Kancharla, M. Civelli, M. Capone, B. Kyung, D. Sénéchal, G. Kotliar and A.-M. Tremblay, *Anomalous superconductivity in doped Mott insulators* (2005), cond-mat/0508205.
[18] T. Maier, M. Jarrell, T. Schulthess, P. Kent and J. White, hys. Rev. Lett. 95, 237001 (2005).
[19] D. Sénéchal, P.-L. Lavertu, M.-A. Marois and A.-M. S. Tremblay, Phys. Rev. Lett. **94**, 156404 (2005).
[20] T. Moriya, *Spin Fluctuations in Itinerant Electron Magnetism* (Springer-Verlag, 1985).
[21] R. S. Markiewicz, Phys. Rev. B **70**, 174518 (2004).
[22] N. Bickers and D. Scalapino, Ann. Phys. (USA)**193**, 206 (1989).
[23] Y. Vilk and A.-M. Tremblay, J. Phys I (France) **7**, 1309 (1997).
[24] Y. Vilk and A.-M. Tremblay, Europhys. Lett. **33**, 159 (1996).
[25] N. Furukawa, T. M. Rice and M. Salmhofer, Phys. Rev. Lett. **81**, 3195 (1998).
[26] C. Halboth and W. Metzner, Phys. Rev. B **61**, 7364 (2000).
[27] V. Hankevych and F. Wegner, Eur. Phys. J. B **31**, 333 (2003).
[28] C. Honerkamp and M. Salmhofer, Phys. Rev. Lett. **87**, 187004 (2001).
[29] F. Wegner and V. Hankevych, Acta Phys. Pol. B **34**, 497 (2003).
[30] A. A. Katanin and A. P. Kampf, Phys. Rev. Lett. **93**, 106406 (2004).
[31] D. Zanchi, Europhys. Lett. **55**, 376 (2001).
[32] G. Mahan, *Many-Particle Physics, 3rd edition, Section 6.4.4* (Kluwer/Plenum, 2000).
[33] Y. Vilk and A.-M. Tremblay, J. Phys. Chem. Solids (UK) **56**, 1769 (1995).
[34] S. Allen, A.-M. Tremblay and Y. M. Vilk, in *Theoretical Methods for Strongly Correlated Electrons*, edited by D. Sénéchal, C. Bourbonnais and A.-M. Tremblay, CRM Series in Mathematical Physics, Springer (2003).
[35] G. Baym, Physical review **127**, 1391 (1962).
[36] S. Moukouri, S. Allen, F. Lemay, B. Kyung, D. Poulin, Y. Vilk and A.-M. Tremblay, Phys. Rev. B **61**, 7887 (2000).
[37] Y. Vilk, L. Chen and A.-M. Tremblay, Phys. Rev. B **49**, 13267 (1994).
[38] S. Roy, C. Brillon and A.-M. Tremblay, unpublished.
[39] S. Allen, B. Kyung and A.-M. Tremblay, unpublished.
[40] B. Kyung, J.-S. Landry and A. M. S. Tremblay, Phys. Rev. B **68**, 174502 (2003).
[41] S. Allen and A.-M. Tremblay, Phys. Rev. B **64**, 075115 (2001).
[42] L. Lilly, A. Muramatsu and W. Hanke, Phys. Rev. Lett. **65**, 1379 (1990).
[43] B. Kyung, S. Allen and A.-M. Tremblay, Phys. Rev. B **64**, 075116 (2001).
[44] S. White, D. J. Scalapino and R. Sugar, Phys. Rev. B **39**, 839 (1989).
[45] N. Bulut, D. Scalapino and S. White, Phys. Rev. B **47**, 2742 (1993).
[46] A. Daré, Y. Vilk and A.-M. Tremblay, Phys. Rev. B **53**, 14236 (1996).
[47] L. Chen, C. Bourbonnais, T. Li and A.-M.-S. Tremblay, Phys. Rev. Lett. **66**, 369 (1991).
[48] V. Hankevych, B. Kyung and A.-M. Tremblay, Phys. Rev. B **68**, 214405 (2003).
[49] N. Trivedi and M. Randeria, Phys. Rev. Lett. **75**, 312 (1995).
[50] A. Moreo, D. J. Scalapino and S. R. White, Phys. Rev. B **45**, 7544 (1992).
[51] G . Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, C.A. Marianetti, *Electronic Structure Calculations with Dynamical Mean-Field Theory: A Spectral Density Functional Approach* (2005), cond-mat/0511085.

[52] C. Gros and R. Valenti, Phys. Rev. B **48**, 418 (1993).
[53] D. Sénéchal, D. Perez and M. Pioro-Ladriere, Phys. Rev. Lett. **84**, 522 (2000).
[54] A. Schiller and K. Ingersent Phys. Rev. Lett. **75**, 113 (1975)
[55] M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke and H. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998).
[56] M. Potthoff, Eur. Phys. J. B **36**, 335 (2003).
[57] M. Potthoff, M. Aichhorn and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003).
[58] G. Kotliar, S. Savrasov, G. Pálsson and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001).
[59] T. Stanescu and G. Kotliar, *Cluster Dynamical Mean Field Theories: A strong coupling perspective* (2005), cond-mat/0508302.
[60] T. D. Stanescu and G. Kotliar, Phys. Rev. B **70**, 205112 (2004).
[61] T. Maier, M. Jarrell, T. Pruschke and M. H. Hettler, Rev. Mod. Phys. **77**, 1027 (2005).
[62] D. Sénéchal, D. Perez and D. Plouffe, Phys. Rev. B **66**, 075129 (2002).
[63] M. Potthoff, *Non-perturbative construction of the Luttinger-Ward functional* (2004), cond-mat/0406671.
[64] M. Potthoff, Adv. Solid State Phys. **45**, 135 (2005).
[65] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994).
[66] C. J. Bolech, S. S. Kancharla and G. Kotliar, Phys. Rev. B **67**, 075110 (2003).
[67] B. Kyung, S. Kancharla, D. Sénéchal, A. M. S. Tremblay, M. Civelli and G. Kotliar, *Short-range correlation induced pseudogap in doped Mott insulators* (2005), cond-mat/0502565, Phys. Rev. B in press.
[68] D. Sénéchal, in *High Performance Computing Systems and Applications and OSCAR symposium*, edited by D. Sénéchal (2003).
[69] K. Aryanpour, T. A. Maier and M. Jarrell, Phys. Rev. B **71**, 037101 (2005).
[70] G. Biroli and G. Kotliar, Phys. Rev. B **65**, 155112 (2002).
[71] G. Biroli and G. Kotliar, Phys. Rev. B **71**, 037102 (2005); B. Kyung, G. Kotliar, A. -M. S. Tremblay, *Quantum Monte Carlo Study of Strongly Correlated Electrons: Cellular Dynamical Mean-Field Theory*, (2006) cond-mat/0601271.
[72] J. Favand, S. Haas, K. Penc, F. Mila and E. Dagotto, Phys. Rev. B **55**, R4859 (1997).
[73] C. Dahnken, M. Aichhorn, W. Hanke, E. Arrigoni and M. Potthoff, Phys. Rev. B **70**, 245110 (2004).
[74] M. Capone, M. Civelli, S. S. Kancharla, C. Castellani and G. Kotliar, Phys. Rev. B **69**, 195105 (2004).
[75] T. A. Maier, O. Gonzalez, M. Jarrell and T. Schulthess, *Two quantum cluster approximations* (2002), cond-mat/0205460.
[76] K. Pozgajcic, *Quantitative aspects of the dynamical impurity approach* (2004), cond-mat/0407172.
[77] D. Sénéchal and A.-M. S. Tremblay, Phys. Rev. Lett. **92**, 126401 (2004).
[78] C. Huscroft, M. Jarrell, T. Maier, S. Moukouri and A. Tahvildarzadeh, Phys. Rev. Lett. **86**, 139 (2001).
[79] C. Grober, R. Eder and W. Hanke, Phys. Rev. B **62**, 4336 (2000).
[80] A. Moreo, S. Haas, A. Sandvik and E. Dagotto, Phys. Rev. B **51**, 12045 (1995).
[81] S. Moukouri and M. Jarrell, Phys. Rev. Lett. **87**, 167010 (2001).
[82] A.Brooks Harris and Robert V. Lange, Phys. Rev. **157**, 295 (1967).
[83] M.B.J. Meinders, G.A. Sawatzky and G.A. Sawatzky, Phys. Rev. B **48**, 3916 (1993).
[84] B. Kyung, J. Landry, D. Poulin and A.-M.S. Tremblay, Phys. Rev. Lett. **90**, 099702 (2003).
[85] F. Ronning et al., Phys. Rev. B **67**, 165101 (2003).
[86] N. Armitage et al., Phys. Rev. Lett. **88**, 257001 (2002).
[87] B. Kyung, V. Hankevych, A.-M. Daré and A.-M. S. Tremblay, Phys. Rev. Lett. **93**, 147004 (2004).
[88] Q. Yuan, F. Yuan and C. S. Ting, Phys. Rev. B **72**, 054504 (2005).
[89] H. Kusunose and T. M. Rice, Phys. Rev. Lett. **91**, 186407 (2003).
[90] C. Kusko, R. S. Markiewicz, M. Lindroos and A. Bansil, Phys. Rev. B **66**, 140513 (2002).
[91] K. Borejsza and N. Dupuis, Phys. Rev. B **69**, 085119 (2004).
[92] J. Carbotte, E. Schachinger and D. Basov, Nature **401**, 354 (1999).
[93] D. Scalapino, Physics Reports **250**, 329 (1995).
[94] B. Tobijaszewska and R. Micnas, Physica Status Solidi B **242**, 468 (2005).
[95] A. Moreo and D. J. Scalapino, Phys. Rev. B **43**, 8211 (1991).
[96] R. Scalettar and R. Singh, Phys. Rev. Lett. **67**, 370 (1991).
[97] A. Paramekanti, M. Randeria and N. Trivedi, Phys. Rev. Lett. **87**, 217002 (2001).
[98] S. Sorella, G. B. Martins, F. Becca, C. Gazza, L. Capriotti, A. Parola and E. Dagotto, Phys. Rev. Lett. **88**, 117002 (2002).
[99] W. Hanke, M. Aichhorn, E. Arrigoni and M. Potthoff, *Correlated band structure and the ground-state phase diagram in high-T_c cuprates* (2005), cond-mat/0506364.
[100] C. Pao and N. Bickers, Phys. Rev. B **49**, 1586 (1994).
[101] J. Deisz, D. Hess and J. Serene, Phys. Rev. Lett. **76**, 1312 (1996).
[102] F. Zhang and T. M. Rice, Phys. Rev. B **37**, 3759 (1988).
[103] O. Andersen, A. Liechtenstein, O. Jepsen and F. Paulsen, Journal of the Physics and Chemistry of Solids **56**, 1573 (1995).
[104] A. Macridin, T. A. Maier, M. S. Jarrell and G. Sawatzky, Phys. Rev. B **71**, 134527 (2005).
[105] R. Coldea, S. M. Hayden, G. Aeppli, T. G. Perring, C. D. Frost, T. E. Mason, S.-W. Cheong and Z. Fisk, Phys. Rev. Lett. **86**, 5377 (2001).
[106] J.-Y. Delannoy, Ph.D. thesis, École normale supérieure de Lyon (2005).
[107] J.-Y. P. Delannoy, M. J. P. Gingras, P. C. W. Holdsworth and A.-M. S. Tremblay, Phys. Rev. B **72**, 115114 (2005).
[108] L. Raymond, G. Albinet and A.-M. Tremblay, cond-mat/0510811.
[109] A. M. Toader, J. P. Goff, M. Roger, N. Shannon, J. R. Stewart and M. Enderle, Phys. Rev. Lett. **94**, 197202 (2005).
[110] A. Damascelli, Z. Hussain and Z.-X. Shen, Rev. Mod. Phys. **75**, 473 (2003).
[111] A. Damascelli, *Probing the low-energy electronic structure of complex systems by ARPES* (2003), cond-mat/0307085.
[112] A. Kordyuk and S. V. Borisenko, *ARPES on HTSC:*

simplicity vs. complexity (2005), cond-mat/0510218.
[113] P. Preuss, W. Hanke and W. von der Linden, Phys. Rev. Lett. **75**, 1344 (1995).
[114] S. Pairault, D. Sénéchal and A.-M. Tremblay, Eur. Phys. J. B (France) **16**, 85 (2000).
[115] S. Pairault, D. Sénéchal and A.-M. S. Tremblay, Phys. Rev. Lett. **80**, 5389 (1998).
[116] R. Gooding, K. Vos and P. Leung, Phys. Rev. B **50**, 12866 (1994).
[117] K. Shen, F. Ronning, D. Lu, F. Baumberger, N. Ingle, W. Lee, W. Meevasana, Y. Kohsaka, M. Azuma, M. Takano et al., Science **307**, 901 (2005).
[118] T. Yoshida, X. J. Zhou, K. Tanaka, W. L. Yang, Z. Hussain, Z.-X. Shen, A. Fujimori, S. Komiya, Y. Ando, H. Eisaki et al., cond-mat/0510608.
[119] X. Zhou, T. Yoshida, A. Lanzara, P. V. Bogdanov, S. A. Kellar, K. M. Shen, W. L. Yang, F. Ronning, T. Sasagawa, T. Kakeshita et al., Nature **423**, 398 (2003).
[120] T. A. Maier, M. Jarrell, A. Macridin and C. Slezak, Phys. Rev. Lett. **92**, 027005 (2004); T. Koretsune and M. Ogata, J. Phys. Soc. Japan, **74**, 1390 (2005).
[121] A. D. Gromko, A. V. Fedorov, Y.-D. Chuang, J. D. Koralek, Y. Aiura, Y. Yamaguchi, K. Oka, Y. Ando and D. S. Dessau, Phys. Rev. B **68**, 174520 (2003).
[122] N. P. Armitage, D. H. Lu, C. Kim, A. Damascelli, K. M. Shen, F. Ronning, D. L. Feng, P. Bogdanov, X. J. Zhou, W. L. Yang et al., Phys. Rev. B **68**, 064517 (2003).
[123] V. Hankevych, B. Kyung, A.-M. Daré, D. Sénéchal and A.-M. Tremblay, in *Proceedings of SNS2004* (2005), J. Chem. Phys. Sol., **67**, 189 (2006).
[124] P. K. Mang, O. P. Vajk, A. Arvanitaki, J. W. Lynn and M. Greven, Phys. Rev. Lett. **93**, 027002 (2004).
[125] M. Matsuda, Y. Endoh, K. Yamada, H. Kojima, I. Tanaka, R. J. Birgeneau, M. A. Kastner and G. Shirane, Phys. Rev. B **45**, 12548 (1992).
[126] Y. Onose, Y. Taguchi, K. Ishizaka and Y. Tokura, Phys. Rev. B. **69**, 024504 (2004).
[127] Y. Onose, Y. Taguchi, K. Ishizaka and Y. Tokura, Phys. Rev. Lett. **87**, 217001 (2001).
[128] H. Matsui, K. Terashima, T. Sato, T. Takahashi, S.-C. Wang, H.-B. Yang, H. Ding, T. Uefuji and K. Yamada, Phys. Rev. Lett. **94**, 047005 (2005).
[129] R. P. S. M. Lobo, N. Bontemps, A. Zimmers, Y. Dagan, R. Greene, P. Fournier, C. Homes and A. Millis, unpublished.
[130] A.J. Millis, A. Zimmers, R.P.S.M. Lobo, N. Bontemps, C.C. Homes Phys. Rev. B **72** 224517 (2005).
[131] T. Hanaguri, C. Lupien, Y. Kohsaka, D. Lee, M. Azuma, M. Takano, H. Takagi and J. Davis, Nature **430**, 1001 (2004).
[132] C. Stock, W. J. L. Buyers, R. Liang, D. Peets, Z. Tun, D. Bonn, W. N. Hardy and R. J. Birgeneau, Phys. Rev. B **69**, 014502 (2004).
[133] M. Norman, D. Pines and C. Kallin, Adv. Phys. **54**, 715 (2005).
[134] P. Fournier, P. Mohanty, E. Maiser, S. Darzens, T. Venkatesan, C. Lobb, G. Czjzek, R. Webb and R. Greene, Phys. Rev. Lett. **81**, 4720 (1998).
[135] E. Motoyama, P. Mang, D. Petitgrand, G. Yu, O. Vajk, I. Vishik and M. Greven, unpublished.
[136] J. Sonier, K. Poon, G. Luke, P. Kyriakou, R. Miller, R. Liang, C. Wiebe, P. Fournier and R. Greene, Phys. Rev. Lett. **91**, 147002 (2003).
[137] J. Sonier, K. Poon, G. Luke, P. Kyriakou, R. Miller, R. Liang, C. Wiebe, P. Fournier and R. Greene, Physica C: Superconductivity and its Applications **408-410**, 783 (2004).
[138] S. Wakimoto, R. J. Birgeneau, M. A. Kastner, Y. S. Lee, R. Erwin, P. M. Gehring, S. H. Lee, M. Fujita, K. Yamada, Y. Endoh et al., Phys. Rev. B **61**, 3699 (2000).
[139] S. Wakimoto, J. M. Tranquada, T. Ono, K. M. Kojima, S. Uchida, S.-H. Lee, P. M. Gehring and R. J. Birgeneau, Phys. Rev. B **64**, 174505 (2001).
[140] T. A. Maier, M. S. Jarrell and D. J. Scalapino, Phys. Rev. Lett. **96**, 047005 (2006).
[141] M. Maple, MRS Bulletin **15**, 60 (1990).
[142] R. Haslinger and A. Chubukov, Phys. Rev. B **67**, 140504 (2003).
[143] F. Marsiglio, Phys. Rev. B **73**, 064507 (2006).
[144] G. Deutscher, A. F. Santander-Syro and N. Bontemps, Phys. Rev. B **72**, 092504 (2005).
[145] H. Molegraaf, C. Presura, D. van der Marel, P. Kes and M. Li, Science **295**, 2239 (2002).
[146] A. Santander-Syro, R. Lobo, N. Bontemps, Z. Konstatinovic, Z. Li and H. Raffy, Europhys. Lett. **62**, 568 (2003).
[147] M. Norman, M. Randeria, B. Janko and J. Campuzano, Phys. Rev. B **61**, 14742 (2000).
[148] B. Kyung, A. Georges and A.-M. Tremblay, *Potential-energy (BCS) to kinetic-energy (BEC)-driven pairing in the attractive Hubbard model* (2005), cond-mat/0508645.
[149] M. Sutherland, D. G. Hawthorn, R. W. Hill, F. Ronning, S. Wakimoto, H. Zhang, C. Proust, E. Boaknin, C. Lupien, L. Taillefer et al., Phys. Rev. B **67**, 174520 (2003).
[150] P. Anderson, *Present status of the theory of high-T_c cuprates* (2005), cond-mat/0510053 published in this issue of Journal of Low Temperature Physics.
[151] N.E. Bickers, D.J. Scalapino and R.T. Scalettar, Int. J. Mod. Phys. B **1**, 687 (1987).
[152] S. Allen, B. Kyung, D. Bergeron, V. Hankevych and A.-M. Tremblay, unpublished.
[153] D. M. Broun, P. J. Turner, W. A. Huttema, S. Ozcan, B. Morgan, L. R., W. N. Hardy and D. A. Bonn, *In-plane superfluid density of highly underdoped $YBa_2Cu_3O_{6+x}$* (2005), cond-mat/0509223.
[154] M. Hettler, M. Mukherjee, M. Jarrell and H. Krishnamurthy, Phys. Rev. B **61**, 12739 (2000).
[155] R. Blankenbecler, D. J. Scalapino and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).
[156] M. Potthoff, Eur. Phys. J. B (France) **32**, 429 (2003).
[157] This additional constraint leads to a degree of consistency between one- and two-particle quantities that is absent from the standard Hartree-Fock factorization.
[158] The constant self-energy is *not* equal to $Un_{-\sigma}$ as it would in the trivial Hartree-Fock factorization (see previous note).
[159] Th. Maier, M. Potthoff and D. Sénéchal, unpublished.
[160] The latter result is disputed by a recent DCA calculation by Alexandru Macridin, Mark Jarrell, Thomas Maier, P. R. C. Kent, cond-mat/0509166.
[161] Kyle Shen, Private communication.
[162] M. Greven, private communication.
[163] I. Herbut, private communication

J. Phys. I France **7** (1997) 1309–1368 NOVEMBER **1997**, PAGE 1309

Non-Perturbative Many-Body Approach to the Hubbard Model and Single-Particle Pseudogap

Y.M. Vilk ([1,2]) and A.-M.S. Tremblay ([2,*])

([1]) Materials Science Division, Bldg. 223, 9700 S. Case Ave., Argonne National Laboratory, Argonne IL 60439, USA

([2]) Département de Physique and Centre de Recherche en Physique du Solide, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1

(Received 7 July 1997, accepted 23 July 1997)

PACS.71.10.Fd – Lattice fermion models (Hubbard model, etc.)

PACS.71.10.Hf – Non-Fermi-liquid ground states, electron phase diagrams and phase transitions in model systems

PACS.71.10.Ca – Electron gas, Fermi gas

Abstract. — A new approach to the single-band Hubbard model is described in the general context of many-body theories. It is based on enforcing conservation laws, the Pauli principle and a number of crucial sum-rules. More specifically, spin and charge susceptibilities are expressed, in a conserving approximation, as a function of two irreducible vertices whose values are found by imposing the local Pauli principle $\langle n_\uparrow^2 \rangle = \langle n_\uparrow \rangle$ as well as the local-moment sum-rule and consistency with the equations of motion in a local-field approximation. The Mermin-Wagner theorem in two dimensions is automatically satisfied. The effect of collective modes on single-particle properties is then obtained by a paramagnon-like formula that is consistent with the two-particle properties in the sense that the potential energy obtained from Tr ΣG is identical to that obtained using the fluctuation-dissipation theorem for susceptibilities. Since there is no Migdal theorem controlling the effect of spin and charge fluctuations on the self-energy, the required vertex corrections are included. It is shown that the theory is in quantitative agreement with Monte Carlo simulations for both single-particle and two-particle properties. The theory predicts a magnetic phase diagram where magnetic order persists away from half-filling but where ferromagnetism is completely suppressed. Both quantum-critical and renormalized-classical behavior can occur in certain parameter ranges. It is shown that in the renormalized classical regime, spin fluctuations lead to precursors of antiferromagnetic bands (shadow bands) and to the destruction of the Fermi-liquid quasiparticles in a wide temperature range above the zero-temperature phase transition. The upper critical dimension for this phenomenon is three. The analogous phenomenon of pairing pseudogap can occur in the attractive model in two dimensions when the pairing fluctuations become critical. Simple analytical expressions for the self-energy are derived in both the magnetic and pairing pseudogap regimes. Other approaches, such as paramagnon, self-consistent fluctuation exchange approximation (FLEX), and pseudo-potential parquet approaches are critically compared. In particular, it is argued that the failure of the FLEX approximation to reproduce the pseudogap and the precursors AFM bands in the weak coupling regime and the Hubbard bands in the strong coupling regime is due to inconsistent treatment of vertex corrections in the expression for the self-energy. Treating the spin fluctuations as if there was a Migdal's theorem can lead not only to quantitatively wrong results but also to qualitatively wrong predictions, in particular with regard to the single-particle pseudogap.

([*]) Author for correspondence (e-mail: tremblay@physique.usherb.ca)

© Les Éditions de Physique 1997

1. Introduction

Understanding all the consequences of the interplay between band structure effects and electron-electron interactions remains one of the present-day goals of theoretical solid-state Physics. One of the simplest model that contains the essence of this problem is the Hubbard model. In the more than thirty years [1, 2] since this model was formulated, much progress has been accomplished. In one dimension [3, 4], various techniques such as diagrammatic resummations [5], bosonization [6], renormalization group [7, 8] and conformal approaches [9, 10] have lead to a very detailed understanding of correlation functions, from weak to strong coupling. Similarly, in infinite dimensions a dynamical mean-field theory [11] leads to an essentially exact solution of the model, although many results must be obtained by numerically solving self-consistent integral equations. Detailed comparisons with experimental results on transition-metal oxides have shown that three-dimensional materials can be well described by the infinite-dimensional self-consistent mean-field approach [11]. Other methods, such as slave-boson [12] or slave-fermion [13] approaches, have also allowed one to gain insights into the Hubbard model through various mean-field theories corrected for fluctuations. In this context however, the mean-field theories are not based on a variational principle. Instead, they are generally based on expansions in the inverse of a degeneracy parameter [14], such as the number of fermion flavors N, where N is taken to be large despite the fact that the physical limit corresponds to a small value of this parameter, say $N = 2$. Hence these theories must be used in conjunction with other approaches to estimate their limits of validity [15]. Expansions around solvable limits have also been explored [16]. Finally, numerical solutions [17], with proper account of finite-size effects, can often provide a way to test the range of validity of approximation methods independently of experiments on materials that are generally described by much more complicated Hamiltonians.

Despite all this progress, we are still lacking reliable theoretical methods that work in arbitrary space dimension. In two dimensions in particular, it is believed that the Hubbard model may hold the key to understanding normal state properties of high-temperature superconductors. But even the simpler goal of understanding the magnetic phase diagram of the Hubbard model in two dimensions is a challenge. Traditional mean-field techniques, or even slave-boson mean-field approaches, for studying magnetic instabilities of interacting electrons fail in two dimensions. The Random Phase Approximation (RPA) for example does not satisfy the Pauli principle, and furthermore it predicts finite temperature antiferromagnetic or Spin Density Wave (SDW) transitions while this is forbidden by the Mermin-Wagner theorem. Even though one can study universal critical behavior using various forms of renormalization group treatments [18–22] or through the self-consistent-renormalized approach of Moriya [23] which all satisfy the Mermin-Wagner theorem in two dimensions, cutoff-dependent scales are left undetermined by these approaches. This means that the range of interactions or fillings for which a given type of ground-state magnetic order may appear is left undetermined.

Amongst the recently developed theoretical methods for understanding both collective and single-particle properties of the Hubbard model, one should note the fluctuation exchange approximation [24] (FLEX) and the pseudo-potential parquet approach [25]. The first one, FLEX, is based on the idea of conserving approximations proposed by Baym and Kadanoff [26,27]. This approach starts with a set of skeleton diagrams for the Luttinger-Ward functional [28] to generate a self-energy that is computed self-consistently. The choice of initial diagrams however is arbitrary and left to physical intuition. In the pseudo-potential parquet approach, one parameterizes response functions in all channels, and then one iterates crossing-symmetric many-body integral equations. While the latter approach partially satisfies the Pauli principle, it violates conservation laws. The opposite is true for FLEX.

In this paper, we present the formal aspects of a new approach that we have recently developed for the Hubbard model [29, 30]. The approach is based on enforcing sum rules and conservation laws, rather than on diagrammatic perturbative methods that are not valid for interaction U larger than hopping t. We first start from a Luttinger-Ward functional that is parameterized by two irreducible vertices $U_{\rm sp}$ and $U_{\rm ch}$ that are local in space-time. This generates RPA-like equations for spin and charge fluctuations that are conserving. The local-moment sum rule, local charge sum rule, and the constraint imposed by the Pauli principle, $\langle n_\uparrow^2\rangle = \langle n_\uparrow\rangle$ then allow us to find the vertices as a function of double occupancy $\langle n_\uparrow n_\downarrow\rangle$ (see Eqs. (37, 38)). Since $\langle n_\uparrow n_\downarrow\rangle$ is a local quantity it depends very little on the size of the system and, in principle, it could be obtained reliably using numerical methods, such as for example Monte Carlo simulations. Here, however, we adopt another approach and find $\langle n_\uparrow n_\downarrow\rangle$ self-consistently [29] without any input from outside the present theory. This is done by using an ansatz equation (40) for the double-occupancy $\langle n_\uparrow n_\downarrow\rangle$ that has been inspired by ideas from the local field approach of Singwi *et al.* [31]. Once we have the spin and charge fluctuations, the next step is to use them to compute a new approximation, equation (46), for the single-particle self-energy. This approach to the calculation of the effect of collective modes on single-particle properties [30] is similar in spirit to paramagnon theories [32]. Contrary to these approaches however, we do include vertex corrections in such a way that, if $\Sigma^{(1)}$ is our new approximation for the self-energy while $G^{(0)}$ is the initial Green's function used in the calculation of the collective modes, and $\langle n_\uparrow n_\downarrow\rangle$ is the value obtained from spin and charge susceptibilities, then $\frac{1}{2}{\rm Tr}\left[\Sigma^{(1)}G^{(0)}\right] = U\langle n_\uparrow n_\downarrow\rangle$ is satisfied exactly. The extent to which $\frac{1}{2}{\rm Tr}\left[\Sigma^{(1)}G^{(1)}\right]$ (computed with $G^{(1)}$ instead of $G^{(0)}$) differs from $U\langle n_\uparrow n_\downarrow\rangle$ can then be used both as an internal accuracy check and as a way to improve the vertex corrections.

If one is interested only in two-particle properties, namely spin and charge fluctuations, then this approach has the simple physical appeal of RPA but it satisfies key constraints that are always violated by RPA, namely the Mermin-Wagner theorem and the Pauli principle. To contrast it with usual RPA, that has a self-consistency only at the single-particle level, we call it the Two-Particle Self-Consistent approach (TPSC) [29, 30, 33]. The TPSC gives a quantitative description of the Hubbard model not only far from phase transitions, but also upon entering the critical regime. Indeed we have shown quantitative agreement with Monte Carlo simulations of the nearest-neighbor [29] and next-nearest neighbor [34] Hubbard model in two dimensions. Quantitative agreement is also obtained as one enters the narrow critical regime accessible in Monte Carlo simulations. We also have shown [33] in full generality that the TPSC approach gives the $n\to\infty$ limit of the $O(n)$ model, while $n=3$ is the physically correct (Heisenberg) limit. In two dimensions, we then recover both quantum-critical [19] and renormalized classical [18] regimes to leading order in $1/n$. Since there is no arbitrariness in cutoff, given a microscopic Hubbard model no parameter is left undetermined. This allows us to go with the same theory from the non-critical to the beginning of the critical regime, thus providing quantitative estimates for the magnetic phase diagram of the Hubbard model, not only in two dimensions but also in higher dimensions [33].

The main limitation of the approach presented in this paper is that it is valid only from weak to intermediate coupling. The strong-coupling case cannot be treated with frequency-independent irreducible vertices, as will become clear later. However, a suitable ansatz for these irreducible vertices in a Luttinger-Ward functional might allow us to apply our general scheme to this limit as well.

Our approach predicts [30] that in two dimensions, Fermi liquid quasiparticles disappear in the renormalized classical regime $\xi_{\rm AFM}\propto\exp({\rm const}/T)$, which always precedes the zero-temperature phase transition in two-dimensions. In this regime the antiferromagnetic correlation length becomes larger than the single-particle thermal de Broglie wave length

$\xi_{\rm th}(= v_{\rm F}/T)$, leading to the destruction of Fermi liquid quasiparticles with a concomitant appearance of precursors of antiferromagnetic bands ("shadow bands") with no quasi-particle peak between them. We stress the crucial role of the classical thermal spin fluctuations and low dimensionality for the existence of this effect and contrast our results with the earlier results of Kampf and Schrieffer [35] who used a susceptibility separable in momentum and frequency $\chi_{\rm sp} = f(\mathbf{q})g(\omega)$. The latter form of $\chi_{\rm sp} = f(\mathbf{q})g(\omega)$ leads to an artifact that dispersive precursors of antiferromagnetic bands can exist at $T = 0$ (for details see [36]). We also contrast our results with those obtained in the fluctuation exchange approximation (FLEX), which includes self-consistency in the single particle propagators but neglects the corresponding vertex corrections. The latter approach predicts only the so-called "shadow feature" [36, 37] which is an enhancement in the incoherent background of the spectral function due to antiferromagnetic fluctuations. However, it does not predict [38] the existence of "shadow bands" in the renormalized classical regime. These bands occur when the condition $\omega - \epsilon_{\mathbf{k}} - \Sigma_\sigma(\mathbf{k}, \omega) + \mu = 0$ is satisfied. FLEX also predicts no pseudogap in the spectral function $A(\mathbf{k}_{\rm F}, \omega)$ at half-filling [38]. By analyzing temperature and size dependence of the Monte Carlo data and comparing them with the theoretical calculations, we argue that the Monte Carlo data supports our conclusion that the precursors of antiferromagnetic bands and the pseudogap do appear in the renormalized classical regime. We believe that the reason for which the FLEX approximation fails to reproduce this effect is essentially the same reason for which it fails to reproduce Hubbard bands in the strong coupling limit. More specifically, the failure is due to an inconsistent treatment of vertex corrections in the self-energy ansatz. Contrary to the electron-phonon case, these vertex corrections have a strong tendency to cancel the effects of using dressed propagators in the expression for the self-energy.

Recently, there have been very exciting developments in photoemission studies of the High-$T_{\rm c}$ materials [39, 40] that show the opening of the pseudogap in single particle spectra above the superconducting phase transition. At present, there is an intense debate about the physical origin of this phenomena and, in particular, whether it is of magnetic or of pairing origin. From the theoretical point of view there are a lot of formal similarities in the description of antiferromagnetism in repulsive models and superconductivity in attractive models. In Section 5 we use this formal analogy to obtain a simple analytical expressions for the self-energy in the regime dominated by critical pairing fluctuations. We then point out on the similarities and differences in the spectral function in the case of magnetic and pairing pseudogaps.

Our approach has been described in simple physical terms in references [29, 30]. The plan of the present paper is as follows. After recalling the model and the notation, we present our theory in Section 3. There we point out which exact requirements of many-body theory are satisfied, and which are violated. Before Section 3, the reader is urged to read Appendix A that contains a summary of sum rules, conservation laws and other exact constraints. Although this discussion contains many original results, it is not in the main text since the more expert reader can refer to the appendix as need be. We also illustrate in this appendix how an inconsistent treatment of the self-energy and vertex corrections can lead to the violation of a number of sum rules and inhibit the appearance of the Hubbard bands, a subject also treated in Section 6. Section 4 compares the results of our approach and of other approaches to Monte Carlo simulations. We study in more details in Section 5 the renormalized classical regime at half-filling where, in two dimensions, Fermi liquid quasiparticles are destroyed and replaced by precursors of antiferromagnetic bands well before the $T = 0$ phase transition. We also consider in this section the analogous phenomenon of pairing pseudogap which can appear in two dimensions when the pairing fluctuations become critical. The following section (Sect. 6) explains other attempts to obtain precursors of antiferromagnetic bands and points out why approaches such as FLEX fail to see the effect. We conclude in Section 7 with a discussion

of the domain of validity of our approach and in Section 8 with a critical comparison with FLEX and pseudo-potential parquet approaches, listing the weaknesses and strengths of our approach compared with these. A more systematic description and critique of various many-body approaches, as well as proofs of some of our results, appear in appendices.

2. Model and Definitions

We first present the model and various definitions. The Hubbard model is given by the Hamiltonian

$$H = -\sum_{<ij>\sigma} t_{i,j}\left(c^{\dagger}_{i\sigma}c_{j\sigma} + c^{\dagger}_{j\sigma}c_{i\sigma}\right) + U\sum_{i} n_{i\uparrow}n_{i\downarrow}. \tag{1}$$

In this expression, the operator $c_{i\sigma}$ destroys an electron of spin σ at site i. Its adjoint $c^{\dagger}_{i\sigma}$ creates an electron and the number operator is defined by $n_{i\sigma} = c^{\dagger}_{i\sigma}c_{i\sigma}$. The symmetric hopping matrix $t_{i,j}$ determines the band structure, which here can be arbitrary. Double occupation of a site costs an energy U due to the screened Coulomb interaction. We work in units where $k_{\mathrm{B}} = 1$, $\hbar = 1$ and the lattice spacing is also unity, $a = 1$. As an example that occurs later, the dispersion relation in the d -dimensional nearest-neighbor model is given by

$$\epsilon_{\mathbf{k}} = -2t\sum_{i=1}^{d}\left(\cos k_i\right). \tag{2}$$

2.1. Single-Particle Propagators, Spectral Weight and Self-Energy. — We will use a "four"-vector notation $k \equiv (\mathbf{k}, ik_{\mathrm{n}})$ for momentum-frequency space, and $1 \equiv (\mathbf{r}_1, \tau_1)$ for position-imaginary time. For example, the definition of the single-particle Green's function can be written as

$$G_{\sigma}(1,2) \equiv -\left\langle T_{\tau}c_{1\sigma}(\tau_1)\, c^{\dagger}_{2\sigma}(\tau_2)\right\rangle \equiv -\left\langle T_{\tau}c_{\sigma}(1)\, c^{\dagger}_{\sigma}(2)\right\rangle \tag{3}$$

where the brackets $\langle\rangle$ represent a thermal average in the grand canonical ensemble, T_{τ} is the time-ordering operator, and τ is imaginary time. In zero external field and in the absence of the symmetry breaking $G_{\sigma}(1,2) = G_{\sigma}(1-2)$ and the Fourier-Matsubara transforms of the Green's function are

$$G_{\sigma}(k) = \sum_{\mathbf{r}_1} \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{r}_1}\int_0^{\beta}\mathrm{d}\tau\ \mathrm{e}^{ik_{\mathrm{n}}\tau_1}G_{\sigma}(\mathbf{r}_1,\tau_1) \equiv \int \mathrm{d}(1)\mathrm{e}^{-ik(1)}G_{\sigma}(1) \tag{4}$$

$$G_{\sigma}(1) = \frac{T}{N}\sum_{k}\mathrm{e}^{ik(1)}G_{\sigma}(k). \tag{5}$$

As usual, experimentally observable retarded quantities are obtained from the Matsubara ones by analytical continuation $ik_{\mathrm{n}} \to \omega + i\eta$. In particular, the single-particle spectral weight $A(\mathbf{k},\omega)$ is related to the single-particle propagator by

$$G_{\sigma}(\mathbf{k}, ik_{\mathrm{n}}) = \int \frac{\mathrm{d}\omega}{2\pi}\frac{A_{\sigma}(\mathbf{k},\omega)}{ik_{\mathrm{n}} - \omega} \tag{6}$$

$$A_{\sigma}(\mathbf{k},\omega) = -2\mathrm{Im}\, G^{\mathrm{R}}_{\sigma}(\mathbf{k},\omega). \tag{7}$$

The self-energy obeys Dyson's equation, leading to

$$G_{\sigma}(\mathbf{k}, ik_{\mathrm{n}}) = \frac{1}{ik_{\mathrm{n}} - (\epsilon_{\mathbf{k}} - \mu) - \Sigma_{\sigma}(\mathbf{k}, ik_{\mathrm{n}})}\,. \tag{8}$$

It is convenient to use the following notation for real and imaginary parts of the analytically continued retarded self-energy

$$\Sigma^{\mathrm{R}}_{\sigma}(\mathbf{k}, ik_{\mathrm{n}} \to \omega + i\eta) = \Sigma'_{\sigma}(\mathbf{k}, \omega) + i\Sigma''_{\sigma}(\mathbf{k}, \omega). \tag{9}$$

Causality and positivity of the spectral weight imply that

$$\Sigma''_{\sigma}(\mathbf{k}, \omega) < 0. \tag{10}$$

Finally, let us point out that for nearest-neighbor hopping, the Hamiltonian is particle-hole symmetric at half-filling, ($c_{\mathbf{k}\sigma} \to c^{\dagger}_{\mathbf{k}+\mathbf{Q}\sigma}$; $c^{\dagger}_{\mathbf{k}\sigma} \to c_{\mathbf{k}+\mathbf{Q}\sigma}$) with $\mathbf{Q} = (\pi, \pi)$, implying that $\mu = U/2$ and that,

$$G_{\sigma}(\mathbf{k}, \tau) = -G_{\sigma}(\mathbf{k} + \mathbf{Q}, -\tau) \tag{11}$$

$$\left[\Sigma(\mathbf{k}, ik_{\mathrm{n}}) - \frac{U}{2}\right] = -\left[\Sigma(\mathbf{k} + \mathbf{Q}, -ik_{\mathrm{n}}) - \frac{U}{2}\right]. \tag{12}$$

2.2. Spin and Charge Correlation Functions. — We shall be primarily concerned with spin and charge fluctuations, which are the most important collective modes in the repulsive Hubbard model. Let the charge and z components of the spin operators at site i be given respectively by

$$\rho_i(\tau) \equiv n_{i\uparrow}(\tau) + n_{i\downarrow}(\tau) \tag{13}$$

$$S^z_i \equiv n_{i\uparrow}(\tau) - n_{i\downarrow}(\tau). \tag{14}$$

The time evolution here is again that of the Heisenberg representation in imaginary time.

The charge and spin susceptibilities in imaginary time are the responses to perturbations applied in imaginary-time. For example, the linear response of the spin to an external field that couples linearly to the z component

$$\mathrm{e}^{-\beta H} \to \mathrm{e}^{-\beta H} T_{\tau} \mathrm{e}^{\int \mathrm{d}\tau S^z_i(\tau')\phi^S_i(\tau')} \tag{15}$$

is given by

$$\chi_{\mathrm{sp}}(\mathbf{r}_i - \mathbf{r}_j, \tau_i - \tau_j) = \frac{\delta \langle S_j(\tau_j)\rangle}{\delta \phi^S_i(\tau_i)} = \left\langle T_{\tau} S^z_i(\tau_i) S^z_j(\tau_j)\right\rangle. \tag{16}$$

In an analogous manner, for charge we have

$$\chi_{\mathrm{ch}}(\mathbf{r}_i - \mathbf{r}_j, \tau_i - \tau_j) = \frac{\delta \langle \rho_j(\tau_j)\rangle}{\delta \phi^{\rho}_i(\tau_i)} = \langle T_{\tau} \rho_i(\tau_i)\rho_j(\tau_j)\rangle - n^2. \tag{17}$$

Here $n \equiv \langle \rho_i \rangle$ is the filling so that the disconnected piece is denoted n^2. It is well known that when analytically continued, these susceptibilities give physical retarded and advanced response functions. In fact, the above two expressions are the imaginary-time version of the fluctuation-dissipation theorem.

The expansion of the above functions in Matsubara frequencies uses even frequencies. Defining the subscript ch, sp to mean either charge or spin, we have

$$\chi_{\mathrm{ch,sp}}(\mathbf{q}, iq_{\mathrm{n}}) = \int \frac{\mathrm{d}\omega'}{\pi} \frac{\chi''_{\mathrm{ch,sp}}(\mathbf{q}, \omega')}{\omega' - iq_{\mathrm{n}}} \tag{18}$$

$$\chi''_{\mathrm{ch}}(\mathbf{q}, t) = \frac{1}{2}\langle [\rho_{\mathbf{q}}(t), \rho_{-\mathbf{q}}(0)]\rangle; \quad \chi''_{\mathrm{sp}}(\mathbf{q}, t) = \frac{1}{2}\left\langle \left[S^z_{\mathbf{q}}(t), S^z_{-\mathbf{q}}(0)\right]\right\rangle. \tag{19}$$

The fact that $\chi''_{\mathrm{ch,sp}}(\mathbf{q},\omega')$ is real and odd in frequency in turn means that $\chi_{\mathrm{ch,sp}}(\mathbf{q},iq_{\mathrm{n}})$ is real

$$\chi_{\mathrm{ch,sp}}(\mathbf{q},iq_{\mathrm{n}}) = \int \frac{\mathrm{d}\omega'}{\pi} \frac{\omega'\chi''_{\mathrm{ch,sp}}(\mathbf{q},\omega')}{(\omega')^2 + (q_{\mathrm{n}})^2} \tag{20}$$

a convenient feature for numerical calculations. The high-frequency expansion has $1/q_{\mathrm{n}}^2$ as a leading term so that there is no discontinuity in $\chi_{\mathrm{ch,sp}}(\mathbf{q},\tau)$ as $\tau \to 0$, contrary to the single-particle case.

3. Formal Derivation

To understand how to satisfy as well as possible the requirements imposed on many-body theory by exact results, such as those in Appendix A, it is necessary to start from a general non-perturbative formulation of the many-body problem. We thus first present a general approach to many-body theory that is set in the framework introduced by Martin and Schwinger [42], Luttinger and Ward [28] and Kadanoff and Baym [26, 27]. This allows one to see clearly the structure of the general theory expressed in terms of the one-particle irreducible self-energy and of the particle-hole irreducible vertices. These quantities represent projected propagators and there is a great advantage in doing approximations for these quantities rather than directly on propagators.

Our own approximation to the Hubbard model is then described in the subsection that follows the formalism. In our approach, the irreducible quantities are determined from various consistency requirements. The reader who is interested primarily in the results rather than in formal aspects of the theory can skip the next subsection and refer back later as needed.

3.1. General Formalism. — Following Kadanoff and Baym [27], we introduce the generating function for the Green's function

$$\ln Z\,[\phi] = \ln \left\langle T_\tau \mathrm{e}^{-c^\dagger_{\overline{\sigma}}(\overline{1})c_{\overline{\sigma}}(\overline{2})\phi_{\overline{\sigma}}(\overline{1},\overline{2})} \right\rangle \tag{21}$$

where, as above, a bar over a number means summation over position and imaginary time and, similarly, a bar over a spin index means a sum over that spin index. The quantity Z is a functional of ϕ_σ, the position and imaginary-time dependent field. Z reduces to the usual partition function when the field ϕ_σ vanishes. The one-particle Green's function in the presence of this external field is given by

$$G_\sigma(1,2;[\phi]) = -\frac{\delta \ln Z\,[\phi]}{\delta\phi_\sigma(2,1)} \tag{22}$$

and, as shown by Kadanoff and Baym, the inverse Green's function is related to the self-energy through

$$G^{-1} = G_0^{-1} - \phi - \Sigma. \tag{23}$$

The self-energy in this expression is a functional of ϕ.

Performing a Legendre transform on the generating functional $\ln Z\,[\phi]$ in equation (21) with the help of the last two equations, one can find a functional $\Phi\,[G]$ of G that acts as a generating function for the self-energy

$$\Sigma_\sigma(1,2;[G]) = \frac{\delta\Phi\,[G]}{\delta G_\sigma(2,1)}\,. \tag{24}$$

The quantity $\Phi\,[G]$ is the Luttinger-Ward functional [28]. Formally, it is expressed as the sum of all connected skeleton diagrams, with appropriate counting factors. Conserving approximations

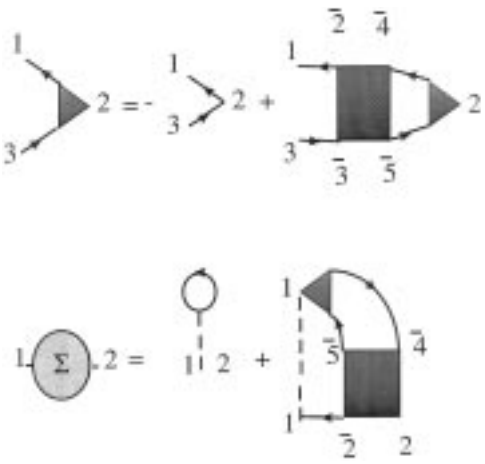

Fig. 1. — The first line is a diagrammatic representation of the Bethe-Salpeter equation (26) for the three point susceptibility and the second line is the corresponding equation (27) for the self-energy. In the Hubbard model, the Fock contribution is absent, but in general it should be there. Solid lines are Green's functions and dashed lines represent the contact interaction U. The triangle is the three point vertex, while the three-point susceptibility $\chi(1,3;2)$ is the triangle along with the attached Green's function. The usual two-point susceptibility is obtained by identifying points 1 and 3 in the Bethe-Salpeter equation. The rectangular box is the irreducible four-point vertex in the selected particle-hole channel.

start from a subset of all possible connected diagrams for $\Phi\,[G]$ to generate both the self-energy and the irreducible vertices entering the integral equation obeyed by response functions. These response functions are then guaranteed to satisfy the conservation laws. They obey integral equations containing as irreducible vertices

$$\Gamma^{\rm ir}_{\sigma\sigma'}(1,2;3,4) \equiv \frac{\delta\Sigma_\sigma(1,2;[G])}{\delta G_{\sigma'}(3,4)} = \frac{\delta^2\Phi\,[G]}{\delta G_\sigma(2,1)\delta G_{\sigma'}(3,4)} = \Gamma^{\rm ir}_{\sigma'\sigma}(4,3;2,1). \tag{25}$$

A complete and exact picture of one- and two-particle properties is obtained then as follows. First, the generalized susceptibilities $\chi_{\sigma\sigma'}(1,3;2) \equiv -\delta G_\sigma(1,3)/\delta\phi_{\sigma'}\,(2^+,2)$ are calculated by taking the functional derivative of GG^{-1} and using the Dyson equation (23) to compute $\delta G^{-1}/\delta\phi$. One obtains [27]

$$\chi_{\sigma\sigma'}(1,3;2) = -G_\sigma(1,2)\delta_{\sigma,\sigma'}G_\sigma(2,3) + G_\sigma(1,\overline{2})\Gamma^{\rm ir}_{\sigma\overline{\sigma}}(\overline{2},\overline{3};\overline{4},\overline{5})\chi_{\overline{\sigma}\sigma'}(\overline{4},\overline{5};2)G_\sigma(\overline{3},3) \tag{26}$$

where one recognizes the Bethe-Salpeter equation for the three-point susceptibility in the particle-hole channel. The second equation that we need is automatically satisfied in an exact theory. It relates the self-energy to the response function just discussed through the equation

$$\Sigma_\sigma(1,2) = Un_{-\sigma}\delta(1-2) + UG_\sigma(1,\overline{2})\Gamma^{\rm ir}_{\sigma\sigma'}(\overline{2},2;\overline{4},\overline{5})\chi_{\sigma'-\sigma}(\overline{4},\overline{5};1) \tag{27}$$

which is proven in Appendix B.

The diagrammatic representation of these two equations (26, 27) appearing in Figure 1 may make them look more familiar. Despite this diagrammatic representation, we stress that this is only for illustrative purposes. The rest of our discussion will not be diagrammatic.

Because of the spin-rotational symmetry the above equations (26, 27) can be decoupled into symmetric (charge) and antisymmetric (spin) parts, by introducing spin and charge irreducible vertices and generalized susceptibilities:

$$\Gamma_{\rm ch} \equiv \Gamma^{\rm ir}_{\uparrow\downarrow} + \Gamma^{\rm ir}_{\uparrow\uparrow} \quad ; \quad \Gamma_{\rm sp} \equiv \Gamma^{\rm ir}_{\uparrow\downarrow} - \Gamma^{\rm ir}_{\uparrow\uparrow} \tag{28}$$

$$\chi_{\rm ch} \equiv 2(\chi_{\uparrow\downarrow} + \chi_{\uparrow\uparrow}) \quad ; \quad \chi_{\rm sp} \equiv 2(\chi_{\uparrow\uparrow} - \chi_{\uparrow\downarrow}). \tag{29}$$

The usual two-point susceptibilities are obtained from the generalized ones as $\chi_{\rm sp,ch}(1,2) = \chi_{\rm sp,ch}(1,1^+;2)$. The equation (26) for the generalized spin susceptibility leads to

$$\chi_{\rm sp}(1,3;2) = -2G(1,2)G(2,3) - \Gamma_{\rm sp}(\overline{2},\overline{3};\overline{4},\overline{5})G(1,\overline{2})G(\overline{3},3)\chi_{\rm sp}(\overline{4},\overline{5};2) \tag{30}$$

and similarly for charge, but with the plus sign in front of the second term.

Finally, one can write the exact equation (27) for the self-energy in terms of the response functions as

$$\Sigma_\sigma(1,2) = Un_{-\sigma}\delta(1-2) + \frac{U}{4}[\Gamma_{\rm sp}(\overline{2},2;\overline{4},\overline{5})\chi_{\rm sp}(\overline{4},\overline{5};1) + \Gamma_{\rm ch}(\overline{2},2;\overline{4},\overline{5})\chi_{\rm ch}(\overline{4},\overline{5};1)]G_\sigma(1,\overline{2}). \tag{31}$$

Our two key equations are thus those for the three-point susceptibilities, equation (30), and for the self-energy, equation (31). It is clear from the derivation in Appendix B that these equations are intimately related.

3.2. Approximations through Local Irreducible Vertices

3.2.1. Conserving Approximation for the Collective Modes. — In formulating approximation methods for the many-body problem, it is preferable to confine our ignorance to high-order correlation functions whose detailed momentum and frequency dependence is not singular and whose influence on the low energy Physics comes only through averages over momentum and frequency. We do this here by parameterizing the Luttinger-Ward functional by two constants $\Gamma^{\rm ir}_{\uparrow\downarrow}$ and $\Gamma^{\rm ir}_{\uparrow\uparrow}$. They play the role of particle-hole irreducible vertices that are eventually determined by enforcing sum rules and a self-consistency requirement at the two-particle level. In the present context, this functional can be also considered as the interacting part of a Landau functional. The ansatz is

$$\Phi\left[G\right] = \frac{1}{2}G_{\overline{\sigma}}\left(\overline{1},\overline{1}^+\right)\Gamma^{\rm ir}_{\overline{\sigma}\overline{\sigma}}G_{\overline{\sigma}}\left(\overline{1},\overline{1}^+\right) + \frac{1}{2}G_{\overline{\sigma}}\left(\overline{1},\overline{1}^+\right)\Gamma^{\rm ir}_{\overline{\sigma}-\overline{\sigma}}G_{-\overline{\sigma}}\left(\overline{1},\overline{1}^+\right). \tag{32}$$

As in every conserving approximation, the self-energy and irreducible vertices are obtained from functional derivatives as in equations (24, 25) and then the collective modes are computed from the Bethe-Salpeter equation (30). The above Luttinger-Ward functional gives a momentum and frequency independent self-energy [43], that can be absorbed in a chemical potential shift. From the Luttinger-Ward functional, one also obtains two local particle-hole irreducible vertices $\Gamma^{\rm ir}_{\sigma\sigma}$ and $\Gamma^{\rm ir}_{\sigma-\sigma}$

$$\Gamma^{\rm ir}_{\sigma\overline{\sigma}}(2,3;4,5) \equiv \frac{\delta\Sigma_\sigma(2,3)}{\delta G_{\sigma'}(4,5)} = \delta(2-5)\delta(3-4)\delta(4^+-5)\Gamma^{\rm ir}_{\sigma\sigma'}\,. \tag{33}$$

We denote the corresponding local spin and charge irreducible vertices as

$$U_{\rm sp} \equiv \Gamma^{\rm ir}_{\sigma-\sigma} - \Gamma^{\rm ir}_{\sigma\sigma}; \quad U_{\rm ch} \equiv \Gamma^{\rm ir}_{\sigma-\sigma} + \Gamma^{\rm ir}_{\sigma\sigma}\,. \tag{34}$$

Notice now that there are only two equal-time, equal-point (*i.e.* local) two-particle correlation functions in this problem, namely $\langle n_\uparrow n_\downarrow\rangle$ and $\langle n_\uparrow^2\rangle = \langle n_\downarrow^2\rangle = \langle n_\downarrow\rangle = n/2$. The last one is

completely determined by the Pauli principle and by the known filling factor, while $U\langle n_\uparrow n_\downarrow\rangle$ is the expectation value of the interaction term in the Hamiltonian. Only one of these two correlators, namely $U\langle n_\uparrow n_\downarrow\rangle$, is unknown. Assume for the moment that it is known. Then, we can use the two sum rules (Eqs. (A.15, A.14)) that follow from the fluctuation-dissipation theorem and from the Pauli principle to determine the two trial irreducible vertices from the known value of this one key local correlation functions. In the present notation, these two sum rules are

$$\chi_{\rm ch}\left(1,1^+\right) = \frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\rm n}}\chi_{\rm ch}(\mathbf{q},iq_{\rm n}) = \langle n_\uparrow\rangle + \langle n_\downarrow\rangle + 2\langle n_\uparrow n_\downarrow\rangle - n^2 \tag{35}$$

$$\chi_{\rm sp}\left(1,1^+\right) = \frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\rm n}}\chi_{\rm sp}(\mathbf{q},iq_{\rm n}) = \langle n_\uparrow\rangle + \langle n_\downarrow\rangle - 2\langle n_\uparrow n_\downarrow\rangle \tag{36}$$

and since the spin and charge susceptibilities entering these equations are obtained by solving the Bethe-Salpeter equation (30) with the constant irreducible vertices equations (33, 34) we have one equation for each of the irreducible vertices

$$n + 2\langle n_\uparrow n_\downarrow\rangle - n^2 = \frac{T}{N}\sum_{q}\frac{\chi_0(q)}{1+\frac{1}{2}U_{\rm ch}\chi_0(q)}, \tag{37}$$

$$n - 2\langle n_\uparrow n_\downarrow\rangle = \frac{T}{N}\sum_{\widetilde{q}}\frac{\chi_0(q)}{1-\frac{1}{2}U_{\rm sp}\chi_0(q)}\,. \tag{38}$$

We used our usual short-hand notation for wave vector and Matsubara frequency $q = (\mathbf{q}, iq_{\rm n})$. Since the self-energy corresponding to our trial Luttinger-Ward functional is constant, the irreducible susceptibilities take their non-interacting value $\chi_0(q)$.

The local Pauli principle $\langle n_\downarrow^2\rangle = \langle n_\downarrow\rangle$ leads to the following important sum-rule

$$\frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\rm n}}\left[\chi_{\rm sp}(\mathbf{q},iq_{\rm n}) + \chi_{\rm ch}(\mathbf{q},iq_{\rm n})\right] = 2n - n^2, \tag{39}$$

which can be obtained by adding equations (38, 37). This sum-rule implies that effective interactions for spin $U_{\rm sp}$ and charge $U_{\rm ch}$ channels must be different from one another and hence that ordinary RPA is inconsistent with the Pauli principle (for details see Appendix A.3).

Equations (37, 38) determine $U_{\rm sp}$ and $U_{\rm ch}$ as a function of double occupancy $\langle n_\uparrow n_\downarrow\rangle$. Since double occupancy is a local quantity it depends little on the size of the system. It could be obtained reliably from a number of approaches, such as for example Monte Carlo simulations. However, there is a way to obtain double-occupancy self-consistently [29] without input from outside of the present theory. It suffices to add to the above set of equations the relation

$$U_{\rm sp} = g_{\uparrow\downarrow}(0)\,U; \qquad g_{\uparrow\downarrow}(0) \equiv \frac{\langle n_\uparrow n_\downarrow\rangle}{\langle n_\downarrow\rangle\langle n_\uparrow\rangle}\,. \tag{40}$$

Equations (38, 40) then define a set of self-consistent equations for $U_{\rm sp}$ that involve only two-particle quantities. This ansatz is motivated by a similar approximation suggested by Singwi *et al.* [31] in the electron gas, which proved to be quite successful in that case. On a lattice we will use it for $n \leq 1$. The case $n > 1$ can be mapped on the latter case using particle-hole transformation. In the context of the Hubbard model with on-site repulsion, the physical meaning of equation (40) is that the effective interaction in the most singular spin

channel, is reduced by the probability of having two electrons with opposite spins on the same site. Consequently, the ansatz reproduces the Kanamori-Brueckner screening that inhibits ferromagnetism in the weak to intermediate coupling regime (see also below). We want to stress, however, that this ansatz is not a rigorous result like sum rules described above. The plausible derivation of this ansatz can be found in references [29, 31] as well as, in the present notation, in Appendix C.

We have called this approach Two-Particle Self-Consistent to contrast it with other conserving approximations like Hartree-Fock or Fluctuation Exchange Approximation (FLEX) [24] that are self-consistent at the one-particle level, but not at the two-particle level. This approach [29] to the calculation of spin and charge fluctuations satisfies the Pauli principle $\langle n_\sigma^2 \rangle = \langle n_\sigma \rangle = n/2$ by construction, and it also satisfies the Mermin-Wagner theorem in two dimensions.

To demonstrate that this theorem is satisfied, it suffices to show that $\langle n_\uparrow n_\downarrow \rangle = g_{\uparrow\downarrow}(0) \langle n_\uparrow \rangle \langle n_\downarrow \rangle$ does not grow indefinitely. (This guarantees that the constant $\widetilde{C}$ appearing in Eq. (A.21) is finite.) To see how this occurs, write the self-consistency condition (Eq. (38)) in the form

$$n - 2\langle n_\uparrow n_\downarrow \rangle = \frac{T}{N} \sum_{\widetilde{q}} \frac{\chi_0(q)}{1 - \frac{1}{2} U \frac{\langle n_\uparrow n_\downarrow \rangle}{\langle n_\uparrow \rangle \langle n_\downarrow \rangle} \chi_0(q)} \,. \tag{41}$$

Consider increasing $\langle n_\uparrow n_\downarrow \rangle$ on the right-hand side of this equation. This leads to a decrease of the same quantity on the left-hand side. There is thus negative feedback in this equation that will make the self-consistent solution finite. A more direct proof by contradiction has been given in reference [29]: suppose that there is a phase transition, in other words suppose that $\langle n_\uparrow \rangle \langle n_\downarrow \rangle = \frac{1}{2} U \langle n_\uparrow n_\downarrow \rangle \chi_0(q)$. Then the zero-Matsubara frequency contribution to the right-hand side of equation (41) becomes infinite and positive in two dimensions as one can see from phase-space arguments (See Eq. (A.21)). This implies that $\langle n_\uparrow n_\downarrow \rangle$ on the left-hand side must become negative and infinite, but that contradicts the starting hypothesis since $\langle n_\uparrow \rangle \langle n_\downarrow \rangle = \frac{1}{2} U \langle n_\uparrow n_\downarrow \rangle \chi_0(q)$ means that $\langle n_\uparrow n_\downarrow \rangle$ is positive.

Although there is no finite-temperature phase transition, our theory shows that sufficiently close to half-filling (see Sect. 4.3) there is a crossover temperature T_{X} below which the system enters the so-called renormalized classical regime, where antiferromagnetic correlations grow exponentially. This will be discussed in detail in Section 5.1.1.

Kanamori-Brueckner screening is also included as we already mentioned above. To see how the screening occurs, consider a case away from half-filling, where one is far from a phase transition. In this case, the denominator in the self-consistency condition can be expanded to linear order in U and one obtains

$$g_{\uparrow\downarrow}(0) = \frac{\langle n_\uparrow n_\downarrow \rangle}{\langle n_\uparrow \rangle \langle n_\downarrow \rangle} = \frac{1}{1 + \Lambda U} \tag{42}$$

where

$$\Lambda = \frac{2}{n^2} \frac{T}{N} \sum_q \chi_0(q)^2. \tag{43}$$

Clearly, quantum fluctuations contribute to the sum appearing above and hence to the renormalization of $U_{\mathrm{sp}} = g_{\uparrow\downarrow}(0)\, U$. The value of Λ is found to be near 0.2 as in explicit numerical calculations of the maximally crossed Kanamori-Brueckner diagrams [44]. At large U, the value of $U_{\mathrm{sp}} = g_{\uparrow\downarrow}(0)\, U \sim 1/\Lambda$ saturates to a value of the order of the inverse bandwidth which corresponds to the energy cost for creating a node in the two-body wave function, in agreement with the Physics described by Kanamori [2].

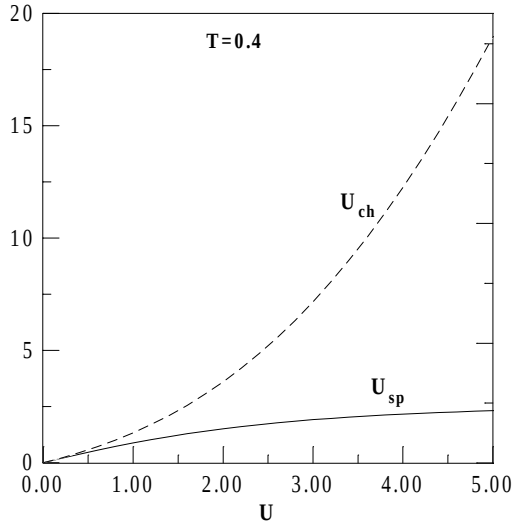

Fig. 2.

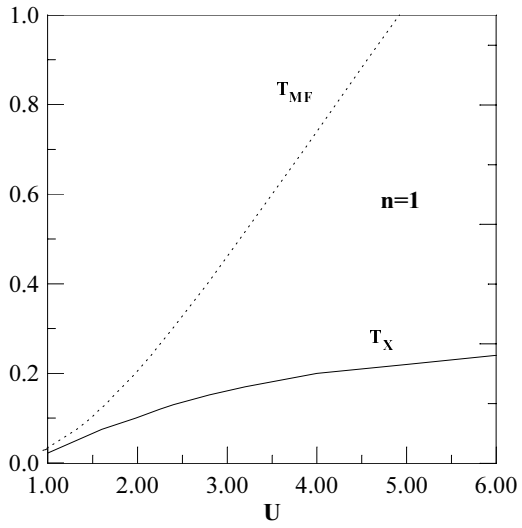

Fig. 3.

Fig. 2. — Dependence on U of the charge and spin effective interactions (irreducible vertices). The temperature is chosen so that for all U, it is above the crossover temperature. In this case, temperature dependence is not significant. The filling is $n = 1$.

Fig. 3. — Crossover temperature at half-filling as function of U compared with the mean-field transition temperature.

To illustrate the dependence of $U_{\rm sp}, U_{\rm ch}$ on bare U we give in Figure 2 a plot of these quantities at half-filling where the correlation effects are strongest. The temperature for this plot is chosen to be above the crossover temperature $T_{\rm X}$ to the renormalized classical regime, in which case the dependence of $U_{\rm sp}$ and $U_{\rm ch}$ on temperature is not significant. As one can see, $U_{\rm sp}$ rapidly saturates to a fraction of the bandwidth, while $U_{\rm ch}$ rapidly increases with U, reflecting the tendency to the Mott transition. We have also shown previously in Figure 2 of reference [29] that $U_{\rm sp}$ depends only weakly on filling. Since $U_{\rm sp}$ saturates as a function of U due to Kanamori-Brueckner screening, the crossover temperature $T_{\rm X}$ also saturates as a function of U. This is illustrated in Figure 3 along with the mean-field transition temperature that, by contrast, increases rapidly with U.

Quantitative agreement with Monte Carlo simulations on the nearest-neighbor [29] and next-nearest-neighbor models [34] is obtained [29] for all fillings and temperatures in the weak to intermediate coupling regime $U < 8t$. This is discussed further below in Section 4. We have also shown that the above approach reproduces both quantum-critical and renormalized-classical regimes in two dimensions to leading order in the $1/n$ expansion (spherical model) [33].

As judged by comparisons with Monte Carlo simulations [45], the particle-particle channel in the repulsive two-dimensional Hubbard model is relatively well described by more standard perturbative approaches. Although our approach can be extended to this channel as well, we do not consider it directly in this paper. It manifests itself only indirectly through the renormalization of $U_{\rm sp}$ and $U_{\rm ch}$ that it produces.

3.2.2. Single-Particle Properties. — As in any implementation of conserving approximations, the initial guess for the self-energy, $\Sigma^{(0)}$, obtained from the trial Luttinger-Ward functional

is inconsistent with the exact self-energy formula (Eq. (31)). The latter formula takes into account the feedback of the spin and charge collective modes actually calculated from the conserving approximation. In our approach, we use this self-energy formula (Eq. (31)) in an iterative manner to improve on our initial guess of the self-energy. The resulting formula for an improved self-energy $\Sigma^{(1)}$ has the simple physical interpretation of paramagnon theories [46].

As another way of Physically explaining this point of view, consider the following: the bosonic collective modes are weakly dependent on the precise form of the single-particle excitations, as long as they have a quasiparticle structure. In other words, zero-sound or paramagnons exist, whether the Bethe-Salpeter equation is solved with non-interacting particles or with quasiparticles. The details of the single-particle self-energy by contrast can be strongly influenced by scattering from collective modes because these bosonic modes are low-lying excitations. Hence, we first compute the two-particle propagators with Hartree-Fock single-particle Green's functions, and then we improve on the self-energy by including the effect of collective modes on single-particle properties. The fact that collective modes can be calculated first and self-energy afterwards is reminiscent of renormalization group approaches [8, 47], where collective modes are obtained at one-loop order while the non-trivial self-energy comes out only at two-loop order.

The derivation of the general self-energy formula (Eq. (31)) given in Appendix B shows that it basically comes from the definition of the self-energy and from the equation for the collective modes (Eq. (30)). This also stands out clearly from the diagrammatic representation in Figure 1. By construction, these two equations (Eqs. (30, 31)) satisfy the consistency requirement $\frac{1}{2}\mathrm{Tr}\ \Sigma G = U\langle n_\uparrow n_\downarrow\rangle$ (see Appendix B), which in momentum and frequency space can be written as

$$\lim_{\tau\to 0^-}\frac{T}{N}\sum_k \Sigma_\sigma(k)G_\sigma(k)\mathrm{e}^{-ik_\mathrm{n}\tau} = U\langle n_\uparrow n_\downarrow\rangle\cdot \tag{44}$$

The importance of the latter sum rule, or consistency requirement, for approximate theories should be clear from the appearance of the correlation function $\langle n_\uparrow n_\downarrow\rangle$ that played such an important role in determining the irreducible vertices and in obtaining the collective modes. Using the fluctuation dissipation theorem (Eqs. (36, 35)) this sum-rule can be written in form that explicitly shows the relation between the self-energy and the spin and charge susceptibilities

$$\frac{T}{N}\sum_k \left[\Sigma_\sigma(k) - Un_{-\sigma}\right]G_\sigma(k) = \frac{U}{4}\frac{T}{N}\sum_q \left[\chi_\mathrm{ch}(q) - \chi_\mathrm{sp}(q)\right]. \tag{45}$$

To keep as much as possible of this consistency, we use on the right-hand side of the self-energy expression (Eq. (31)) the same irreducible vertices and Green's functions as those that appear in the collective-mode calculation (Eq. (30)). Let us call $G^{(0)}$ the initial Green's function corresponding to the initial Luttinger-Ward self-energy $\Sigma^{(0)}$. Our new approximation for the self-energy $\Sigma^{(1)}$ then takes the form

$$\Sigma_\sigma^{(1)}(k) = Un_{-\sigma} + \frac{U}{4}\frac{T}{N}\sum_q \left[U_\mathrm{sp}\chi_\mathrm{sp}(q) + U_\mathrm{ch}\chi_\mathrm{ch}(q)\right]G_\sigma^{(0)}(k+q). \tag{46}$$

Note that $\Sigma_\sigma^{(1)}(k)$ satisfies particle-hole symmetry (Eq. (12)) where appropriate. This self-energy expression (Eq. (46)) is physically appealing since, as expected from general skeleton diagrams, one of the vertices is the bare one U, while the other vertex is dressed and given by U_sp or U_ch depending on the type of fluctuation being exchanged. It is because Migdal's theorem does not apply for this problem that U_sp and U_ch are different from the bare U at one of the vertices. U_sp and U_ch here take care of vertex corrections [48].

The use of the full $G_\sigma(k+q)$ instead of $G^0_\sigma(k+q)$ in the above expression (Eq. (46)) would be inconsistent with frequency-independent irreducible vertices. For the collective mode (Eq. (30)) this is well known to lead to the violation of the conservation laws as was discussed in detail in the previous subsection. Here we insist that the same is true in the calculation of the effect of electronic collective modes on the single-particle properties. Formally, this is suggested by the similarity between the equation for the susceptibility (Eq. (30)) and that for the self-energy (Eq. (31)) in terms of irreducible vertices. More importantly, two physical effects would be absent if one were to use full G and frequency independent irreducible vertices. First, upper and lower Hubbard bands would not appear because the U^2/ω high-frequency behavior in equation (68) that is necessary to obtain the Hubbard bands would set in too late, as we discuss in Sections 1.2 and 6.1. This result is also apparent from the fact that FLEX calculations in infinite dimension do not find upper and lower Hubbard bands [49] where the exact numerical solution does. The other physical effect that would be absent is precursors of antiferromagnetic bands, Section 5 and the pseudogap in $A(\mathbf{k}_\mathrm{F}, \omega)$, that would not appear for reasons discussed in Section 6. We also will see in Section 4 below that FLEX calculations of the single-particle Green's function, significantly disagree with Monte Carlo data, even away from half-filling, as was already shown in Figure 1 of reference [30].

The chemical potential for interacting electrons μ is found from the usual condition on particle number

$$n = \frac{T}{N}\sum_k G^{(1)}_\sigma(k)\exp(-ik_\mathrm{n}0^-) = \frac{T}{N}\sum_k \frac{\exp(-ik_\mathrm{n}0^-)}{i\omega_\mathrm{n} - \varepsilon_\mathbf{k} + \mu^{(1)} - \Sigma^{(1)}(\mathbf{k}, k_\mathrm{n})}\cdot \tag{47}$$

This chemical potential μ is, of course, different from μ_0 but the Luttinger sum rule $\sum\theta(-\varepsilon_\mathbf{k} + \mu - \Sigma^{(1)}) = n_\sigma$ is satisfied to a high accuracy (about few percent) for all fillings and temperatures $T_\mathrm{X} \leq T \ll W$. As usual this occurs because the change in $\mu^{(1)} - \mu_0$ is compensated by the self-energy shift on the Fermi surface $\Sigma^{(1)}(\mathbf{k}_\mathrm{F}, 0)$. For $T < T_\mathrm{X}$ there is some deviation from the Luttinger sum rule which is due to the appearance of the precursors of the antiferromagnetic bands below T_X (Sect. 5) which develop into true SDW bands at $T = 0$.

It is important to realize that $G^{(0)}$ on the right hand side of the equation for the self-energy Σ cannot be calculated as $G^{(0)} = 1/(\omega - \varepsilon_\mathbf{k} + \mu^{(1)})$, because otherwise it would not reduce to zero-temperature perturbation theory when it is appropriate. As was pointed out by Luttinger, (see also Sect. A.4) the "non-interacting" Green's function used in the calculation for Σ should be calculated as $G^{(0)} = 1/(\omega - \varepsilon_\mathbf{k} - \Sigma^{(\mathrm{n})}(\mathbf{k}_\mathrm{F}, 0) + \mu^{(\mathrm{n})})$, where $\mu^{(\mathrm{n})}$ is calculated on the same level of accuracy as $\Sigma^{(\mathrm{n})}(\mathbf{k}_\mathrm{F}, 0)$, *i.e.* from equation (47) with $\Sigma^{(\mathrm{n})}(\mathbf{k}, ik_\mathrm{n})$. In our calculation below, we approximate $\mu^{(1)} - \Sigma^{(1)}(\mathbf{k}_\mathrm{F}, 0)$ by μ_0 because for the coupling strength and temperatures considered in this paper ($U \leq W/2$, $T_\mathrm{X} \leq T \ll W$) the Luttinger theorem is satisfied to high accuracy and the change of the Fermi surface shape is insignificant. In addition, at half-filling the condition $\mu - \Sigma(\mathbf{k}_\mathrm{F}, 0) = \mu_0$ is satisfied exactly at any U and T because of particle-hole symmetry. For somewhat larger coupling strengths and away from half-filling, one may try to improve the theory by using $G^{(0)} = 1/(\omega - \varepsilon_\mathbf{k} - \Sigma^{(1)}(\mathbf{k}_\mathrm{F}, 0) + \mu^{(1)})$, with $\Sigma^{(1)}$ and μ found self-consistently. However, the domain of validity of our approach is limited to the weak-to-intermediate coupling regime since the strong-coupling regime requires frequency-dependent pseudopotentials (see below).

Finally, let us note that, in the same spirit as Landau theory, the only vertices entering our theory are of the type $\Gamma_{\uparrow\downarrow}$ and $\Gamma_{\uparrow\uparrow}$, or, through equation (34), U_sp and U_ch. In other words, we look at the problem from the longitudinal spin and charge particle-hole channel. Consequently, in the contact pseudopotential approximation the exact equation for the self-energy (Eq. (31)) reduces to our expression (Eq. (46)) which does not have the factor 3 in

the front of the spin susceptibility. This is different from some paramagnon theories, in which such factor was introduced to take care of rotational invariance. However, we show in Appendix E.1 that these paramagnon theories are inconsistent with the sum-rule (Eq. (45)) which relates one and two-particle properties. In our approach, questions about transverse spin fluctuations are answered by invoking rotational invariance $\chi_{\rm sp}^{xx} = \chi_{\rm sp}^{yy} = \chi_{\rm sp}^{zz}$. In particular, one can write the expression for the self-energy (Eq. (46)) in an explicitly rotationally invariant form by replacing $\chi_{\rm sp}$ by $(1/3){\rm Tr}[\chi_{\rm sp}^{\nu\nu}]$. If calculations had been done in the transverse channel, it would have been crucial to do them while simultaneously enforcing the Pauli principle in that channel. In functional integration methods, it is well known that methods that enforce rotational invariance without enforcing the Pauli principle at the same time give unphysical answers, such as the wrong factor 2/3 in the RPA susceptibility [23] $\chi_{\rm sp} = \chi^0/(1-(2/3)U\chi^0)$ or wrong Hartree-Fock ground state [50].

3.2.3. Internal Accuracy Check. — The quantitative accuracy of the theory will be discussed in detail when we compare with Monte Carlo calculations in the next section. Here we show that we can use the consistency requirement between one- and two-particle properties (Eq. (44)) to gauge the accuracy of the theory from within the theory itself.

The important advantage of the expression for the self-energy $\Sigma_\sigma^{(1)}(k)$ given by equation (46) is that, as shown in Appendix (B), it satisfies the consistency requirement between one- and two-particle properties (Eq. (44)), in the following sense

$$\lim_{\tau\to 0^-} \frac{T}{N}\sum_k \Sigma_\sigma^{(1)}(k) G_\sigma^{(0)}(k) {\rm e}^{-ik_{\rm n}\tau} = U\,\langle n_\uparrow n_\downarrow\rangle\,. \tag{48}$$

Let $G_\sigma^{(1)}$ be defined by $[G_\sigma^{(1)}]^{-1} \equiv G_0^{-1} - \Sigma^{(1)}$. We can use the fact that in an exact theory we should have ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(1)}]$ in the above expression instead of ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(0)}]$ to check the accuracy of the theory. It suffices to compute by how much ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(0)}]$ differs from ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(1)}]$. In the parameter range $U < 4t$ and n, T arbitrary but not too deep in the, soon to be described, renormalized-classical regime, we find that ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(0)}]$ differs from ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(1)}]$ by at most 15%. Another way to check the accuracy of our approach is to evaluate the right-hand side of the f-sum rule (Eqs. (A.22)) with $n_{\mathbf{k}\sigma} = G_\sigma^{(1)}(\mathbf{k}, 0^-)$ and to compare with the result that had been obtained with $f_{\mathbf{k},\sigma}$. Again we find the same 15% disagreement, at worse, in the same parameter range. As one can expect, this deviation is maximal at half-filling and becomes smaller away from it.

Equation (46) for the self-energy $\Sigma^{(1)}$ already gives good agreement with Monte Carlo data but the accuracy can be improved even further by using the general consistency condition (Eq. (44)) on ${\rm Tr}[\Sigma_\sigma^{(1)} G_\sigma^{(1)}]$ to improve on the approximation for vertex corrections. To do so we replace $U_{\rm sp}$ and $U_{\rm ch}$ on the right-hand side of equation (46) by $\alpha U_{\rm sp}$ and $\alpha U_{\rm ch}$ with α determined self-consistently in such a way that equation (48) is satisfied with $G_\sigma^{(0)}(k)$ replaced by $G_\sigma^{(1)}(k)$. For $U < 4$, we have $\alpha < 1.15$. The slight difference between the irreducible vertices entering the collective modes and the vertex corrections entering the self-energy formula can be understood from the fact that the replacement of irreducible vertices by constants is, in a way, justified by the mean-value theorem for integrals. Since the averages are not taken over the same variables, it is clear that the vertex corrections in the self-energy formula and irreducible vertices in the collective modes do not need to be strictly identical when they are approximated by constants.

Before we move on to comparisons with Monte Carlo simulations, we stress that $\Sigma^{(1)}$ given by equation (46) cannot be substituted back into the calculation of $\chi_{\rm sp,ch}$ by simply replacing

$\chi_0 = G_0 G_0$ with the dressed bubble $\tilde{\chi}_0 = GG$. Indeed, this would violate conservation of spin and charge and f-sum rule. In particular, the condition $\chi_{\rm sp,ch}(\mathbf{q} = 0, iq_{\rm n} \neq 0) = 0$ that follows from the Ward identity (A.28) would be violated as we see in equation (A.23). In the next order, one is forced to work with frequency-dependent irreducible vertices that offset the unphysical behavior of $\tilde{\chi}_0$ at non-zero frequencies.

4. Numerical Results and Comparisons with Monte Carlo Simulations

In this section, we present a few numerical results and comparisons with Monte Carlo simulations. We divide this section in two parts. In the first one we discuss data sufficiently far from half-filling, or at high enough temperature, where size effects are unimportant for systems sizes available in Monte Carlo simulations. In the second part, we discuss data at half-filling. There, size effects become important below the crossover temperature $T_{\rm X}$ where correlations start to grow exponentially (Sect. 5). All single-particle properties are calculated with our approximation (Eq. (46)) for the self-energy using the vertex renormalization α explained in the previous section. The results would differ at worse by 15% if we had used $\alpha = 1$.

4.1. Far from the Crossover Temperature $T_{\rm X}$

4.1.1. Two-Particle Properties. — We have shown previously in Figures 4a-d of reference [29] and in Figures 2-4 and Figure 6 of reference [34] that both spin and charge structure factor sufficiently away from the crossover temperature $T_{\rm X}$ are in quantitative agreement with Monte Carlo data for values of U as large as the bandwidth. On the qualitative level, the decrease in charge fluctuations as one approaches half-filling has been explained [29] as a consequence of the Pauli principle embodied in the calculation of the irreducible vertex $U_{\rm ch}$ [51].

Here we present in Figures 4 and 5 comparisons with a dynamical quantity, namely the spin susceptibility. Similar comparisons, but with a phenomenological value of $U_{\rm sp}$, have been done by Bulut *et al.* [52]. Figure 4 shows the staggered spin susceptibility as a function of Matsubara frequencies for $n = 0.87$, $T = 0.25$ and $U = 4$. The effect of interactions is already quite large for the zero-frequency susceptibility. It is enhanced by a factor of over 5 compared with the non-interacting value. Nevertheless, one can see that the theory and Monte Carlo simulations are in good agreement.

Figure 5 shows the temperature dependence of the zero-frequency staggered spin susceptibility for the same filling and interaction as in the previous figure. Symbols represent Monte Carlo simulations from references [53, 99], the solid line is for our theory while dotted and dashed lines are for two versions of FLEX. Surprisingly, the fully conserving FLEX theory, (dashed line) compares worse with Monte Carlo data than the non-conserving version of this theory that neglects the so-called Aslamasov-Larkin diagrams (dotted line). By contrast, our theory is in better agreement with the Monte Carlo data than FLEX for the staggered susceptibility $\chi_{\rm sp}\left(\mathbf{q} = (\pi, \pi), i\omega_{\rm n} = 0\right)$, and at the same time it agrees exactly with the conservation law that states that $\chi_{\rm sp,ch}(\mathbf{q} = 0, i\omega_{\rm n} \neq 0) = 0$.

Finally, Figure 6 shows the double occupancy $\langle n_\uparrow n_\downarrow \rangle$ as a function of filling for various values of U. The symbols again represent Monte Carlo data for $T = 1/6$, and the lines are the results of our theory. Everywhere the agreement is very good, except for $n = 1, U = 4$. In the latter case, the system is already below the crossover temperature $T_{\rm X}$ to the renormalized classical regime. As explained in Section 7, the appropriate procedure for calculating double occupancy in this case is to take for $\langle n_\uparrow n_\downarrow \rangle$ its value (dotted line) at $T_{\rm X}$ instead of using the ansatz equation (40). In any case, the difference is not large.

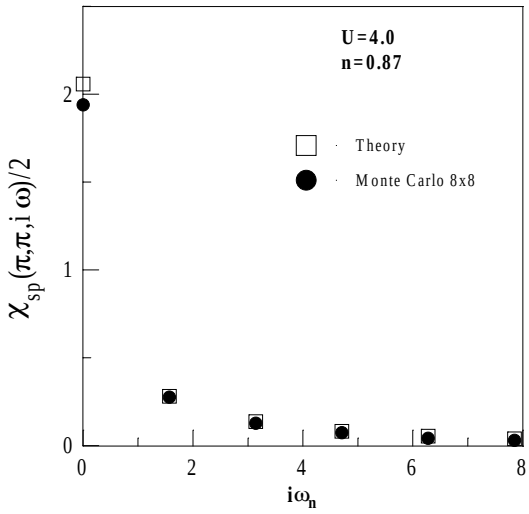

Fig. 4.

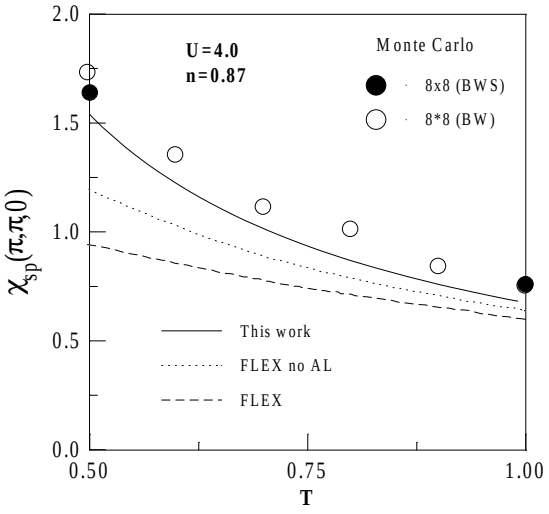

Fig. 5.

Fig. 4. — Comparisons between Monte Carlo simulations [99] and our theory for the spin susceptibility at $Q = (\pi, \pi)$ as a function of Matsubara frequency. The temperature is $T = 0.25$, and the system size 8×8. The factor $1/2$ on the vertical axis is due to the fact that the susceptibility in [99] is χ_{+-} a quantity that is by definition twice smaller then ours and that of [53].

Fig. 5. — Comparisons between the Monte Carlo simulations (BW) and FLEX calculations presented in Figure 19 of reference [53] and our theory for the spin susceptibility at $Q = (\pi, \pi)$ as a function of temperature at zero Matsubara frequency. The filled circles (BWS) are from reference [99].

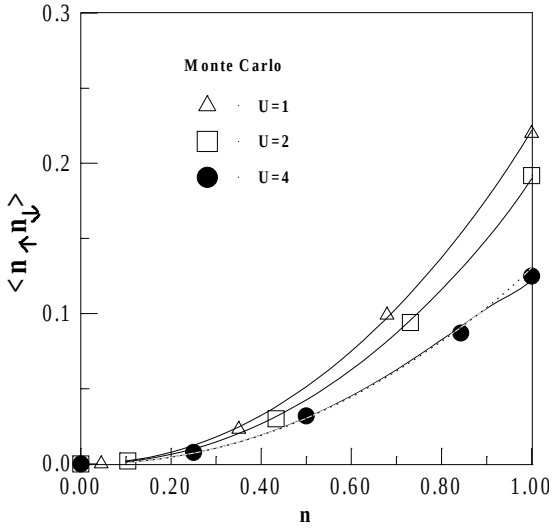

Fig. 6. — Comparisons between the Monte Carlo simulations of reference [57] and our theory (solid lines) for the filling dependence of the double occupancy. The results are are for $T = 1/6$ as a function of filling and for various values of U except for $U = 4$ where the dotted line shows the results of our theory at the crossover temperature $T = T_X$.

4.1.2. Single-Particle Properties. — Figure 1a of reference [30] shows $G(\mathbf{k}, \tau)$ for filling $n = 0.875$, temperature $T = 0.25$ and $U = 4$ for the wave vector on the 8×8 lattice which is closest to the Fermi surface, namely $(\pi, 0)$. Our theory is in agreement with Monte Carlo data and with the parquet approach [53] but in this regime second-order perturbation theory for the self-energy gives the same result. Surprisingly, FLEX is the only theory that disagrees significantly with Monte Carlo data. The good performance of perturbation theory (see also [54]) can be explained in part by compensation between the renormalized vertices and susceptibilities ($U_{\rm sp} < U$, $\chi_{\rm sp}(q) > \chi_0(q)$; $U_{\rm ch} > U$, $\chi_{\rm ch}(q) < \chi_0(q)$).

We have also calculated $\mathrm{Re}(\Sigma(ik_{\rm n})/ik_{\rm n})$ and compared with the Monte Carlo data in Figure 2a of reference [52] obtained at $n = 0.87$, $U = 4$, $\beta = 6$. Our approach agrees with Monte Carlo data for all frequencies, but again second-order perturbation theory gives similar results.

4.2. Close to Crossover Temperature $T_{\rm X}$ at Half-Filling

4.2.1. Two-Particle Properties. — The occurrence of the crossover temperature $T_{\rm X}$ at half-filling is perhaps best illustrated in the upper part of Figure 7 by the behavior of the static structure factor $S_{\rm sp}(\pi, \pi)$ for $U = 4$ as a function of temperature. When the correlation length becomes comparable to the size of the system used in Monte Carlo simulations [55], the static structure factor starts to increase rapidly, saturating to a value that increases with system size. The solid line is calculated from our theory for an infinite lattice. The Monte Carlo data follow our theoretical curve (solid line) until they saturate to a size-dependent value. The theory correctly describes the static structure factor not only above $T_{\rm X}$ but also as we enter the renormalized classical regime at $T_{\rm X}$. Analytical results for this regime are given in Section 5.1.1. Note that the RPA mean-field transition temperature for this value of U is more than three times larger than $T_{\rm X} \sim 0.2$. The size-dependence of Monte Carlo data for $S_{\rm sp}(\mathbf{q})$ at all other values of $\mathbf{q} \neq (\pi, \pi)$ available in simulations is negligible and our calculation for infinite system reproduces this data (not shown).

4.2.2. Single-Particle Properties. — Equal-time (frequency integrated) single-particle properties are much less sensitive to precursor effects than dynamical quantities as we now proceed to show. For example, $n(\mathbf{k}) = G(\mathbf{k}, 0^-)$ is a sum of $G(\mathbf{k}, ik_{\rm n})$ over all Matsubara frequencies. We have verified (figure not shown) that $\frac{1}{N}\sum_{\mathbf{k}\sigma} n_{\mathbf{k}\sigma}\partial^2\epsilon_{\mathbf{k}}/\partial k_x^2$ obtained from Monte Carlo simulations [56] is given quite accurately by either second-order perturbation theory or by our theory. This has very important consequences since, for this quantity, the non-interacting value differs from second-order perturbation theory by at most 15%. This means that the numerical value of the right-hand side of the f sum-rule (Eq. (A.22)) is quite close to that obtained from the left-hand side using our expression for the spin and charge susceptibility.

One can also look in more details at $n(\mathbf{k})$ itself instead of focusing on a sum rule. Figure 8 shows a comparison of our theory and of second order perturbation theory with Monte Carlo data for $n(\mathbf{k})$ obtained for a set of lattice sizes from 6×6 to 16×16 at $n = 1$, $T = 1/6$, $U = 4$. Size effects appear unimportant for this quantity at this temperature. These Monte Carlo data have been used in the past [57] to extract a gap by comparison with mean field SDW theory. Our theory for the same set of lattice sizes is in excellent agreement with Monte Carlo data and predicts a pseudogap at this temperature, as we will discuss below. However, for available values of $\mathbf{k}$ on finite lattices, second order perturbation theory is also in reasonable agreement with Monte Carlo data for $n(\mathbf{k})$. Since second order perturbation theory does not predict a pseudogap, this means that $n(\mathbf{k})$ is not really sensitive to the opening of a pseudogap. This is so both because of the finite temperature and because the wave vectors closest to the Fermi surface are actually quite far on the appropriate scale. For this filling, the value of $n(\mathbf{k})$ is fixed to $1/2$ on the Fermi surface itself.

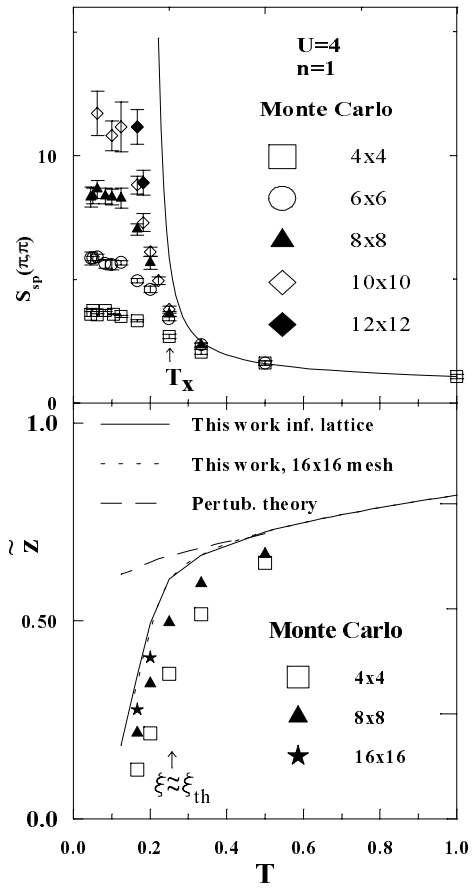

Fig. 7. — The upper part of the figure, adapted from reference [29], shows the temperature dependence of $S_{\rm sp}(\pi,\pi)$ at half-filling $n = 1$. The solid line is our theory for an infinite system while symbols are Monte Carlo data from reference [56]. The bottom part of the figure, adapted from reference [30], shows the behavior of $z(\tilde{T}) = -2G(\mathbf{k}_{\rm F}, \beta/2)$ in equation (49), as a function of temperature as obtained from Monte Carlo [53] simulations (symbols), from second order pertrubation theory (dashed line) and from our theory for an infinite system (solid line) and for a 16×16 lattice (dashed line).

It is thus necessary to find a dynamical quantity defined on the Fermi surface whose temperature dependence will allow us to unambiguously identify the pseudogap regime in both theory and in Monte Carlo data. The most dramatic effect is illustrated in the lower part of Figure 7 where we plot the quantity $\tilde{z}\,(T)$ defined by [30, 58]

$$\tilde{z}\,(T) = -2G\,(\mathbf{k}_{\rm F}, \beta/2) = \int \frac{{\rm d}\omega}{2\pi} \frac{A\,(\mathbf{k}_{\rm F}, \omega)}{\cosh\,(\beta\omega/2)}\,. \tag{49}$$

The physical meaning of this quantity $\tilde{z}\,(T)$ is that it is an average of the single-particle spectral weight $A\,(\mathbf{k}_{\rm F}, \omega)$ within $T \equiv 1/\beta$ of the Fermi level ($\omega = 0$). When quasiparticles exist, this is the best estimate of the usual zero-temperature quasiparticle renormalization factor $z \equiv 1/(1 - \partial\Sigma/\partial\omega)$ that can be obtained directly from imaginary-time Monte Carlo data. For non-interacting particles $\tilde{z}\,(T)$ is unity. For a normal Fermi liquid it becomes equal

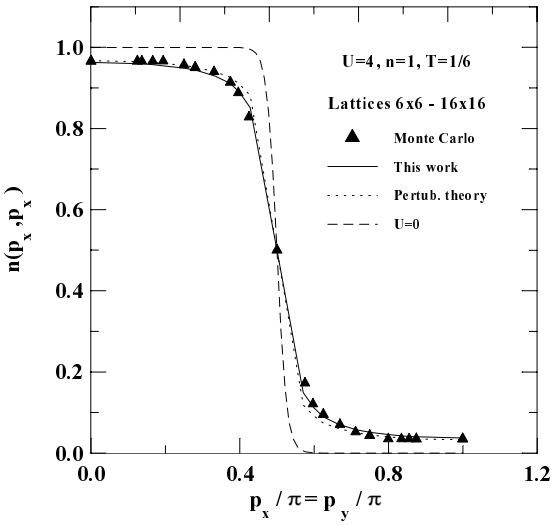

Fig. 8. — Occupation number $n(\mathbf{k})$ as a function of wave vector $\mathbf{k}$ at half-filling for $T = 1/6$, $U = 4$, and system sizes 6×6 to 16×16. The symbols are Monte Carlo results from reference [57] while the solid line is our theory and the dotted line is the prediction from second order perturbation theory. The dashed line shows the result for $U = 0$ as a reference.

to a constant less than unity as the temperature decreases since the width of the quasiparticle peak scales as T^2 and hence lies within T of the Fermi level. However, contrary to the usual $z \equiv 1/(1 - \partial\Sigma/\partial\omega)$ this quantity gives an estimate of the spectral weight $A(\mathbf{k}_{\mathrm{F}}, \omega)$ around the Fermi level, even if quasiparticles disappear and a pseudogap forms, as in the present case, (see Sect. 5).

One can clearly see from the lower part of Figure 7 that while second-order perturbation theory exhibits typical Fermi-liquid behavior for $\tilde{z}(T)$, both Monte Carlo data [53] and a numerical evaluation of our expression for the self-energy lead to a rapid fall-off of $\tilde{z}(T)$ below T_{X} (for $U = 4$, $T_{\mathrm{X}} \approx 0.2$ [29]). The rapid decrease of $\tilde{z}(T)$ clearly suggests non Fermi-liquid behavior. We checked also that our theory reproduces the Monte Carlo size-dependence. This dependence is explained analytically in Section 5.1.2. In reference [30] we have shown that at half-filling, our theory gives better agreement with Monte Carlo data [53] for $G(\mathbf{k}_{\mathrm{F}}, \tau)$ than FLEX, parquet or second order perturbation theory.

To gain a qualitative insight into the meaning of this drop in $\tilde{z}(T)$, we use the analytical results of the next section to plot in Figure 9 the value of $A(\mathbf{k}_{\mathrm{F}}, \omega)$. This plot is obtained by retaining only the contribution of classical fluctuations (Eq. (59)) to the self-energy. One sees that above T_{X}, there is a quasiparticle but that at $T \sim T_{\mathrm{X}}$ a minimum instead of a maximum starts to develop at the Fermi surface $\omega = 0$. Below T_{X}, the quasiparticle maximum is replaced by two peaks that are the precursors of antiferromagnetic bands. This is discussed in detail in much of the rest of this paper.

4.3. Phase Diagram. — The main features predicted by our approach for the magnetic phase diagram of the nearest-neighbor hopping model have been given in reference [29]. Needless to say, all our considerations apply in the weak to intermediate coupling regime. Note also that both quantum critical and renormalized classical properties of this model have been studied in another publication [33]. The shape of the phase diagram that we find is illustrated in Figure 10 for $U = 2.5$ and $U = 4$.

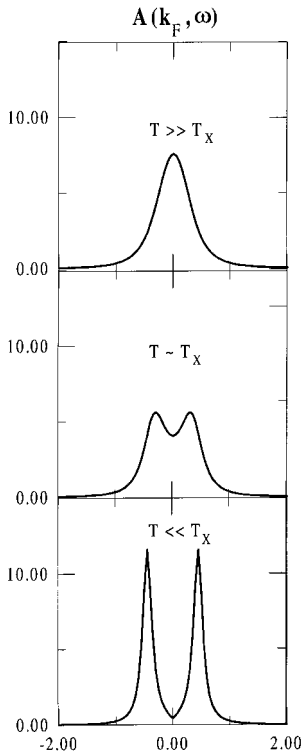

Fig. 9.

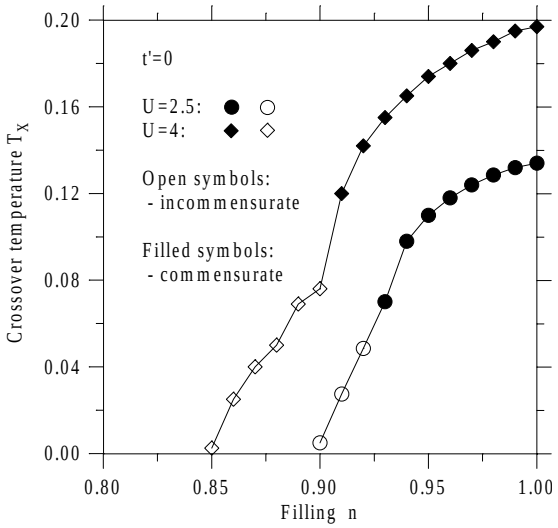

Fig. 10.

Fig. 9. — Qualitative sketch of the spectral weight at the Fermi wave vector at half-filling for three temperatures. This plot is obtained by retainig only the classical contribution to the self-energy using parameters corresponding to the typical $U = 4$ of Monte Carlo simulations. The top plot is for $T > T_{\rm X}$, the middle one for $T \sim T_{\rm X}$ and the bottom one for $T < T_{\rm X}$. The precursors of antiferromagnetic bands would look like this last figure.

Fig. 10. — Crossover temperature $T_{\rm X}$ as a function of filling for $U = 4$ and $U = 2.5$. On this crossover line, ξ^2 is enhanced by a factor of 500 over the bare value. Filled symbols indicate that the crossover is at the antiferromagnetic wave vector, while open symbols indicate a crossover at an incommensurate wave vector. Reproduced with permission from reference [100].

At zero temperature and small filling, the system is a paramagnetic Fermi liquid, whatever the value of the interaction U ($< W$). Then, as one moves closer to half-filling, one hits a quantum critical point at a value of filling $n_{\rm c}$. Since, $U_{\rm sp}$ in our theory saturates with increasing U, the value of $n_{\rm c}$ is necessarily larger than about $n_{\rm c}(U = \infty) = 0.68$. At this point, incommensurate order sets in at a wave vector $(q_{\rm c}, \pi)$ or at symmetry-related points. Whatever the value of U, the value of $q_{\rm c}$ is contained [29] in the interval $0.74\pi < q_{\rm c} \leq \pi$, increasing monotonously towards 0.74π as U increases. Since our approach applies only in the paramagnetic phase, at zero temperature we cannot move closer to half-filling. Starting from finite-temperature then, the existence of long-range order at low temperature is signaled by the existence of a crossover temperature $T_{\rm X}\,(n, U)$ below which correlations start to grow exponentially. We have already discussed the meaning of $T_{\rm X}\,(n, U)$ at half-filling. This crossover temperature becomes smaller

and smaller as one moves away from half-filling, until it reaches the quantum-critical point that we just discussed. The correlations that start to grow at $T_{\rm X}(n,U)$ when $n=1$ are at the antiferromagnetic wave vector, and they stay at this wave vector for a range of fillings n. Finally, at some filling, the correlations that start to grow at $T_{\rm X}(n,U)$ are at an incommensurate value until the quantum-critical point is reached.

Note that the above phase diagram is quite different from the predictions of Hartree-Fock theory mostly because of the strong renormalization of $U_{\rm sp}$. This quantitative change leads to qualitative changes in the Hartree-Fock phase diagram since, for example, Stoner ferromagnetism never occurs in our picture. While the existence of ferromagnetism in the strong coupling limit has been proven only recently [59], the absence of Stoner ferromagnetism in the Hubbard model was already suggested by Kanamori [2] a long time ago and was verified by more recent studies [44, 60, 61]. More relevant to the present debate though, is the fact that SDW order persists away from half-filling for a finite range of dopings. While this is in agreement with slave-boson approaches [62] and studies [63] using the infinite-dimension methodology [11], it is in clear disagreement with Monte Carlo simulations [64]. Our approach certainly fails sufficiently below $T_{\rm X}$, but given the successes described above, we believe that it can correctly predict the exponential growth of fluctuations at $T_{\rm X}$. It would be difficult to imagine how one could modify the theory in such a way that the growth of magnetic fluctuations does not occur even at incommensurate wave vectors. Also, such an approach would also need to stop the growth of fluctuations that we find as we approach the quantum critical point along the zero temperature axis, from the low-filling, paramagnetic side, where $T_{\rm X}(n<n_{\rm c},U)=0$.

It could be that Monte Carlo simulations [64] fail to see long-range order at zero temperature away from half-filling because at zero temperature, in the nearest-neighbor model, this order has a tendency to being incommensurate everywhere except at $n=1$. Furthermore, as we saw above, this incommensuration is in general far from one of the available wave vectors on an 8×8 lattice. It comes close to $(0.75\pi,\pi)$ only for the largest values of U available by Monte Carlo. Hence, incommensurate order on small lattices is violently frustrated not only by the boundary conditions, but also by the fact that there is no wave vector on what would be the Fermi surface of the infinite system. This means that the electron-electron interaction scatters the electrons at wave-vectors that are not those where the instability would show up, rendering these scatterings not singular. This is clearly an open problem.

5. Replacement of Fermi Liquid Quasiparticles by a Pseudogap in Two Dimensions below $T_{\rm X}$

One of the most striking consequences of the results discussed in the context of Monte Carlo simulations is the fall of the spectral weight below the temperature $T_{\rm X}$ where antiferromagnetic fluctuations start to grow exponentially in two dimensions. We have already shown in a previous publication [30] that this corresponds to the disappearance of Fermi liquid quasiparticles at the Fermi surface, well above the zero temperature phase transition. We also found that, simultaneously, precursors of the antiferromagnetic bands develop in the single-particle spectrum. Given the simplicity of our approach, it is possible to demonstrate this phenomenon analytically. This is particularly important here because size effects and statistical errors make numerical continuation of the Monte Carlo data to real frequencies particularly difficult. Such analytic continuations using the maximum entropy method [55] have, in the past, lead to a conclusion different from the one obtained later using singular value decomposition [65].

In this section then, we will consider the conditions for which Fermi liquid quasiparticles can be destroyed and replaced by a pseudogap in two dimensions. The major part of this section will be concerned with the single particle pseudogap and the precursors of antiferromagnetic

bands in the vicinity of the zero temperature antiferromagnetic phase transition in the positive U Hubbard model. However, it is well known that the problem of superconductivity is formally related to the problem of antiferromagnetism, in particular at half-filling where the nearest-neighbor hopping positive U Hubbard model maps exactly onto the nearest-neighbor negative U Hubbard model. The corresponding canonical transformation maps the $\mathbf{q} = (\pi, \pi)$ spin correlations of the repulsive model onto the $\mathbf{q} = \mathbf{0}$ pairing and $\mathbf{q} = (\pi, \pi)$ charge correlations of the attractive model while the single-particle Green's functions of both models are identical. Thus all our results below concerning the opening of the pseudogap in $A(\mathbf{k}_{\mathrm{F}}, \omega)$ in the repulsive U half-filled Hubbard model are directly applicable to the attractive U model at half-filling, the only difference being in the physical interpretation. While in the case of repulsive interaction the pseudogap is due to the critical thermal spin fluctuation, in the case of attractive interactions it is, obviously, due to the critical thermal pairing and charge fluctuations. Away from half-filling the mapping between two models is more complicated and the single particle spectra in the pairing pseudogap regime $A(\mathbf{k}_{\mathrm{F}}, \omega)$ have important qualitative differences with the single particle spectra in the magnetic pseudogap regime. However, even in this case there are very useful formal similarities between two problems so that in Section 5.6 we will give some simple analytical results for the self-energy in the regime dominated by critical pairing fluctuations.

The problem of precursor effects in the repulsive Hubbard model has been first studied by Kampf and Schrieffer [35]. Their analysis however was done at zero temperature and although the precursor effect that they found, called "shadow bands", looks similar to what we find, there are a number of important differences. For example, they find a quasiparticle between the precursors of antiferromagnetic bands, while we do not. Also, one does not obtain precursors at zero temperature when one uses our more standard expression for the dynamical susceptibility instead of the phenomenological form $\chi_{\mathrm{K\,Shr}} = f(\mathbf{q})g(\omega)$ that they use. The physical reason why a function that is separable in both momentum and frequency, such as $\chi_{\mathrm{K\,Shr}}$, leads to qualitatively different results than the conventional one has been explained in reference [36]. The microscopic justification for $\chi_{\mathrm{K\,Shr}}$ is unclear. We comment below on this problem as well as on some of the large related literature that has appeared lately.

Repeating some of the arguments of reference [30], we first show by general phase space arguments that the feedback of antiferromagnetic fluctuations on quasiparticles has the potential of being strong enough to destroy the Fermi liquid only in low enough dimension, the upper critical dimension being three. Then we go into more detailed analysis to give explicit analytic expressions for the quasi-singular part of the self-energy, first in Matsubara frequency. The analysis of the self-energy expression directly in real-frequencies is in Appendix (D). The latter analysis is useful to exhibit in the same formalism both the Fermi liquid limit and the non-Fermi liquid limit.

For simplicity we give asymptotics for $n = 1$ at the Fermi wave vector, where $\varepsilon(\mathbf{k}_{\mathrm{F}}) = 0$, but similar results apply for $n \neq 1$ as long as there is long-range order at $T = 0$ and one is below T_{X}. This case is also discussed briefly, but for more details the reader is referred to reference [36].

5.1. Upper Critical Dimension for the Destruction of Quasiparticles by Critical Fluctuations. —

Before describing the effect of spin fluctuations on quasiparticles, we first describe the so-called renormalized classical regime of spin fluctuations that precedes the zero-temperature phase transition in two dimensions.

5.1.1. Renormalized Classical Regime of Spin Fluctuations. — The spin susceptibility $\chi_{\mathrm{sp}}(\mathbf{q}, 0)$ below T_{X} is almost singular at the antiferromagnetic wave vector $\mathbf{Q}_2 = (\pi, \pi)$ because

the energy scale $\delta U \equiv U_{\rm mf,c} - U_{\rm sp}$ ($U_{\rm mf,c} \equiv 2/\chi_0(\mathbf{Q},0)$) associated with the proximity to the SDW instability becomes exponentially small [29]. This small energy scale, $\delta U \ll T$, leads to the so-called renormalized classical regime for the fluctuations [66]. In this regime, the main contribution to the sum over Matsubara frequencies entering the local-moment sum rule (Eq. (38)) comes from $iq_{\rm n} = 0$ and wave vectors $(\mathbf{q}-\mathbf{Q})^2 \leq \xi^{-2}$ near $\mathbf{Q}$. Approximating $\chi_{\rm sp}(\mathbf{q},0)$ by its asymptotic form

$$\chi_{\rm sp}(\mathbf{q},0) \approx \frac{1}{U_{\rm sp}\xi_0^2}\frac{2}{((\mathbf{q}-\mathbf{Q}_d)^2+\xi^{-2})} \tag{50}$$

where $\mathbf{Q}_2 = (\pi,\pi)$, $\mathbf{Q}_3 = (\pi,\pi,\pi)$ and

$$\xi_0^2 \equiv \frac{-1}{2\chi_0(Q)}\left.\frac{\partial^2\chi_0(\mathbf{q})}{\partial q_x^2}\right|_{\mathbf{q}=\mathbf{Q}_d} \quad ; \quad \xi \equiv \xi_0(U_{\rm sp}/\delta U)^{1/2} \tag{51}$$

we obtain, in d dimensions

$$\tilde{\sigma}^2 = \frac{2T}{U_{\rm sp}\xi_0^2}\int\frac{{\rm d}^d q}{(2\pi)^d}\frac{1}{q^2+\xi^{-2}} \tag{52}$$

where $\tilde{\sigma}^2 \equiv n - 2\langle n_\uparrow n_\downarrow\rangle - C < 1$ is the left-hand side of equation (38) minus corrections C that come from the sum over non-zero Matsubara frequencies (quantum effects) and from $(\mathbf{q}-\mathbf{Q})^2 \gg \xi^{-2}$. There is an upper cutoff to the integral which is less than or of the order of the Brillouin zone size. The important point is that the left-hand side of the above equation (Eq. (52)) is bounded and weakly dependent on temperature. This implies, as discussed in detail in reference [33], that the above equation leads to critical exponents for the correlation length that are in the spherical model ($n \to \infty$) universality class. For our purposes, it suffices to notice that the integral converges even when $\xi \to \infty$ in more than two dimensions. This leads to a finite transition temperature. In two dimensions, the transition temperature is pushed down to zero temperature and, doing the integral, one is left with a correlation length ξ that grows exponentially below $T_{\rm X}$

$$\xi \sim \exp\left(\pi\tilde{\sigma}^2\xi_0^2\frac{U_{\rm sp}}{T}\right). \tag{53}$$

The important consequence of this is that, below $T_{\rm X}$, the correlation length quickly becomes larger than the single-particle thermal de Broglie wave length $\xi_{\rm th} = v_{\rm F}/(\pi T)$. This has dramatic consequences on quasiparticles in two dimensions.

5.1.2. Effect of Critical Spin Fluctuations on Quasiparticles. — When the classical fluctuations ($iq_{\rm n} = 0$) become critical, they also give, in two dimensions, a dominant contribution to the self-energy *at low frequency*. To illustrate what we mean by the classical frequency contribution, neglect the contribution of charge fluctuations and single out the zero Matsubara frequency component from equation (46) to obtain

$$\Sigma(\mathbf{k},ik_{\rm n}) \approx Un_{-\sigma} + \frac{U}{4}\frac{T}{N}\sum_{\mathbf{q}} U_{\rm sp}\chi_{\rm sp}(\mathbf{q},0)\frac{1}{ik_{\rm n}-\tilde{\epsilon}_{\mathbf{k}+\mathbf{q}}}$$

$$+\frac{U}{4}\frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\rm n}\neq 0} U_{\rm sp}\chi_{\rm sp}(\mathbf{q},iq_{\rm n})\frac{1}{ik_{\rm n}+iq_{\rm n}-\tilde{\epsilon}_{\mathbf{k}+\mathbf{q}}}. \tag{54}$$

Here, $\tilde{\epsilon}_{\mathbf{k}}$ is measured relative to the chemical potential. The last term is the contribution from quantum fluctuations. In this last term, the sum over Matsubara frequencies $iq_{\rm n}$ must

be done before the analytical continuation of $ik_{\rm n}$ to real frequencies otherwise this analytical continuation would involve going through complex plane poles of the other terms entering the full sum over $iq_{\rm n}$. The contribution from classical fluctuations, $iq_{\rm n} = 0$, does not have this problem and furthermore it has the correct asymptotic behavior at $ik_{\rm n} \to \infty$. Hence the contribution of classical fluctuations to the retarded self-energy $\Sigma^{\rm R}(\mathbf{k}, \omega)$ can be obtained from the $iq_{\rm n} = 0$ term by trivial analytical continuation $ik_{\rm n} \to \omega + i0$. Note also that the chemical potential entering $G^{(0)}$ in the self-energy formula is $\mu_0 = \mu = 0$ at half-filling.

Doing the same substitution as above for the asymptotic form of the spin susceptibility (Eq. (50)) in the equation for the self-energy (Eq. (46)) one obtains the following contribution to Σ from classical fluctuations

$$\Sigma_{\rm cl}(\mathbf{k}, ik_{\rm n}) \cong \frac{UT}{2\xi_0^2} \int \frac{{\rm d}^d q}{(2\pi)^d} \frac{1}{q^2 + \xi^{-2}} \frac{1}{ik_{\rm n} - \tilde{\epsilon}_{\mathbf{k}+\mathbf{Q}} - \mathbf{q} \cdot \mathbf{v}_{\mathbf{k}+\mathbf{Q}}}, \tag{55}$$

where we have expanded $\tilde{\epsilon}_{\mathbf{k}+\mathbf{Q}+\mathbf{q}} \simeq \tilde{\epsilon}_{\mathbf{k}+\mathbf{Q}} + \mathbf{q} \cdot \mathbf{v}_{\mathbf{k}+\mathbf{Q}}$. In the case that we consider, namely half-filling and $\mathbf{k} = \mathbf{k}_{\rm F}$, we have $\mu_0 = \mu = 0$ and $\tilde{\epsilon}_{\mathbf{k}_{\rm F}+\mathbf{Q}} = 0$. The key point is again that in two dimensions the integral in this equation (Eq. (55)) is divergent at small q for $\xi = \infty$. In a Fermi liquid, the imaginary part of the self-energy at the Fermi surface ($\omega = 0$) behaves as $\Sigma''_{\rm R}(\mathbf{k}_{\rm F}, 0) \sim T^2$. Here instead, we find a singular contribution

$$\Sigma''_{\rm R}(\mathbf{k}_{\rm F}, 0) \propto T \int {\rm d}^{d-1} q_\perp \frac{1}{q_\perp^2 + \xi^{-2}} \propto T\xi^{3-d} \tag{56}$$

that is proportional to ξ in $d = 2$ and hence is very large $\Sigma''_{\rm R}(\mathbf{k}_{\rm F}, 0) \approx -U\xi/(\xi_{\rm th}\xi_0^2) > 1$ when the condition $\xi > \xi_{\rm th}$ is realized. By contrast, for $d = 3$, $\Sigma''_{\rm R}(\mathbf{k}_{\rm F}, 0) \sim -U(\ln \xi) / \left(\xi_0^2 \xi_{\rm th}\right)$, so that the Fermi liquid is destroyed only in a very narrow temperature range close the Néel temperature $T_{\rm N}$. Dimensional analysis again suffices to show that in four dimensions the classical critical fluctuations do not lead to any singular behavior. Three dimensions then is the upper critical dimension. As usual, logarithmic corrections exist at the upper critical dimension. The effect will be very small in three dimensions not only because it is logarithmic, but also because the fluctuation regime is very small, extending only in a narrow temperature range around the Néel temperature. By contrast, in two dimensions the effect extends all the way from the crossover temperature, $T_{\rm X}$, which is of the order of the mean-field transition temperature, to zero temperature where the transition is.

Wave vectors near Van Hove singularities are even more sensitive to classical thermal fluctuations. Indeed, near this point the expansion should be of the type $\epsilon_{\mathbf{k}_{\rm VH}+\mathbf{q}+(\pi,\pi)} \propto q_x^2 - q_y^2$. This leads, in two dimensions, to even stronger divergence in $\Sigma''_{\rm R}(\mathbf{k}_{\rm F}, 0) \propto T\xi^2 \int {\rm d}q_y \left[\left(2q_y^2 + 1\right)|q_y|\right]^{-1}$ [36]. Even if the logarithmic divergence is cutoff the prefactor is larger by a factor of ξ compared with points far from the Van Hove singularities.

5.2. Precursors of Antiferromagnetic Bands in Two Dimensions. — Let us analyze in more details the consequences of this singular contribution of critical fluctuations to the self-energy in two dimensions. The integral appearing in the two-dimensional version of the expression for the self-energy (Eq. (55)), can be performed exactly [67]

$$\Sigma(\mathbf{k}_{\rm F}, ik_{\rm n}) = \frac{U}{2} - i \frac{UT}{8\pi\xi_0^2 \sqrt{k_{\rm n}^2 - v_{\rm F}^2 \xi^{-2}}} \ln \frac{k_{\rm n} + \sqrt{k_{\rm n}^2 - v_{\rm F}^2 \xi^{-2}}}{k_{\rm n} - \sqrt{k_{\rm n}^2 - v_{\rm F}^2 \xi^{-2}}} + \mathcal{R}. \tag{57}$$

Here $\mathcal{R}$ is a regular part.

As a first application, we can use this expression to understand qualitatively both the temperature and size dependence of the Monte Carlo data for $\tilde{z}(T)$ appearing in Figure 2 of reference [30] or in the lower panel of Figure 7. Indeed, $\tilde{z}(T)$ can be written as the alternating series $-2G\left(\mathbf{k}_{\mathrm{F}}, \beta/2\right) = -4T\sum_{n=1}^{\infty}(-1)^n / \left(k_{\mathrm{n}} - \Sigma''\left(\mathbf{k}_{\mathrm{F}}, ik_{\mathrm{n}}\right)\right)$. Even though the series converges slowly, in the beginning of the renormalized classical regime and for qualitative purposes it suffices to use the first term of this series. Then, using the expressions for the correlation length (Eq. (53)) and for the self-energy (Eq. (57)), one finds

$$\tilde{z}(T) \sim \frac{T^2}{\tilde{\sigma}^2 U U_{\mathrm{sp}}} \sqrt{1 - \frac{\xi_{\mathrm{th}}^2}{\xi^2}}, \quad T_{\mathrm{X}} - T \ll T_{\mathrm{X}}. \tag{58}$$

On the infinite lattice, ξ starts growing exponentially below T_{X}, quickly becoming much larger than ξ_{th}. This implies $\tilde{z}(T) \simeq T^2$. On finite lattices $\xi \sim \sqrt{N}$, which explains the size effect observed in Monte Carlo *i.e.* smaller $\tilde{z}$ for smaller size N, ($\xi_{\mathrm{th}}(T_{\mathrm{X}}) \sim 5$ for Fig. 7).

The analytic continuation of $\Sigma\left(\mathbf{k}_{\mathrm{F}}, ik_{\mathrm{n}}\right)$ in equation (57) is

$$\Sigma^{\mathrm{R}}\left(\mathbf{k}_{\mathrm{F}}, \omega\right) = \frac{U}{2} + \frac{UT}{8\pi\xi_0^2\sqrt{\omega^2 + v_{\mathrm{F}}^2\xi^{-2}}} \left[\ln\left|\frac{\omega + \sqrt{\omega^2 + v_{\mathrm{F}}^2\xi^{-2}}}{\omega - \sqrt{\omega^2 + v_{\mathrm{F}}^2\xi^{-2}}}\right| - i\pi\right] + \mathcal{R}. \tag{59}$$

For the wave vectors $\mathbf{k}$ away from the Fermi surface the anomalous contribution due to the classical fluctuation has a similar form but with ω replaced by $(\omega - \tilde{\epsilon}_{\mathbf{k}+\mathbf{Q}})$. When $T > T_{\mathrm{X}}$, the correlation length ξ becomes of order unity and, as we will show in Appendix D, the regular part $\mathcal{R}$ dominates so that one recovers standard Fermi liquid behavior. Furthermore, even for large correlation length the regular part cannot be neglected when $\omega \gg T$ since the term exhibited here becomes small. Hence we concentrate on small frequencies and on $T < T_{\mathrm{X}}$ where the regular part $\mathcal{R}$ can be neglected.

Exactly at the Fermi level ($\omega = 0$) we recover the result of the previous section, namely that the imaginary part of the self-energy for $\xi > \xi_{\mathrm{th}}$ increases exponentially when the temperature decreases, $\Sigma''(\mathbf{k}_{\mathrm{F}}, 0) \sim U\xi/(\xi_{\mathrm{th}}\xi_0^2) \propto T\xi \propto T\exp\left(\pi\tilde{\sigma}^2\xi_0^2 U_{\mathrm{sp}}/T\right)$. The above analysis shows by contradiction that in the paramagnetic state below T_{X} there is no Fermi-liquid quasiparticle at k_{F}, yet the symmetry of the system remains unbroken at any finite T. Indeed, starting from quasiparticles ($G_\sigma^{(0)}$) we found that as temperature decreases, $\Sigma''_{\mathrm{R}}(\mathbf{k}_{\mathrm{F}}, 0)$ increases indefinitely instead of decreasing, in direct contradiction with the starting hypothesis. By contrast, a self-consistent treatment where we use in equation (46) the full G_σ with a large $\Sigma''_{\mathrm{R}}(\mathbf{k}_{\mathrm{F}}, 0)$ shows that, for $T < T_{\mathrm{X}}$, $\Sigma''_{\mathrm{R}}(\mathbf{k}_{\mathrm{F}}, 0)$ remains large in $d = 2$ and does not vanish as $T \to 0$, again confirming that the system is not a Fermi liquid in this regime (See however Sect. 6.2 below). Strong modifications to the usual Fermi liquid picture also persist away from half-filling as long as $T_{\mathrm{X}}(n) > 0$, as we discuss later.

One can check that the large $\Sigma''_{\mathrm{R}}(\mathbf{k}_{\mathrm{F}}, 0)$ in two dimensions (for $T < T_{\mathrm{X}}$) leads to a pseudogap in the infinite lattice, contrary to the conclusion reached in reference [55]. Indeed, instead of a quasiparticle peak, the spectral weight $A\left(\mathbf{k}_{\mathrm{F}}, \omega\right) \equiv -2\mathrm{Im}\, G_{\mathrm{R}}\left(\mathbf{k}_{\mathrm{F}}, \omega\right)$ has a minimum at the Fermi level $\omega = 0$ and two symmetrically located maxima away from it. More specifically, for $v_{\mathrm{F}}/\xi < |\omega| < T$ we have

$$A\left(\mathbf{k}_{\mathrm{F}}, \omega\right) \cong \frac{2\,|\omega|\, UT/(8\xi_0^2)}{[\omega^2 - UU_{\mathrm{sp}}\tilde{\sigma}^2/4]^2 + [UT/(8\xi_0^2)]^2}\,. \tag{60}$$

The maxima are located at $\omega = \pm\sqrt{UU_{\rm sp}}\tilde{\sigma}/2$. These two maxima away from zero frequency correspond to precursors of the zero-temperature antiferromagnetic (or SDW) bands (shadow bands [35]). There is no quasiparticle peak between these two maxima when $\xi > \xi_{\rm th}$. This remains true in the case of no perfect nesting as well [36] (see also Sect. 5.5). We note that this is different from the results of the zero-temperature ($\xi_{\rm th} = \infty$) calculations of Kampf and Schrieffer [35] that were based on a phenomenological susceptibility separable in momentum and frequency $\chi_{\rm K.Sh.} = f(\mathbf{q})g(\omega)$. As was explained in reference [36], the existence of precursors of antiferromagnetic bands (shadow bands in the terminology of Ref. [35]) at zero temperature is an artifact of the separable form of the susceptibility. The third peak between the two precursors of antiferromagnetic bands that was found in reference [35] is due to the fact that at zero temperature the imaginary part of the self-energy $\Sigma''(\mathbf{k}, \omega = 0, T = 0)$ is strictly zero at all $\mathbf{k}$. In our calculations, precursor bands appear only at finite temperature when the system is moving towards a zero-temperature phase transition. In this case, the imaginary part of the self-energy goes to infinity for $\mathbf{k}$ on the "shadow Fermi surface" $\lim_{T\to 0} \Sigma''(\mathbf{k}_{\rm F} + \mathbf{Q}, 0) \propto T\xi \propto T \exp(C_{\rm st}/T) \to \infty$ and to zero at all other wave vectors. This is consistent with the SDW result which we should recover at $T = 0$. Indeed, the latter result can be described by the self-energy $\Sigma^{\rm R}(\mathbf{k}, \omega) = \Delta^2/(\omega - \tilde{\varepsilon}(\mathbf{k} + \mathbf{Q}) + i\eta)$ which implies that the imaginary is a delta function $\Sigma''(\mathbf{k}, \omega) = -\pi\delta(\omega - \tilde{\varepsilon}(\mathbf{k} + \mathbf{Q}))$ instead of zero at all $\mathbf{k}$ as in a Fermi liquid. We note also that analyticity and the zero value of $\Sigma''(\mathbf{k}, \omega = 0)$ in reference [35] automatically implies that the slope of the real part of the self-energy $\partial\Sigma'(\mathbf{k}, \omega)/\partial\omega|_{\omega=0}$ is negative. By contrast, in our case $\partial\Sigma'(\mathbf{k}_{\rm F} + \mathbf{Q}, \omega)/\partial\omega|_{\omega=0}$ is positive and increases with decreasing temperature, eventually diverging at the zero-temperature phase transition. The real part of the self-energy obtained using the asymptotic form equation (59) is at the bottom left corner of Figure 11 with the corresponding spectral function $A(\mathbf{k}_{\rm F}, \omega)$ shown above it. In Figure 9 we have already shown the evolution of the spectral function $A(\mathbf{k}_{\rm F}, \omega)$ with temperature. The positions of the precursors of antiferromagnetic bands scale like $\tilde{\sigma}/2$ which itself, at small coupling in two dimensions, scales like the mean field SDW transition temperature or gap (see Appendix B of Ref. [33]). As U increases, the predicted positions of the maxima obtained from the asymptotic form (Eq. (60)) will be less accurate since they will be at intermediate frequencies and the regular quantum contribution to the self-energy will affect more and more the position of the peaks.

We have predicted [30] that the exponential growth of the magnetic correlation length ξ below $T_{\rm X}$ will be accompanied by the appearance of precursors of SDW bands in $A(\mathbf{k}_{\rm F}, \omega)$ with no quasiparticle peak between them. By contrast with isotropic materials, in quasi-two-dimensional materials this effect should exist in a wide temperature range, from $T_{\rm X}$ ($T_{\rm X} \ll U < E_{\rm F}$) to the Néel temperature $T_{\rm N}$ ($T_{\rm X} - T_{\rm N} \sim 10^2$ K).

5.3. Contrast between Magnetic Precursor Effects and Hubbard Bands. — Although there are some formal similarities between the precursors of antiferromagnetic bands and the Hubbard bands (see Sect. 6) we would like to stress that these are two different physical phenomena. A clear illustration of this is when a four peak structure exists in the spectral function $A(\mathbf{k}, \omega)$, two peaks being precursors of antiferromagnetic bands, and two peaks being upper and lower Hubbard bands. The main differences between these bands are in the $\mathbf{k}$-dependence of the self-energy $\Sigma(\mathbf{k}, \omega)$ and in the conditions for which these bands develop. Precursors of antiferromagnetic bands appear even for small U in the renormalized classical regime $T < T_{\rm X}$, and their dispersion has the quasi-periodicity of the magnetic Brillouin zone. In contrast, upper and lower Hubbard bands are high-frequency features that appear only for sufficiently large $U > W$ and $T < U$ and have the periodicity of the whole Brillouin zone in the paramagnetic state. Furthermore, the existence of Hubbard bands is not sensitive to

dimensionality so they exist even in infinite dimension where the self-energy does not depends on momentum $\mathbf{k}$ at all. In contrast, the upper critical dimension for the precursors of antiferromagnetic bands is three (see Sect. 5.4).

In our theory the precursors of antiferromagnetic bands come from the almost singular behavior of the zero Matsubara frequency susceptibility $\chi_{\rm sp}(\mathbf{q}, 0)$, which leads to the characteristic behavior of $\Sigma(\mathbf{k}, \omega) = \Delta^2_{\rm Sh.B}/(\omega - \varepsilon(\mathbf{k} + \mathbf{Q}))$ with $\Delta^2_{\rm Sh.B} \propto T \ln(\xi)$. On another hand, the Hubbard bands appear in our theory because the high-frequency asymptotics $\Sigma(\mathbf{k}, \omega) \propto \Delta^2_{\rm H.B}/\omega$ has already set in for $\omega > W$, and this leads to the bands at $\omega = \pm\Delta_{\rm H.B}$ for $\Delta > W$ (see for more details Sect. 6). The coefficient $\Delta^2_{\rm H.B}$ is determined by the sum over all Matsubara frequencies and $\mathbf{q}$: $\Delta^2_{\rm H.B} = TUN^{-1}\sum_{\mathbf{q},n}[U_{\rm sp}\chi_{\rm sp}(\mathbf{q}, i\omega_{\rm n}) + U_{\rm ch}\chi_{\rm ch}(\mathbf{q}, i\omega_{\rm n})]$.

It was noticed in Monte Carlo simulations [68,78] that for intermediate U, the spectral weight has four maxima. We think that peaks at $\omega \sim \pm U/2$ are Hubbard bands, while the peaks closer to $\omega = 0$ are precursors of antiferromagnetic bands. If this interpretation is correct, then the latter peaks should disappear with increasing temperature when ξ becomes smaller than $\xi_{\rm th}$, while the Hubbard bands should exist as long as $T < U$.

While the location of the precursors of antiferromagnetic bands should be accurate in our theory, the same will not be true for the location of the upper and lower Hubbard bands. This is because our theory is tuned to the low frequency behavior of the irreducible vertices and does not have the right numerical coefficient in the high-frequency expansion of the self-energy, as shown in equation (E.10) below. Nevertheless, our analytical approach to date is the only one that agrees at least qualitatively with the finding that precursors of antiferromagnetic bands as well as upper and lower Hubbard bands can occur simultaneously. Note however that a four peak structure at $n = 1$ was also obtained in reference [70] but the physical difference between Hubbard bands and precursors of antiferromagnetic bands was not clearly spelled out. We comment on recent findings of the FLEX approach in Section 6 [37, 38, 69].

5.4. Can the Precursors of Antiferromagnetic Bands Exist in Three Dimensions?. — In two dimensions, the finite-temperature phase is disordered, but the zero-temperature one is ordered and has a finite gap, except at the quantum critical point away from half-filling. Hence, precursors of antiferromagnetic bands that appear in the paramagnetic state do so with a finite pseudogap which appears consistent with the finite zero-temperature gap towards which the system is evolving. By contrast, in higher dimensions the gap opens-up with a zero value at the transition temperature. Based on this simple argument, one does not expect precursors of antiferromagnetic bands in dimensions larger than two (see, however, below). Here, we will also show that there is no phase space reasons for the existence of precursors of the antiferromagnetic bands when $d > 2$.

We have already shown that in three dimensions the quasiparticle at the Fermi level at half-filling will have an imaginary part of the self-energy that grows like $T \ln \xi$, an effect that is much weaker than $T\xi$ found in two dimensions. Despite this small effect, in three dimensions the classical fluctuations do not affect the self-energy for energies larger than $v_{\rm F}\xi^{-1}$. Indeed, consider the contribution of classical thermal fluctuations to the self-energy (Eq. (55)). In two dimensions, we have for $|\omega| > v_{\rm F}\xi^{-1}$

$$\mathrm{Re}\left[\Sigma^{2d}_{\rm cl}(\mathbf{k}_{\rm F}, \omega)\right] \cong \frac{UT}{2\xi_0^2}\int \frac{\mathrm{d}^2 q}{(2\pi)^2}\frac{1}{q^2 + \xi^{-2}}\frac{1}{\omega}, \tag{61}$$

which allows us to recover the approximate formula for the spectral weight given in equation (60) above. In three dimensions however, this approximation cannot be done because the integral is not dominated by small values of q. To see this explicitly in three dimensions,

consider the contribution of classical thermal fluctuations

$$\Sigma^{3d}_{\rm cl}\left(\mathbf{k}_{\rm F},\omega+i\eta\right) \cong \frac{UT}{2\xi_0^2}\int\frac{\mathrm{d}q_\parallel}{2\pi}\int\frac{\mathrm{d}^2q_\perp}{(2\pi)^2}\frac{1}{q_\perp^2+q_\parallel^2+\xi^{-2}}\frac{1}{\omega+i\eta+v_{\rm F}q_\parallel} \tag{62}$$

$$\cong \frac{UT}{2\xi_0^2}\frac{1}{4\pi}\int\frac{\mathrm{d}q_\parallel}{2\pi}\ln\left[\frac{\Lambda_\perp^2+q_\parallel^2+\xi^{-2}}{q_\parallel^2+\xi^{-2}}\right]\frac{1}{\omega+i\eta+v_{\rm F}q_\parallel}. \tag{63}$$

As long as $|\omega| > v_{\rm F}\xi^{-1}$, the logarithmic singularity that develops at $q_\parallel = 0$ when $\xi^{-1} \to 0$ is integrable and gives no singular contribution to the self-energy. Hence, unusual effects of classical thermal fluctuations are confined to the range of frequencies $|\omega| < v_{\rm F}\xi^{-1}$. At higher frequencies, $|\omega| > v_{\rm F}\xi^{-1}$, all bosonic Matsubara frequencies in equation (46) need to be taken into account and from phase space considerations alone there is no reason for the existence of precursors of antiferromagnetic bands in the 3D case. However, the existence of such bands in 3D cannot be completely excluded based on dimensional arguments alone because they occur at finite frequencies and strictly speaking they are non-universal. In particular, as discussed in reference [33], one expects to see precursors that look like 2D antiferromagnetic bands (shadow bands) in the vicinity of the finite temperature phase transition in strongly anisotropic quasi-two-dimensional material. On the other hand, such bands do not generically exist in the almost isotropic 3D case, because even in 2D the conditions for such bands are quite stringent. The difference between shadow bands and Hubbard bands has been discussed in the previous subsection and the discussion of non-analyticities sometimes encountered in Fermi liquid theory can be found in Appendix D.

5.5. Away from Half-Filling. — Close to half-filling, in the nearest-neighbor hopping model, one can enter a renormalized classical regime with large antiferromagnetic correlation length, even though the zero-temperature Fermi surface properties may favor incommensurate correlations. This renormalized-classical regime with large (π,π) correlations occurs when $T_{\rm X} \gg \mu_0$. By arguments similar to those above, one finds that in this regime one still has precursors of antiferromagnetic bands. However, the chemical potential is in or near the lower precursor band and the system remains metallic. The high-frequency precursor appears only below $T_{\rm X}$ at $\omega \approx \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}$.

With second-neighbor hopping, the points of the Fermi surface that intersect the magnetic Brillouin zone (hot spots) behave as does the whole Fermi surface of the nearest-neighbor (nested) case discussed above. These questions were discussed in detail in reference [36].

5.6. The Pairing Pseudogap and Precursors of Superconducting Bands in Two Dimensions. — As we have already pointed out above, the results for the single particle spectra obtained for the half-filled nearest-neighbor hopping repulsive Hubbard model can be directly applied to the corresponding attractive Hubbard model, in which case the pseudogap opens up in the renormalized classical regime of pairing and charge fluctuations. Away from half-filling, the symmetry between charge and pair correlations is lost and pair fluctuations dominate, becoming infinite at the Kosterlitz-Thouless transition temperature. This temperature is below the temperature at which the magnitude of the pair order parameter acquires rigidity despite the randomness of its phase. One expects then that a pseudogap will also open in this case when the correlation length for pairing fluctuations becomes larger than the single-particle thermal de Broglie wavelength $\xi_{\rm pairing} > \xi_{\rm th} = v_{\rm F}/T$. This should occur below the crossover temperature to the renormalized classical regime of pairing fluctuations but above the Kosterlitz-Thouless transition temperature.

The quantitative microscopic theory for the negative U Hubbard model will be considered in a separate publication. By contrast with all other sections of this paper, our considerations here will be more phenomenological. Nevertheless, they will allow us to present some analytical results for the self-energy obtained in the critical regime dominated by pairing fluctuations. Details of the model should not be very important since we are in a regime where everything is dominated by long wave length fluctuations.

The derivation of $\Sigma(\mathbf{k},\omega)$ in the pairing case is a straightforward extension of what we did in the antiferromagnetic case (see Sects. 5.1.2, 5.2 and Ref. [36]). In particular, in complete analogy with the magnetic case, the main contribution to the self-energy in the critical regime comes from the classical thermal fluctuations $iq_{\mathrm{n}} = 0$. Assuming some effective coupling constant g' between quasiparticles and pairing fluctuations, which in general can be momentum dependent, one can write in the one loop approximation

$$\Sigma_{\mathrm{cl}}(\mathbf{k}, ik_{\mathrm{n}}) \approx T g'(\mathbf{k}) \int \frac{\mathrm{d}^2 q}{(2\pi)^2} \frac{1}{\xi_{\mathrm{p}}^{-2} + q^2} \frac{1}{ik_{\mathrm{n}} + \tilde{\epsilon}_{-\mathbf{k}+\mathbf{q}}} \,. \tag{64}$$

Here $\tilde{\epsilon}_{\mathbf{k}}$ is the electron dispersion relative to the chemical potential, and all factors in front of integral are reabsorbed into the coupling constant g'. This expression is similar to the expression (Eq. (55)) in the magnetic case but there are two important differences: *i)* instead of $\tilde{\epsilon}_{\mathbf{k}+\mathbf{Q}+\mathbf{q}}$ we have now $\tilde{\epsilon}_{-\mathbf{k}+\mathbf{q}}$; *ii)* there is no minus sign in front of $\tilde{\epsilon}_{-\mathbf{k}+\mathbf{q}}$. The first difference is due to the fact that superconductivity usually occurs with zero center of mass momentum for the pair, and hence the pairing susceptibility in the normal state $\chi_{\mathrm{p}} \propto 1/(\xi_{\mathrm{p}}^{-2} + q^2)$ must be peaked near $\mathbf{q} = 0$, (the integration variable $\mathbf{q}$ in equation (55) was measured relative to $\mathbf{Q} = (\pi, \pi)$). The second difference comes from the fact that we are now considering the contribution to Σ coming from the particle-particle channel instead of the particle-hole channel. Taking the integrals over $\mathbf{q}$ and using the fact that small $\mathbf{q}$ only will contribute we neglect the $\mathbf{q}$ dependence of the coupling constant and obtain for the imaginary part of Σ_{cl} the following expression

$$\Sigma''(\mathbf{k},\omega) = -\frac{g'(\mathbf{k})T}{4\sqrt{(\omega + \tilde{\varepsilon}_{-\mathbf{k}})^2 + v_{-\mathbf{k}}^2 \xi_{\mathrm{p}}^{-2}}} \,. \tag{65}$$

In the renormalized classical regime the pairing correlation length ξ_{p} increases faster with decreasing temperature than $\xi_{\mathrm{th}} = v_{\mathrm{F}}/T$. Consequently, $\Sigma''(\mathbf{k}_{\mathrm{F}}, 0)$ tends to diverge with decreasing temperature and a pairing pseudogap in the spectral function $A(\mathbf{k}_{\mathrm{F}}, \omega)$ opens up over the complete Fermi surface, except maybe at a few points where $g'(\mathbf{k}) = 0$. This is different from the antiferromagnetic case, where the pseudogap in $A(\mathbf{k}_{\mathrm{F}} = \mathbf{k}_{\mathrm{h.sp.}}, \omega)$ opens up only when, so called, "hot spots" ($\tilde{\varepsilon}(\mathbf{k}_{\mathrm{h.sp.}} + \mathbf{Q}) = \tilde{\varepsilon}(\mathbf{k}_{\mathrm{h.sp.}}) = 0$) exist in a given model [36]. The antiferromagnetic pseudogap opens everywhere on the Fermi surface only in the case of perfect nesting, where all points on the Fermi surface are "hot spots".

The real part of the self-energy can be obtained from equation (65) using the Kramers-Kronig relation and has the form:

$$\Sigma'(\mathbf{k},\omega) = \frac{g'(\mathbf{k})T}{4\pi\sqrt{(\omega + \tilde{\varepsilon}_{-\mathbf{k}})^2 + v_{-\mathbf{k}}^2 \xi_{\mathrm{p}}^{-2}}} \ln \left| \frac{\omega + \tilde{\varepsilon}_{-\mathbf{k}} + \sqrt{(\omega + \tilde{\varepsilon}_{-\mathbf{k}})^2 + v_{-\mathbf{k}}^2 \xi_{\mathrm{p}}^{-2}}}{\omega + \tilde{\varepsilon}_{-\mathbf{k}} - \sqrt{(\omega + \tilde{\varepsilon}_{-\mathbf{k}})^2 + v_{-\mathbf{k}}^2 \xi_{\mathrm{p}}^{-2}}} \right| . \tag{66}$$

To understand how precursors of the superconducting bands develop, let us look at $\Sigma'(\mathbf{k},\omega)$ at frequencies $|\omega + \tilde{\varepsilon}_{-\mathbf{k}}| \gg v_{-\mathbf{k}} \xi_{\mathrm{p}}^{-1}$. In this case, using inversion symmetry $\tilde{\varepsilon}_{-\mathbf{k}} = \tilde{\varepsilon}_{\mathbf{k}}$, one can obtain from equation (66) the following asymptotic form

$$\Sigma'(\mathbf{k},\omega) \approx \frac{g'(\mathbf{k})}{2\pi} \frac{T \ln \xi_{\mathrm{p}}}{\omega + \tilde{\varepsilon}_{\mathbf{k}}} \,. \tag{67}$$

When, $\xi_{\rm p} \sim \exp(\text{const}/T)$ (see, more general case below) this form of the self-energy leads to the usual BCS result $\Sigma'(\mathbf{k},\omega) \approx \Delta^2(\mathbf{k})/(\omega+\tilde{\varepsilon}_{\mathbf{k}})$ with the gap $\Delta^2(\mathbf{k}) \approx (g'(\mathbf{k})/2\pi)T\ln\xi_{\rm p}$. On the other hand, the imaginary part $\Sigma''(\mathbf{k},\omega)$, equation (65), vanishes everywhere in the $T=0$ limit, except when $\omega=-\tilde{\varepsilon}_{\mathbf{k}}$ where it becomes infinite. The results for Σ' and Σ'' can thus be combined to write for the corresponding limit of the retarded self-energy $\Sigma^{\rm R} = \Delta^2(\mathbf{k})/(\omega+\tilde{\varepsilon}_{\mathbf{k}}+i\eta)$. This limit leads to the standard BCS expression for the normal Green's function when substituted back into the Dyson equation $G^{\rm R} = 1/\left(\omega+i\eta-\tilde{\varepsilon}_{\mathbf{k}}-\Sigma^{\rm R}(\mathbf{k},\omega)\right)$. Above the transition temperature, the anomalous Green's function remains zero since there is no broken symmetry. The qualitative picture for the development of the pairing pseudogap and of the precursors of superconducting bands at $\mathbf{k}=\mathbf{k}_{\rm F}$ is illustrated in Figure 9 and in the left part of Figure 11. While in the case of magnetic critical fluctuations these figures describe the precursor effect in $A(\mathbf{k}_{\rm F},\omega)$ for perfect-nesting or for the "hot spots" (when such points exist), in the case of pairing fluctuations they describe the spectra for all $\mathbf{k}_{\rm F}$ and for all fillings where the ground state is superconducting.

We need to comment on a subtle difference between the antiferromagnetic and the pairing precursor effects in the single particle spectra. While the magnetic order parameter has three components and can order only at zero temperature in the two-dimensional repulsive model, away from half-filling in the attractive model the pairing order parameter becomes the only relevant order parameter at low temperature. Since it has only two components, a finite temperature Kosterlitz-Thouless phase transition is then allowed in two dimensions. The critical behavior in vicinity of this transition is given by $\xi_{\rm p} \propto \exp[\text{const}/(T-T_{\rm KT})^{1/2}]$ instead of $\xi \propto \exp(\text{const}/T)$ as in the magnetic case. To take this properly into account one would need a treatment of the problem that is more sophisticated than that given above. In particular, one would have to take into account corrections to the simple form that we used for the pairing susceptibility $\chi_{\rm p}(q,0) \propto 1/(\xi_{\rm p}^{-2}+q^2)$. This Lorentzian form of the susceptibility in the critical regime is strictly valid only in the $n=\infty$ limit (n is the number of the components of the order parameter) and is, clearly, a less accurate approximation in the case of pairing fluctuations ($n=2$) than in the case of the antiferromagnetic fluctuations ($n=3$). Nevertheless, we believe that qualitatively the picture given above is correct for two reasons. First, because in the Kosterlitz-Thouless picture the magnitude of the order parameter is locally non-zero starting below a crossover temperature $T_{\rm X}$ that is larger than the transition temperature $T_{\rm KT}$. It is only the phase that is globally decorrelated above $T_{\rm KT}$. This means that locally the quasiparticles are basically in a superconducting state even above $T_{\rm KT}$. A second reason to believe in the precursor effects is that the superfluid density and the gap are finite as $T \to T_{\rm KT}^-$ and, hence, the two peak structure in $A(\mathbf{k}_{\rm F},\omega)$ exists even as the phase transition point is approached from the low-temperature side. By analogy with the antiferromagnetic case, this two peak structure should not immediately disappear when one increases the temperature slightly above $T_{\rm KT}$.

Finally, we point out that the precursor phenomenon described above has to be distinguished from, so-called, pre-formed pairs considered first by Nozières and Schmitt-Rink [71] (see also [72]). These pre-formed pairs exist in any dimension when the coupling strength is sufficiently large, while the precursor effect considered above can be caused by arbitrarily small attractive interactions but only in two dimensions. We think that recent Monte Carlo data [73] on the negative $U=-W/2$ Hubbard model illustrates the opening of the single-particle pseudogap due to critical fluctuations, rather than a strong-coupling effect. In these simulations, the drop in the density of states at the Fermi level should be accompanied by a simultaneous rapid increase of the pairing structure factor $S_{\rm p}(\mathbf{q}=0,T)$. The latter must be exponential in the infinite 2D lattice and a size analysis of Monte Carlo data similar to the one shown in Figure 7 would be extremely helpful to clarify this issue.

6. Absence of the Precursors of Antiferromagnetic Bands and Upper and Lower Hubbard Bands in Eliashberg-Type Self-Consistent Theories

In this section, we explain why the theories that use self-consistent propagators but neglect the corresponding frequency-dependent vertex corrections fail to see two important physical effects: namely upper and lower Hubbard bands, as well as the precursors of antiferromagnetic bands that we just discussed. The failure of this type of self-consistent schemes to correctly predict upper and lower Hubbard bands has been realized a long time ago in the context of calculations in infinite dimension [11, 74]. While one may brush aside this failure by claiming that high-energy phenomena are not so relevant to low-energy physics, we show that in fact these schemes also fail to reproduce the low-energy pseudogap and the precursors of antiferromagnetic bands for essentially the same reasons that they *fail* to see Hubbard bands. It is thus useful to start by a discussion of the better understood phenomenon of upper and lower Hubbard bands and then to move to precursors of antiferromagnetic bands.

6.1. Why Eliashberg-Type Self-Consistency for the Electronic Self-Energy Kills Hubbard Bands. — We first note that ordinary perturbation theory satisfies the correct high-frequency behavior (Eq. (68)) for the self-energy namely, for $k_{\rm n} \gg W$

$$\lim_{ik_{\rm n}\to\infty} \Sigma_\sigma\left(\mathbf{k}, ik_{\rm n}\right) = Un_{-\sigma} + \frac{U^2 n_{-\sigma}\left(1-n_{-\sigma}\right)}{ik_{\rm n}} + \cdots \tag{68}$$

It is the latter property that guarantees the existence of the Hubbard bands for $U > W$. To see this, consider the half-filled case. In this case, $n_{-\sigma} = 1/2$, $\mu = U/2$ and one finds for the spectral weight

$$A\left(\mathbf{k},\omega\right) \sim \frac{-2\Sigma''}{\left(\omega - \frac{U^2}{4\omega}\right)^2 + \Sigma''^2} \tag{69}$$

which has pronounced maxima at the upper and lower Hubbard bands, namely $\omega = \pm U/2$, has long as Σ'' is not too large. Since these results are obtained using high-frequency asymptotics, they are valid only when the asymptotic equation (68) has already set in when $\omega \sim U/2$. In the exact theory and in ordinary perturbation theory in terms of bare Green functions $G^{(0)}$, equation (68) is valid for $|\omega| \gg W$and the Hubbard bands appears as soon as U becomes larger than W.

The fact that this simple high-frequency behavior sets in at the energy scale given by W rather than U, even when $W < U$, is a non-trivial consequence of the Pauli principle. To see this we first recall the exact result for the self-energy $\Sigma_\sigma\left(\mathbf{k}, ik_{\rm n}\right)$ in the atomic limit [1]

$$\Sigma_\sigma^{\rm atomic}\left(\mathbf{k}, ik_{\rm n}\right) = U{\rm n}_{-\sigma} + \frac{U^2 n_{-\sigma}\left(1-n_{-\sigma}\right)}{ik_{\rm n} + \mu - U\left(1-n_{-\sigma}\right)}\,. \tag{70}$$

Formally, the atomic limit means that hopping is the smallest of all energy scales in the problem, including the temperature, $t \ll T$, which is not a very interesting case. However, the same arguments that have been used to derive the expression (70) in the atomic limit can be used to show that equation (70) is valid at any T/t when $k_{\rm n} \gg W$. Indeed, in the equations of motion for two-particle correlators [1] one can neglect hopping terms when $k_{\rm n} \gg W$. This is where the asymptotic behavior (70) sets in since the equations of motion then immediately lend themselves to a solution without any additional approximation for the interacting term. This solution is possible because the Pauli principle $n_{i\sigma}^2 = n_{i\sigma}$ allows us to collapse three-particle correlation function which enters equation of motion to the two-particle one

$U\left\langle T_\tau\left(n_{i-\sigma}(\tau_i)\, n_{i-\sigma}(\tau_i)\, c_{i\sigma}(\tau_i)\, c^\dagger_{j\sigma}(\tau_j)\right)\right\rangle = U\left\langle T_\tau\left(n_{i-\sigma}(\tau_i)\, c_{i\sigma}(\tau_i)\, c^\dagger_{j\sigma}(\tau_j)\right)\right\rangle$. Hence, the expression for atomic limit (Eq. (70)) is also a general result for the self-energy that is valid for $k_\mathrm{n} \gg W$. At half-filling $n_{-\sigma} = 1/2$, $\mu = U/2$ and the asymptotic (68) sets in at $k_\mathrm{n} \sim W$, as was pointed out above. Away from half-filling, as long as $|\mu - \Sigma(\infty)|$ and $|\mu - U(1 - n_{-\sigma})|$ are both much smaller than W, (they both vanish at half-filling), the asymptotic behavior will also start at $k_\mathrm{n} \sim W$.

The situation is qualitatively different when one uses dressed Green functions, but does not take into account the frequency dependence of the vertex, as it is done in FLEX (see Eq. (E.9)) or for second-order perturbation theory with dressed G. For example, the second-order expression for $\Sigma_\sigma(\mathbf{k}, ik_\mathrm{n})$ in terms of full G does satisfy the asymptotics equation (68), but it sets in too late, namely for $k_\mathrm{n} \gg U$, instead of $k_\mathrm{n} \gg W$. Indeed, when $k_\mathrm{n} \gg W$, the equation for the self-energy at half-filling in this type of theories reduces to

$$\Sigma(ik_\mathrm{n}) = \frac{\Delta^2}{ik_\mathrm{n} - \Sigma(ik_\mathrm{n})} \tag{71}$$

where $\Delta^2 = cU^2/4$ with c a constant of proportionality involving the sum over all wave vectors and Matsubara frequencies of the self-consistent dynamical susceptibilities. In a given theory the value of c may differ from its value $c = 1$ obtained from the exact result (Eq. (70)), but its always of order unity. The solution of equation (71)

$$\Sigma(ik_\mathrm{n}) = \frac{1}{2} ik_\mathrm{n} - \frac{1}{2}\sqrt{(ik_\mathrm{n})^2 - 4\Delta^2} \tag{72}$$

has the analytically continued form

$$\mathrm{Re}\,\Sigma^\mathrm{R}(\omega) = \frac{\omega}{2} - \frac{\omega}{2|\omega|}\theta(|\omega| - 2\Delta)\sqrt{\omega^2 - 4\Delta^2} \tag{73}$$

$$\mathrm{Im}\,\Sigma^\mathrm{R}(\omega) = -\frac{1}{2}\theta(2\Delta - |\omega|)\sqrt{4\Delta^2 - \omega^2}. \tag{74}$$

From this one can immediately see that a U^2/ω regime exists for $\mathrm{Re}\,\Sigma^\mathrm{R}(\omega)$ only when $|\omega| \gg U$, (with $2\Delta = U$).

This means that such regime sets in too late to give the Hubbard bands described by equation (69), because the Hubbard bands occur at $\omega = \pm U/2$ and for such ω the asymptotic form $\Sigma^\mathrm{R} \propto U^2/\omega$ is not valid yet in FLEX and similar theories. Consequently, instead of well defined peaks at $\omega = \pm U/2$ in the half-filled case, one obtains only long tails in the spectral function $A_\sigma(\mathbf{k}, \omega)$, no matter how large U is [74] (see also following subsection).

This explains why there is no Hubbard bands in any theory that uses self-consistent Green's functions, but neglects the frequency dependence of the vertex. This is an explicit example that illustrates what seems to be a more general phenomenon when there is no Migdal theorem for vertex corrections: a calculation with dressed Green's functions but no frequency dependent vertex correction often gives worse results than a calculation done with bare Green's functions and a frequency independent vertex.

6.2. Why FLEX Fails to See Precursors of Antiferromagnetic Bands. — In this subsection we describe the qualitative differences between our results and the results of FLEX approximations given by equation (E.9) with regards to the "shadow bands" and explain why we believe that the failure of the FLEX to reproduce these bands is an artifact of that approximation. To avoid any confusion, we first clarify the terminology, because the term "shadow

bands" has been used previously to describe different physical effects (see for details Ref. [36]). We note that the so-called shadow features discussed in [36, 37] as well as the pseudogap in the total density of states $N(\omega) = (1/N)\sum_{\mathbf{k}} A(\mathbf{k},\omega)$)exist in both theories and we will not discuss them here. Instead, we concentrate on the precursors of antiferromagnetic bands in the spectral function $A(\mathbf{k}_{\mathrm{F}},\omega)$which correspond to two new solutions of the quasi-particle equation

$$\omega - \epsilon(\mathbf{k}) + \mu - \Sigma(\mathbf{k},\omega) = 0. \tag{75}$$

We start by recalling a simple physical argument why the precursors of antiferromagnetic bands must exist at finite temperatures in the vicinity of the zero-temperature phase transition in two dimensions. This can be best understood by contrasting this case with isotropic 3D case where such precursor effect are highly unlikely (for a discussion of the strongly anisotropic case see Sect. 5.4). Indeed, in three dimensions there is a *finite* temperature phase transition and the gap is equal to zero at this temperature $\Delta(T_{\mathrm{N}}) = 0$. Consequently at T_{N} there is only one peak in the $A(\mathbf{k}_{\mathrm{F}},\omega)$ at $\omega = 0$ which starts to split into two peaks only below T_{N}. Based on this simple physical picture, one would not expect to see precursors of antiferromagnetic bands above T_{N} in this case. The situation is qualitatively different in two-dimensions where classical thermal fluctuations suppress long-range order at any finite temperature while at the $T = 0$ phase transition the system goes directly into the ordered state with a finite gap. Clearly, the two peak structure in $A(\mathbf{k}_{\mathrm{F}},\omega)$ at $T = 0$ cannot disappear as soon as we raise the temperature.

For simplicity we again consider half-filling. As we have seen in Section 5.2 two new quasi-particle peaks do appear in the renormalized classical regime $T < T_{\mathrm{X}}$ in our theory. We have also found a pseudogap with the minimum at $\omega = 0$ in this regime. In contrast, the numerical solution of the FLEX equations [38] found a spectral function with a single maximum in $A(\mathbf{k}_{\mathrm{F}},\omega)$ at $\omega = 0$ even when $\tilde{\chi}^{\mathrm{RPA}}(\mathbf{q},0)$ becomes strongly peaked at $\mathbf{q} = \mathbf{Q}$. With decreasing temperature this central maximum becomes anomalously broad, but the two peak structure does not appear. The clear deviation from the Fermi liquid is signaled by the positive sign of $\partial\Sigma'(\mathbf{k}_{\mathrm{F}},\omega)/\partial\omega > 0$. However the value of $\partial\Sigma'(\mathbf{k}_{\mathrm{F}},\omega)/\partial\omega$ does not become larger than unity. The latter would unavoidably lead to the existence of two new quasi-particle peaks away from $\omega = 0$ as is clear from the graphical solution of the quasiparticle equation (Eq. (75)) shown on the bottom left panel of Figures 11.

We now explain analytically the origin of these qualitative differences in the two theories. In our theory $\partial\Sigma'(\mathbf{k}_{\mathrm{F}},\omega)/\partial\omega|_{\omega=0} \propto T\xi^2$ and hence it quickly becomes larger than unity in the renormalized classical regime $\xi \propto \exp(\mathrm{const}/T)$. In addition, for $\omega > v_{\mathrm{F}}\xi^{-1}$ the real part of the self-energy has the same behavior as in the ordered state $\Sigma(\mathbf{k}_{\mathrm{F}},\omega) \propto \Delta^2/\omega$ with $\Delta^2 \propto T\ln\xi = \mathrm{const}$. The important point is that this asymptotic behavior $\Sigma(\mathbf{k}_{\mathrm{F}},\omega) \propto \Delta^2/\omega$ of the self-energy already sets in for $\omega \sim \Delta \gg v_{\mathrm{F}}\xi^{-1}$. It is this property that leads to the appearance of the precursors of antiferromagnetic bands at $\omega = \pm\Delta$ in a manner analogous to the appearance of the Hubbard bands in the strong coupling limit that is discussed in the previous subsection. Let's now try to understand analytically what happens in the FLEX approximation. As in our theory, the main contribution to the self-energy in the strongly fluctuating regime comes from the zero-frequency term in the Matsubara sum in the equation for the self-energy (Eq. (54) in our theory and Eq. (E.9) in FLEX). An upper bound of the effect of the critical spin fluctuations can be obtained by approximating $T\tilde{\chi}^{\mathrm{RPA}}(\mathbf{q},0) \propto \delta(\mathbf{q})$. Then one immediately obtains the same expression for the self-energy as the one obtained in FLEX in the context of Hubbard bands (Eq. (71)). (The only difference is that the parameter Δ is now defined by the zero-frequency Matsubara contribution of $\tilde{\chi}^{\mathrm{RPA}}(\mathbf{q},0)$, rather than by the sum over all Matsubara frequencies.) As we have already discussed in the context of Hubbard bands, such a form for Σ does not lead to the appearance of two new quasiparticle solution away from $\omega = 0$ because the characteristic behavior $\Sigma(\mathbf{k}_{\mathrm{F}},\omega) \propto \Delta^2/\omega$ sets in too

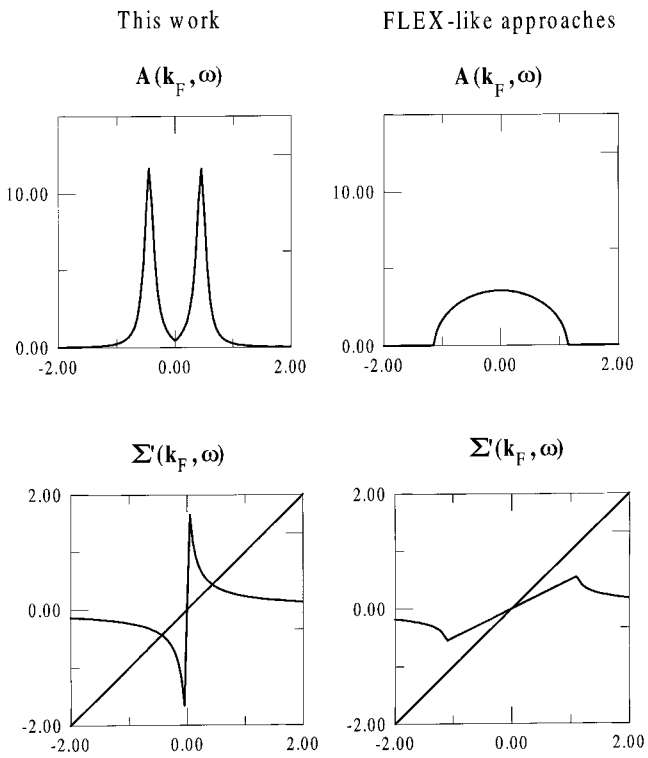

Fig. 11. — Top two panels are qualitative sketches of the spectral weight at the Fermi wave vector at half-filling. The plots are obtained by retainig only the classical contribution to the self-energy for $T < T_{\mathrm{X}}$ using parameters corresponding to the typical $U = 4$, of Monte Carlo simulations. The two bottom panels are the corresponding plots of $\mathrm{Re}\,\Sigma(\omega)$. The left-hand side of this figure is obtained using our approximation while the right-hand side is obtained from the FLEX-like approach. The intersection with the 45 degree line ω in the bottom-left panel gives rise to the precursors of antiferromagnetic bands seen right above it.

late, namely for $\omega \gg \Delta$. In addition, the slope of $\Sigma'(\mathbf{k}_{\mathrm{F}}, \omega)$ at $\omega \to 0$ does not diverge with decreasing temperature as in our theory but instead saturates to its value given by the analog of equation (73), *i.e.* $\partial\Sigma'(\mathbf{k}_{\mathrm{F}}, \omega)/\partial\omega < 1/2$. As we mentioned above, a value larger than unity $\partial\Sigma'(\mathbf{k}_{\mathrm{F}}, \omega)/\partial\omega > 1$ would guarantee the existence of two new solutions of the quasiparticle equation (Eq. (75)) away from $\omega = 0$. The right-hand side of Figure 11 illustrates clearly what happens in a FLEX-like approach such as equation (71). The contribution of classical fluctuations to the spectral weight does not lead to a Fermi liquid since $A(\mathbf{k}_{\mathrm{F}}, \omega)$ saturates to a finite width as temperature decreases, but nevertheless precursors of shadow bands do not occur because $\partial\Sigma'(\mathbf{k}_{\mathrm{F}}, \omega)/\partial\omega$ is bounded below unity. (Note that the spectral weight would not vanish so steeply at large frequencies if we had taken into account the quantum contribution of the spin fluctuations, as in full FLEX calculations.)

We just saw that the self-consistency in the propagators without corresponding self-consistency in the vertices inhibits the existence of the shadow bands in essentially the same way as it inhibits the existence of the Hubbard bands. It thus seems to us very likely that the absence of the precursors of antiferromagnetic bands below T_{X} in FLEX is an artifact. This conclusion

can be reliably verified by comparison with Monte Carlo data despite the fact that the latter is done for finite lattices and in the Matsubara formalism. This was discussed in more detail in Section 5.2. Here we just note that the temperature dependence of Matsubara quantities such as $G(\mathbf{k}_F, \tau = \beta/2)$ and $\Sigma(\mathbf{k}_F, ik_1)$ have a very characteristic form in the pseudogap regime. For example, $\Sigma(\mathbf{k}, ik_1) \propto 1/(i\pi T)$ in the pseudogap regime, while in FLEX we would expect a much weaker temperature dependence of this quantity (the upper bound being given by the analog of Eq. (72)).

We also would like to comment on the 1D model [75] which describes the interaction of electrons with static spin fluctuations characterized by the susceptibility $\chi_{sp} \propto \delta(\omega)[\xi^{-1}/(\mathbf{q} - \mathbf{Q})^2 + \xi^{-2}]$. The nice thing about this model is that it has an exact solution which shows the development of shadow bands and of the pseudogap in $A(\mathbf{k}_F, \omega)$. A treatment similar to ours which uses non-interacting Green's functions in the one-loop approximation also reproduces this feature [75]. However, the analogous approximation with dressed Green's functions leads to equation (71) and hence inhibits the existence of the "shadow bands" and of the pseudogap in $A(\mathbf{k}_F, \omega)$.

In closing we comment on semantics and on the physical interpretation of some results obtained in the FLEX approximation. The expression "conserving approximation" has been widely used to describe FLEX calculations of the single particle properties and, in particular, in the context of the shadow bands and of the failure of Luttinger's theorem [37, 38, 69]. The conserving aspect has been emphasized, but in fact the only desirable feature in the calculation of the single-particle properties is that the self-energy Σ is obtained from a functional derivative of the Luttinger-Ward functional $\Sigma = \delta\Phi/\delta G$ and hence it is guaranteed to satisfy Luttinger's theorem whenever appropriate. Only on the next level does this scheme lead to a calculation of the "true" susceptibilities [24] and of collective modes that satisfy conservation laws (Ward identities). However, these "true" susceptibilities are never substituted back in the calculation of the self-energy and the effect of "true" collective modes on the single-particle spectrum is an open question in FLEX. In fact, the RPA propagators $\tilde{\chi}_{RPA}$ appearing in the self-energy expression are different from susceptibilities from which collective modes should be computed and further they explicitly break conservation laws, as can be seen from the fact that RPA-like expressions $\tilde{\chi}_{RPA} = \tilde{\chi}_0/(1 - U\tilde{\chi}_0)$ with a dressed bubble $\tilde{\chi}_0$ have the unphysical properties that are mentioned in equations (A.23, A.24) of Appendix A. The fact that there are in effect two susceptibilities in the FLEX approximation leads, in our opinion, to some confusion and incorrect physical interpretation of the results in the literature. In particular, it was argued that the non-Fermi-liquid behavior and deviations from Luttinger theorem found in FLEX [37,38,69] are not due to critical thermal fluctuation in the vicinity of the phase transition but are rather the result of large U. The reasoning for such claim was that although the RPA susceptibilities $\tilde{\chi}_{RPA}$ is very strongly peaked at $\mathbf{q} = \mathbf{Q}$, the "true" FLEX susceptibility is not. In our opinion, such claim could be justified only if one would substitute the "true" susceptibility back in the calculation of Σ (for example using the exact Eq. (31)) and found that the deviation from the Luttinger theorem and other qualitative changes in $A(\mathbf{k}, \omega)$ increase with decreasing temperature without almost divergent behavior of the conserving susceptibility $\chi_{sp}(\mathbf{Q}, 0)$ and of the static structure factor $S_{sp}(\mathbf{Q})$.

The Monte Carlo data in Figure 7 are also instructive since they clearly show that qualitative changes in the single-particle spectra occur when the system enters the renormalized classical regime with rapidly growing $S_{sp}(\mathbf{Q})$. The fact that the FLEX "true" susceptibility does not show such behavior at half-filling [38] tells us that it even more drastically disagrees with the Monte Carlo data than the RPA-like $\tilde{\chi}$ which enters the expression for self-energy. Moreover, even away from half-filling the "true" susceptibility in FLEX at $\mathbf{q} = \mathbf{Q}$ significantly underestimates the strength of the spin fluctuations, as is clear from the comparison with Monte Carlo

data in Figure 5. In our opinion the, so-called, "true susceptibility" in FLEX is the key element in the confusion surrounding the interpretation of FLEX results for the self-energy because the "true susceptibility" never comes in the calculation of the self-energy. For all practical purposes these calculations of the self-energy should be considered as consistent with Luttinger's theorem at $T = 0$ but based on a non-conserving susceptibility. Consistency with conservation laws and consistency with Luttinger's theorem are not identical requirements because to satisfy rigorously Luttinger's theorem one needs that $\Sigma = \delta\Phi/\delta G$, while to have conserving susceptibilities one needs that the irreducible vertices used in Bethe-Salpeter (Eq. (26)) should be obtained from $\Gamma = \delta^2\Phi/\delta G\delta G$.

7. Domain of Validity of our Approach

Our approach is not valid beyond intermediate coupling. That is perhaps best illustrated by Figure 3 that shows that the crossover temperature first increases with U and then saturates instead of decreasing. The decrease is expected on general grounds from the fact that at strong coupling the tendency to antiferromagnetism should decrease roughly as $J \sim t^2/U$. The reason for this failure of our approach is clear. As we know from studies in infinite dimension [11], to account for strong-coupling effects it is necessary to include at least a frequency dependence to the self-energy and to the corresponding irreducible vertices.

Our theory also fails at half-filling deep in the renormalized classical regime, *i.e.* $T \ll T_X$ mainly for two reasons. First, the ansatz $U_{sp} = U g_{\uparrow\downarrow}(0)$, equation (40), fails in the sense that $g_{\uparrow\downarrow}(0)$ eventually reaches zero at $T \to 0$ because of the $\log^2 T$ divergence in the irreducible susceptibility $\chi_0(\pi, \pi)$ due to perfect nesting. The physically appropriate choice for $g_{\uparrow\downarrow}(0)$ in the renormalized classical regime is to keep its value fixed to its crossover-temperature value (See Fig. 6 and Sect. 4). The more serious reason why our approach fails for $T \ll T_X$ is that, as we just saw, critical fluctuations destroy completely the Fermi liquid quasiparticles and lead to a pseudogap. This invalidates our starting point. It is likely that in a more self-consistent theory, the logarithmic divergence of the appropriate irreducible susceptibility will be cutoff by the pseudogap. However, just a simple dressing of the Green's function is not the correct solution to the problem because it would make the theory non-conserving, as we discussed in Section A.3. One needs to take into account wave vector and energy dependent vertex corrections similar to those discussed by Schrieffer [76, 77].

8. Comparisons with other Approaches

In Appendix E, we discuss in detail various theories, pointing out limitations and advantages based on the criteria established in Appendices A.2 and A.3. More specifically, we include in our list of desirable properties, the local Pauli principle $\langle n_\uparrow^2 \rangle = \langle n_\uparrow \rangle$, the Mermin-Wagner theorem (Eq. (A.14)), the Ward identities (Eq. (A.28)), and f-sum rule (Eq. (A.22)), one-particle *versus* two-particle consistency $\Sigma_\sigma\left(1, \overline{1}\right) G_\sigma\left(\overline{1}, 1^+\right) = U\langle n_\uparrow n_\downarrow \rangle$ (Eq. (44)), Luttinger's theorem, and the large frequency asymptotic for the self-energy (Eq. (68)), which is important for the existence of the Hubbard bands. In the present section, we only state without proof where each theory has strengths and weaknesses.

In standard paramagnon theories [32, 46], the spin and charge fluctuations are computed by RPA, using either bare or dressed Green's functions. Then the fluctuations are feedback in the self-energy. When RPA with bare Green's functions are used for the collective modes, these satisfy the f-sum rule, but that is the only one of our requirements that is satisfied by such theories.

In conserving approximation schemes [24, 26] the Mermin-Wagner theorem, the Luttinger theorem and conservation laws are satisfied, but none of the other above requirements are fulfilled.

In the parquet approach [25, 53], one enforces complete antisymmetry of the four point function by writing down fully crossing-symmetric equations for these. However, in actual calculations, the local Pauli principle, the Mermin Wagner theorem, and the consistency between one and two particle properties are only approximately satisfied, while nothing enforces the other requirements.

In our approach, the high-frequency asymptotics and Luttinger's theorem are satisfied to a very good degree of approximation while all other properties in our list are exactly enforced. Let us specify the level of approximation. Luttinger's theorem is trivially satisfied with our initial approximation for the self-energy $\Sigma_\sigma^{(0)}$, but at the next level of approximation, $\Sigma_\sigma^{(1)}$, one needs a new chemical potential to keep the electron density $\mathrm{Tr}[G_\sigma^{(1)}(1, 1^+)]$ fixed. With this new chemical potential the Fermi surface volume is preserved to a very high accuracy. Finally, consider the high-frequency asymptotics. Since we use bare propagators, the high-frequency asymptotics comes in at the appropriate frequency scale, namely $ik_\mathrm{n} \sim W$, which is crucial for the existence of the Hubbard bands. However, the coefficient of the $1/ik_\mathrm{n}$ term in the high-frequency expansion (Eq. (68)) is incorrect because our irreducible vertices U_sp and U_ch are tuned to the low frequencies. If one would take into account the frequency dependence of U_sp and U_ch and assume that at high frequency they become equal to the bare interaction U, then one would recover the exact result, provided the Pauli principle in the form of equation (39) is satisfied. The difficulty with such a procedure is that frequency dependent irreducible vertices requires frequency dependent self-energy in the calculation of collective modes and that would make the theory much more complicated. Yet it is, probably, the only way to extend the theory to strong coupling.

9. Conclusion

We have presented a new simple approach [29,30] to the repulsive single-band Hubbard model. We have also critically compared competing approaches, such as paramagnon, fluctuation exchange approximation, and pseudo-potential parquet approaches. Our approach is applicable for arbitrary band structure [34] and gives us not only a quantitative description of the Hubbard model, but also provide us with some qualitatively new results. Let us summarize our theory again. We first obtain spin and charge fluctuations by a self-consistent parameterization of the two-particle effective interactions (irreducible vertices) that satisfies a number of exact constraints usually not fulfilled by standard diagrammatic approaches to the many-body problem. Then the influence of collective modes on single-particle properties is taken into account in such a way that single-particle properties are consistent with two-particle correlators, which describe these collective modes. More specifically, our approach satisfies the following constraints:

1. Spin and charge susceptibilities, through the fluctuation-dissipation theorem, satisfy the Pauli principle in the form $\langle n_\uparrow^2 \rangle = \langle n_\uparrow \rangle$ as well as the local moment sum-rule, conservations laws and consistency with the equations of motion in a local-field-like approximation.

2. In two dimensions, the spin fluctuations satisfy the Mermin-Wagner theorem.

3. The effect of collective modes on single-particle properties is obtained by a paramagnon-like formula that is consistent with the two-particle properties in the sense that

the potential energy obtained from $\mathrm{Tr}\,[\Sigma G]$ is identical to that obtained from applying the fluctuation-dissipation theorem to spin and charge susceptibilities.

4. Vertex corrections are included not only in spin and charge susceptibilities ($U_{\mathrm{sp}} \neq U_{\mathrm{ch}} \neq U$) but also in the self-energy formula. In the latter case, this takes into account the fact that there is no Migdal theorem controlling the effect of spin and charge fluctuations on the self-energy.

The results for both single-particle and two-particle properties are in *quantitative* agreement with Monte Carlo simulations for all fillings, as long as U is less than the bandwidth and T is not much smaller than the crossover temperature T_{X} where renormalized-classical behavior sets in. Both quantum-critical and renormalized-classical behavior can occur in certain parameter ranges but the critical behavior of our approach is that of the $O(n)$ model with $n \to \infty$ [33].

The main predictions of physical significance are as follows:

1. The theory predicts a magnetic phase diagram where magnetic order persists away from half-filling but with completely suppressed ferromagnetism.

2. In the renormalized classical regime above the zero-temperature phase transition, precursors of antiferromagnetic bands (shadow bands) appear in $A(\mathbf{k}_{\mathrm{F}}, \omega)$. These precursors occur when $\xi > \xi_{\mathrm{th}}$ (or $\omega_{\mathrm{SF}} < T$). Between these precursors of antiferromagnetic bands a pseudogap appears at half-filling, so that the Fermi liquid quasiparticles are completely destroyed in a wide temperature range above the zero-temperature phase transition $0 < T < T_{\mathrm{X}}$. The upper critical dimension for this phenomenon is three. We stress the qualitative difference between the Hubbard bands and the precursors of antiferromagnetic bands and we predict that in two dimensions one may see both sets of bands simultaneously in certain parameter ranges. This prediction is consistent with the results of numerical simulations [68,78]. We know of only one other analytic approach [70] which leads to similar four peak structure in the spectral function.

The zero temperature magnetic phase diagram is partly an open question because, despite the qualitative agreement with other analytical approaches, there is still an apparent contradiction with Monte Carlo simulations [64]. Our prediction of precursors of antiferromagnetic bands on the other hand is in agreement with Monte Carlo simulations. Neither this effect nor upper and lower Hubbard bands are observed in self-consistent schemes such as FLEX. This is because of inconsistent treatment of the vertex and self-energy corrections in this approximation, as we have explained in Section 6. However, if there was a Migdal theorem for spin fluctuations, it would be justifiable to neglect the vertex corrections and keep only the self-energy effects as is done in the FLEX approximation. The presence of precursors of antiferromagnetic bands in two-dimensions is then a clear case of qualitatively new Physics that would not appear if there was a Migdal theorem for spin fluctuations. The same is true for the Hubbard bands for large $U > W$ in any dimension.

We would like to state again clearly the nature of our critique of approximation schemes which are based on using Migdal's theorem for systems with electron-electron interactions. We do not imply that one does not need at all to take into account the feedback of the single-particle spectra on collective modes. The only point that we want to make here is that, based on sum rules and comparison with Monte Carlo data, we see that frequency and momentum dependent corrections to the self-energy and to the vertex often tend to cancel one another and that ignoring this leads to qualitatively incorrect results, in particular, with regards to the pseudogap. In this paper we were able to look only at the beginning of the renormalized

classical regime when the pseudogap starts to form. The truly self-consistent treatment of the one-particle and two-particle properties in the pseudogap regime remains an open and very challenging problem. We hope that by extending our approach to the ordered state and looking at how the pseudogap starts to disappear as the temperature is raised, one can better understand how to develop a more self-consistent theory in the pseudogap regime. We now point out how our approach can be extended in other directions.

As we mentioned in Section 5, the pseudogap and precursors of antiferromagnetic bands in the two-dimensional repulsive Hubbard model have interesting analogs in the attractive Hubbard model. In that model, one expects a pairing pseudogap and precursors of superconducting quasiparticle bands above T_c. At half-filling the negative and positive Hubbard models are mapped onto one another by a canonical transformation and the present theory is directly applicable to the attractive case. However, away from half-filling the mapping between the two models is more complicated and the microscopic theory requires additional sum-rule for pairing susceptibilities to find self-consistently the effective pairing interaction. This work is now in progress.

The present approach can be also extended to stronger coupling $U > W$. Again the key idea would be to parameterize the irreducible vertices, which have now to be frequency dependent, and then use the most important sum-rules to find the parameterization coefficients. This will, of course, require solving much more complicated self-consistent equations than in the present approach, but we believe that the problem still can be made tractable.

Finally, we would like to make two comments about the magnetic and the pairing pseudogap in the context of high-T_c superconductors, based on the results of our studies. First, as was stressed in reference [36], to understand clearly the physics of the single-particle pseudogap phenomena it is important to distinguish static short-range order from dynamical short-range order. The former is defined by a nearly Lorentzian form of the corresponding static structure factor $S(\mathbf{q}) \propto 1/((\mathbf{q}-\mathbf{Q})^2 + \xi^{-2})$ ($\mathbf{Q} = (\pi, \pi)$ in magnetic case, $\mathbf{Q} = 0$ in the case of pairing), while the latter means only that the corresponding susceptibility $\chi(\mathbf{q}, 0)$ has such a Lorentzian form. A condition for the existence of the single particle pseudogap in the vicinity of a given phase transition is that the corresponding short-range order is quasi-static (*i.e.* $\omega_{\mathrm{SF}} \ll T$) [36]. Experimentally, one can measure directly the dynamical spin structure factor $S(\mathbf{q}, \omega)$, and then obtain the static structure factor through the integral $S(\mathbf{q}) = \int S(\mathbf{q}, \omega) \mathrm{d}\omega/(2\pi)$. Even if the zero-frequency dynamical structure factor $S_{\mathrm{sp}}(\mathbf{q}, 0)$ is very strongly peaked at $\mathbf{q} \sim \mathbf{Q}$ it is possible that the static structure factor $S_{\mathrm{sp}}(\mathbf{q})$ is only weakly momentum dependent [36]. Thus in order to know whether one should expect to see the precursors of the antiferromagnetic bands and the corresponding pseudogap at a given doping and temperature it is necessary to obtain the static spin structure factor from the experimentally determined dynamical structure factor and then analyze its momentum dependence to see both if it is peaked and if it is quasi-two-dimensional.

The second comment that we would like to make is that both the pairing and the magnetic single-particle pseudogap discussed above are an effect of low dimensionality and hence they exist as long as there is a large two-dimensional fluctuating regime before the real three-dimensional phase transition. In this context, a pairing pseudogap could exist on either side of optimal doping [79]. The much larger temperature range over which a pseudogap appears in the underdoped compounds suggests that, in addition to pairing fluctuations, other thermal fluctuations (charge, spin...) prohibit finite-temperature ordering [80]. An example of this occurs in the attractive Hubbard model where charge fluctuations push the Kosterlitz-Thouless temperature to zero at half-filling, precisely where the crossover temperature to the pseudogap regime is largest.

Acknowledgments

We thank A.E. Ruckenstein, Liang Chen and R. Côté for helpful discussions and A. Georges for pointing out useful references. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the "Fonds pour la formation de chercheurs et l'aide à la recherche" from the Government of Québec (FCAR), and (for Y.M.V.) the National Science Foundation (Grant No. NSF-DMR-91-20000) through the Science and Technology Center for Superconductivity and (for A.-M.S.T.) the Killam Foundation as well as the Canadian Institute for Advanced Research (CIAR). A.-M.S.T. would like to thank the Institute for Theoretical Physics, Santa Barbara for hospitality during the writing of this work. Partial support there was provided by the National Science Foundation under Grant No. PHY94-07194.

Appendix A

Sum Rules, Ward Identities and Consistency Requirements

In this appendix, we recall well known constraints on many-body theory that follow from sum-rules and conservation laws and comment, wherever possible, on their physical meaning and on where commonly used approaches fail to satisfy these constraints. Although we come back on a detailed discussion of various theories in a later appendix, we find it useful to include some of this discussion here to motivate our approach. We consider in turn various results that would be satisfied by any exact solution of the many-body problem. They are all consequences of either anticommutation relations alone (Pauli principle) or of anticommutation relations and the Heisenberg equations of motion. We describe in turn: 1) the relation between self-energy and two-body correlation functions that embodies the details of the Hamiltonian; 2) sum rules for one-particle properties; 3) sum rules and constraints on two-particle properties, in particular *f*-sum rule and Ward identities that express conservation laws; 4) a few relations that are crucial in Fermi liquid theory, namely Luttinger's theorem and the forward scattering sum rule.

A.1. Equations of Motion and the Relation between the Self-Energy Σ and Two-Particle Properties. — The self-energy (we always mean one-particle irreducible self-energy) is related to the potential energy, and hence to two-particle correlations through the expression equation (44), which in the Kadanoff and Baym notation can be written as

$$\Sigma_\sigma\left(1,\overline{1}\right)G_\sigma\left(\overline{1},1^+\right)=U\left\langle n_\uparrow n_\downarrow\right\rangle\cdot \tag{A.1}$$

Here, the index with an overbar, $\overline{1}$, means that there is a sum over corresponding lattice positions and an integral over imaginary time. The notation 1^+ means that the imaginary time implicit in 1 is $\tau_1+\eta$ where η is a positive infinitesimal number. Equation (A.1) is an important consistency requirement between self-energy and double occupancy in the Hubbard model that can easily be proven as follows. From the equations of motion for the single-particle Green's function (Eq. (3)) one finds

$$\begin{aligned}&\left[\left(-\frac{\partial}{\partial\tau_i}+\mu\right)\delta_{i,\ell}+t_{i\ell}\right]G_\sigma\left(\mathbf{r}_\ell-\mathbf{r}_j,\tau_i-\tau_j\right)\\&\quad=\delta_{i,j}\delta\left(\tau_i-\tau_j\right)-U\left\langle T_\tau\left(c^\dagger_{i-\sigma}\left(\tau_i\right)c_{i-\sigma}\left(\tau_i\right)c_{i\sigma}\left(\tau_i\right)c^\dagger_{j\sigma}\left(\tau_j\right)\right)\right\rangle\cdot\end{aligned} \tag{A.2}$$

Using the short-hand notation in equations (3, 4) and the definition of self-energy (Dyson's equation) the above equation is also written in the form,

$$G_0^{-1}(1,\overline{1})G_\sigma(\overline{1},2) = \delta\,(1-2) + \Sigma_\sigma(1,\overline{1})G_\sigma(\overline{1},2). \tag{A.3}$$

Comparing the last two equations, the well known relation equation (A.1) (or Eq. (44)) between self-energy, Green's function and potential energy follows.

So-called conserving [26] approaches to the many-body problem violate the above consistency requirement (Eq. (44)) in the following sense. The right-hand side can be computed from the collective modes using the fluctuation-dissipation theorem. In conserving approximations, this gives a result that is different from what is computed directly from the left-hand side of the equation, namely from the self-energy and from the Green's function. In fact, all many-body approaches satisfy the above consistency requirement at best in an approximate way. However, it is a very important requirement and equation (44) plays a key role in our discussion. Seen in Matsubara frequency, it is a sum rule, or an integral constraint that involves all frequencies, large and small.

A.2. Constraints on Single-Particle Properties. — The spectral weight $A_\sigma\,(\mathbf{k},\omega)$ can be interpreted as a probability of having an electron in a state $(\sigma,\mathbf{k},\omega)$ and it satisfies the normalization sum rule

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}A_\sigma(\mathbf{k},\omega) = \left\langle\left\{c_{\mathbf{k}\sigma},c^\dagger_{\mathbf{k}\sigma}\right\}\right\rangle = 1. \tag{A.4}$$

Formally this is a consequence of the jump in the Green's function at $\tau = 0$, as can be seen from calculating

$$\begin{aligned} G_\sigma(\mathbf{k},0^-) - G_\sigma(\mathbf{k},0^+) = 1 &= T\sum_{ik_n}\left(\mathrm{e}^{ik_\mathrm{n}\eta} - \mathrm{e}^{-ik_\mathrm{n}\eta}\right)\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\frac{A_\sigma(\mathbf{k},\omega)}{ik_\mathrm{n}-\omega} \\ &= \int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}A_\sigma(\mathbf{k},\omega). \end{aligned} \tag{A.5}$$

To do perturbation theory directly for the Green's function to any finite order would require that the interaction U be small not only in comparison with the bandwidth W but also in comparison with the smallest Matsubara frequency $ik_1 = 2\pi T$. Also, the direct perturbation series for the Green's function gives, after analytical continuation, poles of arbitrary high order located at the unperturbed energies. These high-order poles are inconsistent with the simple pole (or branch cut) structure of the Green's function predicted by the spectral representation. Furthermore, the high-order poles lead to a spectral weight that can be negative [81]. The common way to get around these difficulties is to make approximations for the self-energy Σ instead and then calculate the Green's function using Dyson's equation (Eq. (8)).

It is interesting to note that to satisfy the constraint equation (A.4), it suffices that $\Sigma(\mathbf{k},ik_\mathrm{n})$, defined by equation (8), has a finite limit as $ik_\mathrm{n}\to\infty$. More constraints on approximations for the self-energy may be found by continuing this line of thought. A systematic way of doing this is to do a high-frequency expansion for both the Matsubara Green's function and the self-energy and to find coefficients using sum-rules. The sum-rules that we need then are [82]

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\omega A_\sigma(\mathbf{k},\omega) = \left\langle\left\{\left[c_{\mathbf{k}\sigma},(H-\mu N)\right],c^\dagger_{\mathbf{k}\sigma}\right\}\right\rangle = \epsilon_\mathbf{k} - \mu + Un_{-\sigma} \tag{A.6}$$

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}\omega}{2\pi}\omega^2 A_\sigma(\mathbf{k},\omega) = (\epsilon_\mathbf{k}-\mu)^2 + 2U(\epsilon_\mathbf{k}-\mu)n_{-\sigma} + U^2 n_{-\sigma} \tag{A.7}$$

where $n_\sigma = n/2$ since we are in the paramagnetic state.

Using the spectral representation (Eq. (6)) one can easily see that the above sum rules give the coefficients of the high-frequency expansion of the Matsubara Green's function

$$\lim_{ik_n\to\infty} G_\sigma(\mathbf{k}, ik_n) = \frac{1}{ik_n} + \left(\frac{1}{ik_n}\right)^2 \int \frac{d\omega}{2\pi}\omega A_\sigma(\mathbf{k},\omega) + \left(\frac{1}{ik_n}\right)^3 \int \frac{d\omega}{2\pi}\omega^2 A_\sigma(\mathbf{k},\omega) + \cdots \quad \text{(A.8)}$$

The self-energy has the same analytic properties as the Green's function. Using its high frequency expansion in the expression for the Green's function (Eq. (8)), one finds that the first term in equation (A.8), leads to the requirement that the self-energy has a finite limit at $ik_n \to \infty$. The second term fixes the value of this constant to the Hartree-Fock result, and the last and second-term combine to give the leading term in $1/ik_n$ of the self-energy high-frequency expansion. In short, we find the result quoted in equation (68), namely

$$\lim_{ik_n\to\infty} \Sigma_\sigma(\mathbf{k}, ik_n) = Un_{-\sigma} + \frac{U^2 n_{-\sigma}(1-n_{-\sigma})}{ik_n} + \cdots \quad \text{(A.9)}$$

The Kramers-Kronig relation for the self-energy

$$\mathrm{Re}\left[\Sigma^R_\sigma(\mathbf{k},\omega) - \Sigma^R_\sigma(\mathbf{k},\infty)\right] = \mathcal{P}\int \frac{d\omega'}{\pi} \frac{\mathrm{Im}\left[\Sigma^R_\sigma(\mathbf{k},\omega')\right]}{\omega' - \omega}$$

and the high-frequency result (Eq. (A.9)) imply the following sum-rule for the imaginary part of the self-energy

$$-\int \frac{d\omega'}{\pi} \mathrm{Im}\left[\Sigma^R_\sigma(\mathbf{k},\omega')\right] = U^2 n_{-\sigma}(1-n_{-\sigma}).$$

Important consequences of this equation are that for a given U the integrated imaginary part of the self-energy is independent of temperature and is increasing towards half-filling. The right-hand side of this equation is also a measure of the width of the single-particle excitation spectrum, as can be seen from the spectral weight moments (Eqs. (A.6, A.7),

$$\overline{\omega^2} - \overline{\omega}^2 \equiv \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\omega^2 A_\sigma(\mathbf{k},\omega) - \left[\int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\omega A_\sigma(\mathbf{k},\omega)\right]^2 = U^2 n_{-\sigma}(1-n_{-\sigma}).$$

An important physical point is that the asymptotic behavior (Eq. (A.9)) is a necessary condition for the existence of upper and lower Hubbard bands, as has been explained in Section 6.1. However, it is important to realize that it is not a sufficient condition. Indeed, the following paradox has been noticed in explicit calculations in infinite dimensions [11,74]. While ordinary second-order perturbation theory with bare Green functions G_0 reproduces correctly the appearance of the Hubbard bands with increasing U, the perturbation theory with dressed Green function $G = [G_0^{-1} - \Sigma]^{-1}$ does not. The reason for this is that although the second-order expression for $\Sigma_\sigma(\mathbf{k}, ik_n)$ in terms of full G does satisfy the asymptotics (Eq. (A.9)), it sets in too late, namely for $k_n \gg U$, instead of $k_n \gg W$. The fact that the asymptotics should start at $k_n \sim W$ even when $U > W$ is a non-trivial consequence of the Pauli principle, as explained in Section 6.1. Thus there are no Hubbard bands in any theory that uses self-consistent Green functions but neglects the frequency dependence of the vertex. This is an explicit example that illustrates what seems to be a more general phenomena: a calculation with dressed Green's functions but no frequency dependent vertex correction often gives worse results that the one done with bare Green's functions and a frequency independent vertex. We will see in the next

subsection that this also happens in the calculation of the two-particle properties. Also, as we have argued in Section 6, a similar situation occurs with the precursors of antiferromagnetic bands in the renormalized classical regime in two-dimensions.

Finally, we quote two more well known sum-rules that we will need. They involve the Fermi function $f(\omega)$ and the spectral weight. The first one follows from definition of $G_\sigma(\mathbf{k},\tau)$ and the spectral representation

$$\lim_{\tau\to 0^-} G_\sigma(\mathbf{k},\tau) = \int \frac{\mathrm{d}\omega}{2\pi} f(\omega) A_\sigma(\mathbf{k},\omega) = \left\langle c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} \right\rangle \equiv n_{\mathbf{k}\sigma}. \tag{A.10}$$

The quantity $n_{\mathbf{k}\sigma}$ is the distribution function. It is equal to the Fermi function only when the self-energy is frequency independent. The next result, that follows simply from the equations of motion,

$$\begin{aligned} \lim_{\tau\to 0^-} -\frac{1}{N}\sum_{\mathbf{k}} \frac{\partial G_\sigma(\mathbf{k},\tau)}{\partial\tau} &= \frac{1}{N}\sum_{\mathbf{k}} \int \frac{\mathrm{d}\omega}{2\pi} \omega f(\omega) A_\sigma(\mathbf{k},\omega) \\ &= \frac{1}{N}\sum_{\mathbf{k}} (\epsilon_{\mathbf{k}} - \mu) n_{\mathbf{k}\sigma} + U\langle n_\uparrow n_\downarrow\rangle \end{aligned} \tag{A.11}$$

is useful to show to what extent certain dressed-propagator approaches fail to satisfy the f-sum rule.

A.3. Constraints on Two-Particle Properties. — For any one-band model, independently of the Hamiltonian, the Pauli principle (anticommutation relations)

$$\left\langle n_{i\sigma}^2 \right\rangle = \langle n_{i\sigma}\rangle \tag{A.12}$$

implies the following two simple identities:

$$\left\langle (n_{i\uparrow} \pm n_{i\downarrow})^2 \right\rangle = n \pm 2\langle n_{i\uparrow} n_{i\downarrow}\rangle \cdot \tag{A.13}$$

The correlation functions on the left-hand side are equal-time and equal-position spin and charge correlation functions. The susceptibilities $\chi_{\mathrm{ch}}(\mathbf{r}_i - \mathbf{r}_j, \tau)$, $\chi_{\mathrm{sp}}(\mathbf{r}_i - \mathbf{r}_j, \tau)$ in equations (17, 16) are response functions for arbitrary $(\mathbf{r}_i - \mathbf{r}_j, \tau)$ so they must reduce to the above equal-time equal-position correlation functions when $\mathbf{r}_i = \mathbf{r}_j$ and $\tau = 0$. This is one special case of the imaginary-time version of the fluctuation-dissipation theorem (Eqs. (16, 17)). This translates into local-moment and local-charge sum-rules for the susceptibilities

$$\frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\mathrm{n}}} \chi_{\mathrm{sp}}(\mathbf{q}, iq_{\mathrm{n}}) = 2\langle n_\uparrow n_\uparrow\rangle - 2\langle n_\uparrow n_\downarrow\rangle = n - 2\langle n_\uparrow n_\downarrow\rangle \tag{A.14}$$

$$\frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\mathrm{n}}} \chi_{\mathrm{ch}}(\mathbf{q}, iq_{\mathrm{n}}) = 2\langle n_\uparrow n_\uparrow\rangle + 2\langle n_\uparrow n_\downarrow\rangle - n^2 = n + 2\langle n_\uparrow n_\downarrow\rangle - n^2 \tag{A.15}$$

where we have removed the i dependence of $\langle n_{i\uparrow} n_{i\downarrow}\rangle$ using translational invariance. The right-hand side of the local-moment sum-rule is equal to $\langle (S^z)^2\rangle$,while that of the local-charge sum rule is equal to $\left\langle \rho^2 \right\rangle - n^2$.

If arbitrary sets of diagrams are summed, nothing can prevent the right-hand side from taking unphysical values. For example, the Pauli principle may be violated, *i.e.* $\langle n_\uparrow n_\uparrow\rangle \neq \langle n_\uparrow\rangle$.

To see this, notice that when the Pauli principle is satisfied, our two sum rules equations (A.14, A.15) lead to

$$\frac{T}{N}\sum_{\mathbf{q}}\sum_{iq_{\mathrm{n}}}\left[\chi_{\mathrm{sp}}\left(\mathbf{q},iq_{\mathrm{n}}\right)+\chi_{\mathrm{ch}}\left(\mathbf{q},iq_{\mathrm{n}}\right)\right]=2n-n^{2}. \tag{A.16}$$

It is easy to check that well known approaches to the many-body problem, such as RPA, violate this basic requirement. Indeed, the ordinary RPA expressions for spin and charge are

$$\chi_{\mathrm{sp}}^{\mathrm{RPA}}(q)\equiv\frac{\chi_{0}}{1-\frac{U}{2}\chi_{0}} \tag{A.17}$$

$$\chi_{\mathrm{ch}}^{\mathrm{RPA}}(q)\equiv\frac{\chi_{0}}{1+\frac{U}{2}\chi_{0}} \tag{A.18}$$

where

$$\chi_{0}\left(q\right)=-2\frac{T}{N}\sum_{k}G^{(0)}\left(k\right)G^{(0)}\left(k+q\right). \tag{A.19}$$

That RPA does not satisfy the sum rule (Eq. (A.16)) already to second order in U can be easily seen by expanding the denominators.

To satisfy the Mermin-Wagner theorem, approximate theories must also prevent $\langle n_{\uparrow}n_{\downarrow}\rangle$ from taking unphysical values. This quantity is positive and bounded by its value for $U=\infty$ and its value for non-interacting systems, namely $0\leq\langle n_{\uparrow}n_{\downarrow}\rangle\leq n^{2}/4$. Hence, the right-hand side of the local-moment sum-rule (Eq. (A.14)) is contained in the interval $\left[n,n-\frac{1}{2}n^{2}\right]$. Any theory that prevents the right-hand side of the local-moment sum rule from taking infinite values satisfies the Mermin-Wagner theorem.

Proof: Near a magnetic phase transition, the zero Matsubara-frequency component of the spin susceptibility takes the Ornstein-Zernicke form

$$\chi_{\mathrm{sp}}\left(\mathbf{q}+\mathbf{Q},0\right)\sim\frac{1}{q^{2}+\xi^{-2}} \tag{A.20}$$

where q is measured with respect to the ordering wave vector $\mathbf{Q}$ and where ξ^{2} is the square of the correlation length. Near its maximum, the above susceptibility is of order ξ^{2} while all finite Matsubara-frequency components at the ordering wave vector are at most of order $1/\left(2\pi T\right)^{2}$ which is much smaller than ξ^{2}. Hence, one can keep only the zero-Matsubara frequency contribution on the left-hand side of the local-moment sum rule (Eq. (A.14)) obtaining

$$T\int\frac{\mathrm{d}^{d}\mathbf{q}}{\left(2\pi\right)^{d}}\frac{1}{q^{2}+\xi^{-2}}=\widetilde{C} \tag{A.21}$$

where $\widetilde{C}$ contains non-zero Matsubara frequency contributions as well as $n-2\left\langle n_{\uparrow}n_{\downarrow}\right\rangle$. Since $\widetilde{C}$ is finite, this means that in two dimensions ($d=2$), it is impossible to have $\xi^{-2}=0$ on the left-hand side otherwise the integral would diverge logarithmically.

Finally, the f-sum rule on spin and charge susceptibilities follows as usual from the fact that the Hamiltonian conserves particle number. Computing $\left\langle\left[\rho_{\mathbf{q}},\frac{\partial\rho_{-\mathbf{q}}}{\partial\tau}\right]\right\rangle\Big|_{\tau=0}$ and $\left\langle\left[S_{\mathbf{q}},\frac{\partial S_{-\mathbf{q}}}{\partial\tau}\right]\right\rangle\Big|_{\tau=0}$ one obtains for either charge or spin

$$\begin{aligned}\int\frac{\mathrm{d}\omega}{\pi}\omega\chi_{\mathrm{ch,sp}}''\left(\mathbf{q},\omega\right) &= \lim_{\eta\to0}T\sum_{iq_{\mathrm{n}}}\left(\mathrm{e}^{-iq_{\mathrm{n}}\eta}-\mathrm{e}^{iq_{\mathrm{n}}\eta}\right)iq_{\mathrm{n}}\chi_{\mathrm{ch,sp}}\left(\mathbf{q},iq_{\mathrm{n}}\right)\\ &= \frac{1}{N}\sum_{\mathbf{k}\sigma}\left(\epsilon_{\mathbf{k}+\mathbf{q}}+\epsilon_{\mathbf{k}-\mathbf{q}}-2\epsilon_{\mathbf{k}}\right)n_{\mathbf{k}\sigma}.\end{aligned} \tag{A.22}$$

As can be seen from the spectral representations of spin and charge susceptibilities, equation (20), the quantity that obeys the f-sum rule is the coefficient of the leading term in the $1/q_{\rm n}^2$ high-frequency expansion of the susceptibilities.

The single-particle energies $\epsilon_{\mathbf{k}}$ entering explicitly the right-hand side of the f-sum rule are independent of interactions, so interactions influence the f-sum rule only very weakly through the $n_{\mathbf{k}\sigma}$. In fact, in a continuum $\epsilon_{\mathbf{k}} \propto \mathbf{k}^2$ so $n_{\mathbf{k}\sigma}$ enters only in the form $\sum_{\mathbf{k}\sigma} n_{\mathbf{k}\sigma} = n$. In this case, the right-hand side of the f-sum rule is proportional to $q^2 n$ and hence is independent of interactions. On a lattice however, the energies cannot in general be taken out of the sum and interactions influence the value of the right-hand side, but only through the fact that $n_{\mathbf{k}\sigma}$ differs from the non-interacting Fermi function $f_{\mathbf{k}\sigma}$. At strong-coupling, where the self-energy is strongly frequency dependent, this difference between $n_{\mathbf{k}\sigma}$ and $f_{\mathbf{k}\sigma}$ becomes important. But from weak to intermediate coupling, calculations where $f_{\mathbf{k}\sigma}$ appears on the right-hand side should be good approximations. In the explicit examples that we have treated, the U dependence of the f-sum rule becomes important only close to half-filling and for $U > 4$, signaling the breakdown of approximations based on frequency-independent self-energies.

While RPA-like theories that use $f_{\mathbf{k}\sigma}$ instead of $n_{\mathbf{k}\sigma}$ violate only weakly the f-sum rule in the weak to intermediate coupling regime, self-consistent theories that use frequency-dependent self-energies but no frequency-dependent vertices violate conservations laws in general, and the f-sum rule in particular, in a much more dramatic way. The point is that susceptibilities with a dressed bubble, $\tilde{\chi}_{\rm RPA} = \tilde{\chi}_0/(1 - \frac{1}{2}U\tilde{\chi}_0)$, are bad approximations because they have the following properties, for any value of U

$$\tilde{\chi}_{\rm RPA}(\mathbf{q} = 0, iq_{\rm n} \neq 0) \neq 0 \tag{A.23}$$

$$\int \frac{{\rm d}\omega}{2\pi}\omega\tilde{\chi}''_{\rm RPA}(\mathbf{q},\omega) = \frac{1}{N}\sum_{\mathbf{k},\sigma}\left(\epsilon_{\mathbf{k}+\mathbf{q}} + \epsilon_{\mathbf{k}-\mathbf{q}} - 2\epsilon_{\mathbf{k}}\right) n_{\mathbf{k}\sigma} + 4U\left(\langle n_\uparrow\rangle\langle n_\downarrow\rangle - \langle n_\uparrow n_\downarrow\rangle\right). \tag{A.24}$$

The first of these equations explicitly violates the Ward identity, equation (A.28) below, at all frequencies, including small non-zero ones, since at zero wave vector we should have $\chi(\mathbf{q} = 0, iq_{\rm n} \neq 0) = 0$ for all frequencies except zero. The second equation (Eq. (A.24)) violates the f-sum rule (Eq. (A.22)) at all wave vectors, by a constant term $4U\left(\langle n_\uparrow\rangle\langle n_\downarrow\rangle - \langle n_\uparrow n_\downarrow\rangle\right)$ which in practical calculations, say at $U = 4$, is of the same order as the first term, which is the only one that should be there according to the f-sum rule.

Proof: Equations (A.23, A.24) are proven as follows. Consider the standard RPA expression but with dressed bubbles $\tilde{\chi}_0$

$$\tilde{\chi}_{\rm RPA} = \tilde{\chi}_0/(1 - \frac{U}{2}\tilde{\chi}_0). \tag{A.25}$$

Using the spectral representation for the Green's function and inversion symmetry in the Brillouin zone one finds

$$\tilde{\chi}_0(\mathbf{q}, iq_{\rm n}) = \frac{2}{N}\sum_{\mathbf{k}}\int\frac{{\rm d}\omega}{2\pi}\int\frac{{\rm d}\omega'}{2\pi}A(\mathbf{k},\omega)A(\mathbf{k}+\mathbf{q},\omega')\frac{(\omega-\omega')\left(f(\omega') - f(\omega)\right)}{(\omega-\omega')^2 + q_{\rm n}^2}\cdot \tag{A.26}$$

When the bubble is not dressed, the spectral weights are delta functions so that at $\mathbf{q} = 0$ the susceptibility would vanish for all non-zero values of $q_{\rm n}$, as required by the Ward identity. However, here because the spectral weight has a width and because the integrand is even and positive, then the integral will not vanish, resulting in the first anomaly (Eq. (A.23)) we mention. To prove the second equation (Eq. (A.24)), it suffices

to remember from the spectral representation of the susceptibility (Eq. (20)) and the derivation of the f-sum rule (Eqs. (A.22)) that we are looking for the coefficient of the $1/q_{\rm n}^2$ term in the high-frequency expansion. Given the RPA form (Eq. (A.25)), only the numerator contributes to this limit. One obtains, for the coefficient of the $1/q_{\rm n}^2$ term,

$$\frac{2}{N}\sum_{\mathbf{k}}\int\frac{\mathrm{d}\omega}{2\pi}\int\frac{\mathrm{d}\omega'}{2\pi}A(\mathbf{k},\omega)A(\mathbf{k}+\mathbf{q},\omega')\left(\omega-\omega'\right)\left(f\left(\omega'\right)-f\left(\omega\right)\right) \tag{A.27}$$

from which equation (A.24) follows using the sum rules for occupation number (Eq. (A.10)) and for energy (Eq. (A.11)).

Conservation laws have general consequences not only on equal-time correlation functions, as in the f-sum rule above, but also on time-dependent correlation functions. For example, from the Heisenberg equations of motion and anti-commutation relations, follow the Ward identities [45]

$$\sum_{\mathbf{k}}\sum_{\sigma=\pm1}\sum_{\sigma'=\pm1}\left(\frac{\partial}{\partial\tau}+\left(\epsilon_{\mathbf{k}+\mathbf{q}}-\epsilon_{\mathbf{k}}\right)\right)\left\langle T_\tau c^\dagger_{\mathbf{k}\sigma}\left(\tau\right)\sigma^\ell c_{\mathbf{k}+\mathbf{q}\sigma}\left(\tau\right)c^\dagger_{\mathbf{k}'+\mathbf{q}\sigma'}\left(\tau_1\right)\sigma'^\ell c_{\mathbf{k}'\sigma'}\left(\tau_2\right)\right\rangle$$

$$=\delta\left(\tau-\tau_1\right)\sum_{\sigma'=\pm1}\sigma'^\ell G_{\sigma'}\left(\mathbf{k}',\tau_2-\tau\right)-\delta\left(\tau-\tau_2\right)\sum_{\sigma'=\pm1}\sigma'^\ell G_{\sigma'}\left(\mathbf{k}'+\mathbf{q},\tau-\tau_1\right) \tag{A.28}$$

where $\ell=0$ for charge, and $\ell=1$ for spin. The f-sum rule above (Eq. (A.22)) follows from the above identity by simply taking $\tau_1=\tau_2^+$, summing over $\mathbf{k}'$ and subtracting the two results for $\tau\to\tau_1^+$ and $\tau\to\tau_1^-$.

We have seen in this section that there are strong cancelations for two-particle properties between the frequency dependence of self-energy and that of the vertex corrections, so that putting a frequency dependence in only one of them is a bad approximation. We have adopted the Kadanoff-Baym formalism in the main text since it can be used as a guide to make approximations that satisfy conservation laws.

A.4. When there is a Fermi Surface. — When perturbation theory converges (no phase transition) then at zero temperature $T=0$ the imaginary part of the self-energy vanishes, $\Sigma''_\sigma\left(\mathbf{k},\omega=0\right)=0$, for all $\mathbf{k}$ values and the Fermi surface defined by

$$\epsilon_{\mathbf{k}}-\mu-\Sigma'_\sigma\left(\mathbf{k},\omega=0\right)=0 \tag{A.29}$$

encloses a volume that is equal to the volume enclosed by non-interacting particles

$$\frac{1}{N}\sum_{\mathbf{k}}\theta\left(\mu-\epsilon_{\mathbf{k}}-\Sigma'_\sigma\left(\mathbf{k},0\right)\right)=\frac{1}{N}\sum_{\mathbf{k}}\theta\left(\mu_0-\epsilon_{\mathbf{k}}\right)=n_\sigma. \tag{A.30}$$

This is the content of Luttinger's theorem [28,83]. It implies that there is a strong cancelation between the change of the chemical potential and the change of the self-energy on the Fermi surface. In particular, when $\Sigma'_\sigma\left(\mathbf{k}_{\rm F},0\right)$ does not depend on $\mathbf{k}$ or on the direction of $\mathbf{k}_{\rm F}$ (infinite D Hubbard model, electron gas) the change in $(\mu-\mu_0)$ is exactly canceled by $\Sigma'_\sigma\left(\mathbf{k}_{\rm F},0\right)$

$$\mu-\mu_0=\Sigma'_\sigma\left(\mathbf{k}_{\rm F},0\right). \tag{A.31}$$

Luttinger's theorem is satisfied when

$$\lim_{T\to0}\int\frac{\partial\Sigma_\sigma(\mathbf{k},i\nu)}{\partial(i\nu)}G_\sigma(\mathbf{k},i\nu)\mathrm{d}\nu\mathrm{d}\mathbf{k}=0. \tag{A.32}$$

Any theory that calculates its self-energy from a functional derivative of the Luttinger-Ward functional $\Sigma = \delta\Phi[G]/\delta G$ will satisfy Luttinger's theorem [28, 83]. The latter procedure requires self-consistent determination of the self-energy as a function of momentum and frequency $\Sigma_\sigma(\mathbf{k}, ik_\mathrm{n})$ and is usually quite computationally involved. However, even when this procedure to calculate the self-energy is not followed, it turns out to be rather easy to satisfy this theorem to an excellent degree of approximation in the weak to intermediate coupling regime. The reason for this is that any frequency-independent self-energy will preserve Luttinger's theorem and weak frequency dependence will not cause great harm. For the electron gas, Luttinger [28] suggests a way to build a perturbation theory in terms of non-interacting Green's functions which allows to satisfy Luttinger's theorem to very good accuracy. The trick is that the chemical potential for the interacting electrons μ should always enter the calculations in the form of the difference with the shift of the self-energy on the Fermi surface $\tilde{G}_0 = 1/[ik_\mathrm{n} - \epsilon_\mathbf{k} + (\mu - \Sigma'_\sigma(\mathbf{k}_\mathrm{F}, 0))]$. The "non-interacting" Green's function $\tilde{G}_0$ in this formalism is the Green's function of some effective non-interacting system and, in general, it is different from both $1/(ik_\mathrm{n} - \epsilon_\mathbf{k} + \mu)$ and $1/(ik_\mathrm{n} - \epsilon_\mathbf{k} + \mu_0)$. However, when $T \to 0$ Luttinger's theorem requires that $(\mu - \Sigma'_\sigma(\mathbf{k}_\mathrm{F}, 0)) \to \mu_0$ and one can approximate $\tilde{G}_0$ by the Green's function for a non-interacting system of the same density $G_0 = 1/(ik_\mathrm{n} - \epsilon_\mathbf{k} + \mu_0)$. In practice, one can also have a phase transition (or crossover) at a finite temperature T_c (T_X). In these cases Luttinger's theorem is satisfied only approximately since the zero-temperature limit cannot be reached without a breakdown of perturbation theory. Then the relevant question is how well it is satisfied at T_c (T_X) (see also Sect. 3.2.2 for a discussion of Luttinger's theorem in our approach).

When Luttinger's theorem holds, one can usually develop a Landau Fermi liquid theory. In this approach, the Pauli principle is implemented only for momentum states near the Fermi surface by imposing the forward scattering sum rule. This sum rule, in two dimensions, reads

$$\sum_\ell \left[\frac{F_\ell^\mathrm{s}}{1 + F_\ell^\mathrm{s}} + \frac{F_\ell^\mathrm{a}}{1 + F_\ell^\mathrm{a}} \right] = 0 \tag{A.33}$$

where F_ℓ^s and F_ℓ^a are the symmetric and antisymmetric Landau parameters expanded on the $\mathrm{e}^{-i\theta\ell}$ basis instead of the Legendre polynomial basis. Recent renormalization group analysis has however claimed [84] that the forward scattering sum rule comes from an inaccurate use of crossing symmetry and is not the proper way to enforce the Pauli principle. Most approaches to the many-body problem disregard this sum rule anyway, in the same way that they disregard the local Pauli principle.

Appendix B

Proofs of Various Formal Results

In this appendix, we give the proofs of various relations mentioned in Sections 3 and 3.2.3.

1. The general expression for the self-energy (Eq. (27)) can be obtained as follows. Use the equations of motion and the definition of the self-energy (Eqs. (A.2, A.3)) which in the present notation give

$$\Sigma_\sigma(1, \overline{1}) G_\sigma(\overline{1}, 2) = -U \left\langle T_\tau \left[\psi_{-\sigma}^+(1^{++}) \psi_{-\sigma}(1^+) \psi_\sigma(1) \psi_\sigma^+(2) \right] \right\rangle \tag{B.1}$$

$$= -U \left[\frac{\delta G_\sigma(1,2)}{\delta \phi_{-\sigma}(1, 1^+)} - G_{-\sigma}(1, 1^+) G_\sigma(1,2) \right]. \tag{B.2}$$

Substituting the equation for the three-point susceptibility (collective modes) (Eq. (26)) in this last equation and multiplying on both sides by G^{-1} proves [27] the expression (Eq. (27)) for the self-energy.

2. We now show that our approach satisfies the consistency requirement between single-particle properties and collective modes in the form of equation (48). Using our expression (Eq. (46)) for $\Sigma^{(1)}$ and the definition of χ_0 (Eq. (A.19)) we obtain

$$\lim_{\tau\to 0^-}\frac{T}{N}\sum_k \Sigma_\sigma^{(1)}(k)G_\sigma^{(0)}(k)\mathrm{e}^{-ik_n\tau} = Un_{-\sigma}^2 - \frac{U}{4}\frac{T}{N}\sum_q [U_{\mathrm{sp}}\chi_{\mathrm{sp}}(q)+U_{\mathrm{ch}}\chi_{\mathrm{ch}}(q)]\frac{\chi_0(q)}{2}. \quad \text{(B.3)}$$

Using

$$\chi_{\mathrm{sp}}(q)-\chi_0(q) = \frac{U_{\mathrm{sp}}}{2}\chi_0(q)\,\chi_{\mathrm{sp}}(q) \quad \text{(B.4)}$$

$$\chi_0(q)-\chi_{\mathrm{ch}}(q) = \frac{U_{\mathrm{ch}}}{2}\chi_0(q)\,\chi_{\mathrm{ch}}(q) \quad \text{(B.5)}$$

and the local moment (Eq. (38)) and local charge (Eq. (37)) sum rules proves the result. The result is also obvious if we follow the steps in the first part of this appendix to deduce the self-energy expression (Eq. (31)) using the collective mode equation (Eq. (30)) adapted to our approximation.

Appendix C

Ansatz for Relation between U_{sp} and $\langle n_\uparrow n_\downarrow\rangle$

Using the present notation and formalism, we now give a physical derivation of equation (40) that is equivalent to the one already given using the equations of motion approach [29]. (The latter derivation was inspired by the local field approximation of Singwi *et al.* [31]). Since our considerations on collective modes are independent of the precise value of the interaction U, we do have to use the equations of motion, or the equivalent, to feed that information back in the definition of irreducible vertices. The two irreducible vertices that we need are in principle calculable from

$$\Gamma_{\sigma\sigma'}\delta(1-3)\,\delta(2-4)\,\delta\left(2-1^+\right) = \frac{\delta\Sigma_\sigma(1,2)}{\delta G_{\sigma'}(3,4)} = \frac{\delta\left[\Sigma_\sigma\left(1,\overline{1}\right)G_\sigma\left(\overline{1},\overline{2}\right)G_\sigma^{-1}\left(\overline{2},2\right)\right]}{\delta G_{\sigma'}(3,4)}. \quad \text{(C.1)}$$

The rewriting on the right-hand side has been done to take advantage of the fact that in the Hubbard model, the equations of motion (see Eqs. (A.2, A.3)) give us the product $\Sigma_\sigma\left(1,\overline{1}\right)G_\sigma\left(\overline{1},\overline{2}\right)$ as the highly local four field correlation function $-U\left\langle T_\tau\left[\psi_{-\sigma}^+\left(1^{++}\right)\psi_{-\sigma}\left(1^+\right)\psi_\sigma(1)\,\psi_\sigma^+\left(\overline{2}\right)\right]\right\rangle$. Ordinary RPA amounts to a Hartree-Fock factoring of this correlation function. Pursuing the philosophy that the minimum number of approximations should be done on local correlation functions, we do this factoring in such a way that it becomes exact when all points are identical, namely when $\overline{2}=1^+$. In other words, we write

$$-U\left\langle T_\tau\left[\psi_{-\sigma}^+(1^{++})\psi_{-\sigma}(1^+)\psi_\sigma(1)\psi_\sigma^+(\overline{2})\right]\right\rangle \sim U\frac{\langle n_\uparrow(1)\,n_\downarrow(1)\rangle}{\langle n_\uparrow(1)\rangle\langle n_\downarrow(1)\rangle}G_{-\sigma}(1,1^+)G_\sigma(1,\overline{2}). \quad \text{(C.2)}$$

All quantities are evaluated as functionals of G up to this point. We can now evaluate the functional derivative

$$\frac{\delta\Sigma_\sigma(1,2)}{\delta G_{\sigma'}(3,4)} = \frac{\delta\left[U\frac{\langle n_\uparrow(1)n_\downarrow(1)\rangle}{\langle n_\uparrow(1)\rangle\langle n_\downarrow(1)\rangle}G_{-\sigma}(1,1^+)\,\delta(1-2)\right]}{\delta G_{\sigma'}(3,4)} \tag{C.3}$$

$$= \frac{\delta\left[U\frac{\langle n_\uparrow(1)n_\downarrow(1)\rangle}{\langle n_\uparrow(1)\rangle\langle n_\downarrow(1)\rangle}\right]}{\delta G_{\sigma'}(3,4)}G_{-\sigma}\left(1,1^+\right)\delta(1-2) + U\frac{\langle n_\uparrow(1)\,n_\downarrow(1)\rangle}{\langle n_\uparrow(1)\rangle\,\langle n_\downarrow(1)\rangle}\frac{\delta G_{-\sigma}(1,1^+)}{\delta G_{\sigma'}(3,4)}\delta(1-2)\,. \tag{C.4}$$

The functional derivatives are now evaluated for the actual equilibrium value of G. Hence, we can use rotational invariance to argue that the first term is independent of σ and σ' whereas the last one is proportional to $\delta_{-\sigma,\sigma'}$. Since $U_{\rm sp} = \Gamma_{\uparrow\downarrow} - \Gamma_{\uparrow\uparrow}$, only this last term proportional to $\delta_{-\sigma,\sigma'}$ contributes to $U_{\rm sp}$. To obtain this term, it suffices to note that

$$\frac{\delta G_{-\sigma}(1,1^+)}{\delta G_{\sigma'}(3,4)} = \delta_{-\sigma,\sigma'}\delta(1-3)\,\delta\left(4-1^+\right) \tag{C.5}$$

and we obtain the desired result (Eq. (40)) for $U_{\rm sp}$.

Appendix D

Real-Frequency Analysis of the Self-Energy and Fermi Liquid Limit

It is instructive to recover the two-dimensional result for precursors of antiferromagnetic bands using the real-frequency formalism since it also clarifies the limit in which the Fermi liquid result is recovered. Again we neglect the contribution of charge fluctuations. Starting from our expression for the self-energy (Eq. (46)), one uses the spectral representation for the susceptibility and for $G^{(0)}$. The Matsubara frequency sums can be then done and the result is trivially continued to real frequencies [85]. One obtains, for the contribution of classical and quantum spin fluctuations to the self-energy in d dimensions

$$\Sigma^{\rm R}(\mathbf{k},\omega) = \frac{UU_{\rm sp}}{4}\int\frac{{\rm d}^dq}{(2\pi)^d}\int\frac{{\rm d}\omega'}{\pi}\left[n(\omega') + f(\varepsilon_{\mathbf{k}+\mathbf{q}})\right]\frac{\chi''_{\rm sp}(\mathbf{q},\omega')}{\omega + i\eta + \omega' - (\varepsilon_{\mathbf{k}+\mathbf{q}} - \mu_0)} \tag{D.1}$$

where $\mu_0 = 0$ at half-filling in the nearest-neighbor model and where f is, as usual, the Fermi function, while $n(\omega) = \left({\rm e}^{\beta\omega} - 1\right)^{-1}$ is the Bose-Einstein distribution. To analyze this result in various limiting cases we need to know more about the frequency dependence of the spin susceptibility. When the antiferromagnetic correlation length is large, the zero-frequency result (Eq. (50)) mentioned above can be generalized to

$$\chi^{\rm R}_{\rm sp}(\mathbf{q}+\mathbf{Q}_d,\omega) \approx \xi^2\frac{2}{U_{\rm sp}\xi_0^2}\left[\frac{1}{1+\mathbf{q}^2\xi^2 - i\omega/\omega_{\rm SF}}\right] \tag{D.2}$$

where, $\omega_{\rm SF} = D/\xi^2$ is the characteristic spin relaxation frequency. In the notation of reference [33], the microscopic diffusion constant D is defined by

$$\frac{1}{D} \equiv \frac{\tau_0}{\xi_0^2} \tag{D.3}$$

with the microscopic relaxation time,

$$\tau_0 = \frac{1}{\chi_0(\mathbf{Q}_d)}\left.\frac{\partial\chi^{\rm R}_0(\mathbf{Q}_d,\omega)}{\partial i\omega}\right|_{\omega=0}. \tag{D.4}$$

This relaxation-time is non-zero in models where the Fermi surface intersects the magnetic Brillouin zone. Clearly, the frequency dependence of $\chi^{\rm R}_{\rm sp}(\mathbf{q}+\mathbf{Q}_d,\omega)$ is on a scale $\omega_{\rm SF}=D/\xi^2$. The $1/\omega$ decrease of $\chi''_{\rm sp}$ at high-frequency is not enough to ensure that the real frequency version of the local-moment sum rule is satisfied and the simplest way to cure this problem is to introduce [86] a high-frequency cutoff $\Omega_{\rm cut}$. The large correlation length makes the characteristic energy of the spin fluctuations $\omega_{\rm SF}$ a small number (critical slowing down). We consider in turn two limiting cases [87]. The Fermi-liquid regime appears for $\omega_{\rm SF}\gg T$ and the non-Fermi liquid regime in the opposite (renormalized classical) regime $\omega_{\rm SF}\ll T$.

D.1. Fermi Liquid and Nested Fermi Liquid Regime $\omega_{\rm SF}\gg T$. — Perhaps the best known characteristic of a Fermi liquid is that $\Sigma''^{\rm R}(\mathbf{k}_{\rm F},\omega;T=0)\propto\omega^2$ and $\Sigma''^{\rm R}(\mathbf{k}_{\rm F},\omega=0;T)\propto T^2$. To recover this result in the regime $\omega_{\rm SF}\gg T$ far from phase transitions, we start from the above expression (Eq. (D.1)) for the self-energy to obtain

$$\begin{aligned}\Sigma''^{\rm R}\left(\mathbf{k}_{\rm F},\omega\right) &= -\frac{UU_{\rm sp}}{4}\frac{1}{2v_{\rm F}}\int\frac{{\rm d}^{d-1}q_\perp}{(2\pi)^{d-1}}\int\frac{{\rm d}\omega'}{\pi}\\ &\quad\times\left[n\left(\omega'\right)+f\left(\omega+\omega'\right)\right]\chi''_{\rm sp}\left(q_\perp,q_{\parallel}\left(q_\perp,\mathbf{k}_{\rm F},\omega,\omega'\right);\omega'\right)\end{aligned}\tag{D.5}$$

where $q_\parallel$, the component of $\mathbf{q}$ parallel to the Fermi momentum $\mathbf{k}_{\rm F}$, is obtained from the solution of the equation

$$\varepsilon_{\mathbf{k}+\mathbf{q}}-\mu_0=\omega+\omega'.\tag{D.6}$$

The key to understanding the Fermi liquid *versus* non-Fermi liquid regime is in the relative width in frequency of $\chi''_{\rm sp}\left(\mathbf{q},\omega'\right)/\omega'$ *versus* the width of the combined Bose and Fermi functions. In general, the function $n\left(\omega'\right)+f\left(\omega+\omega'\right)$ depends on ω' on a scale ${\rm Max}\left(\omega,T\right)$ while far from a phase transition, the explicit frequency dependence of $\chi''_{\rm sp}\left(\mathbf{q},\omega'\right)/\omega'$ is on a scale $\omega_{\rm SF}\sim E_{\rm F}\gg T$. Hence, in this case we can assume that $\chi''_{\rm sp}\left(\mathbf{q},\omega'\right)/\omega'$ is a constant in the frequency range over which $n(\omega')+f(\omega+\omega')$ differs from zero. Also, since $\chi''_{\rm sp}(\mathbf{q},\omega')/\omega'$ depends on wave vector $\mathbf{q}$ over a scale of order $q_{\rm F}$, one can neglect the $\omega+\omega'$ dependence of $q_\parallel$ obtained from equation (D.6). Hence, we can approximate our expression (Eq. (D.5)) for $\Sigma''^{\rm R}$ by

$$\begin{aligned}\Sigma''^{\rm R}\left(\mathbf{k}_{\rm F},\omega\right) &\simeq -\frac{UU_{\rm sp}}{4}\frac{A\left(\mathbf{k}_{\rm F}\right)}{2v_{\rm F}}\int\frac{{\rm d}\omega'}{\pi}\left[n\left(\omega'\right)+f\left(\omega+\omega'\right)\right]\omega'\\ &= -\frac{UU_{\rm sp}}{4}\frac{A\left(\mathbf{k}_{\rm F}\right)}{4v_{\rm F}}\left[\omega^2+\left(\pi T\right)^2\right]\end{aligned}\tag{D.7}$$

where the substitution $x={\rm e}^{\beta\omega}$ allowed the integral to be done exactly and where

$$A\left(\mathbf{k}_{\rm F}\right)\equiv\int\frac{{\rm d}^{d-1}q_\perp}{(2\pi)^{d-1}}\lim_{\omega\to0}\frac{\chi''_{\rm sp}\left(q_\perp,q_{\parallel}\left(q_\perp,\mathbf{k}_F,0,0\right);\omega'\right)}{\omega'}.\tag{D.8}$$

In general, A depends on the orientation of the Fermi wave vector, $\widehat{\mathbf{k}}_{\rm F}$, because it determines the choice of parallel and perpendicular axis $q_\perp,q_\parallel$. The above result (Eq. (D.7)) for $\Sigma''^{\rm R}$ is the well known Fermi liquid result.

There are known corrections to the Fermi liquid self-energy that come from the non-analytic $\omega'/v_{\rm F}q$ behavior of $\chi''_{\rm sp}\left(\mathbf{q},\omega'\right)/\omega'$ near the ferromagnetic (zone center) wave vector. In three dimensions [88] this non-analyticity leads to subdominant $\omega^3\ln\omega$ corrections, while in two dimensions it leads to the dominant $\omega^2\ln\omega$ behavior [89,90]. In the case under consideration, the antiferromagnetic contribution has a larger prefactor. Even when it dominates however, it can also lead to non-analyticities in the case of a nested Fermi surface. Indeed, we note that

$${\rm Im}\,\chi^{\rm R}_0(\mathbf{Q}_d,\omega)=\pi N_{\rm d}(\frac{\omega}{2}){\rm tanh}\left(\frac{\omega}{4T}\right).\tag{D.9}$$

In two dimensions, the logarithmic divergence of the density of states $N_d(\frac{\omega}{2})$ at the van Hove singularity makes the zero-frequency limit of the microscopic relaxation time (Eq. (D.4)) ill-defined, because of the logarithmic divergence at $\omega = 0$. However, this leads only to logarithmic corrections. If we drop logarithmic dependencies, then for $\omega < T$ one has $\partial\chi_0^{\mathrm{R}}(\mathbf{Q}_d,\omega)/\partial i\omega\big|_{\omega\sim T} \sim 1/T$ and this $1/T$ dependence of $\partial\chi_0^{\mathrm{R}}(\mathbf{Q}_d,\omega)/\partial i\omega\big|_{\omega=0}$ changes the temperature dependence of $\Sigma''^{\mathrm{R}}(\mathbf{k}_{\mathrm{F}},0)$ from T^2 to T as discussed in the "Nested Fermi Liquid" approach [91].

D.2. Non-Fermi Liquid Regime $\omega_{\mathrm{SF}} \ll T$. — Near an antiferromagnetic phase transition, the spin-fluctuation energy becomes much smaller than temperature. This is the renormalized classical regime. The condition $\omega_{\mathrm{SF}} \ll T$ means that $\chi''_{\mathrm{sp}}(q_\perp, q_\parallel;\omega')$ is peaked over a frequency interval $\omega' \ll T$ much narrower than the interval $\omega' \sim T$ over which $n(\omega')+f(\omega+\omega')$ changes. This situation is the opposite of that encountered in the Fermi liquid regime. To evaluate Σ''^{R} (Eq. (D.5)) the Fermi factor can now be neglected compared with the classical limit of the Bose factor, T/ω'. Then the dominant contribution to $\Sigma''^{\mathrm{R}}(\mathbf{k}_{\mathrm{F}},\omega)$ is from classical spin fluctuations $T\int \frac{\mathrm{d}\omega'}{\pi}\frac{1}{\omega'}\chi''_{\mathrm{sp}} = T\chi'_{\mathrm{sp}} \simeq S_{\mathrm{sp}}$ as we see below. More specifically, we take into account that the integral is peaked near $\mathbf{Q} = (\pi,\pi)$ and measure wave vector with respect to the zone center. For simplicity we consider below the half-filled case $\mu_0 = 0$. Then, with the help of $\varepsilon_{\mathbf{k}+\mathbf{q}+\mathbf{Q}} = -\varepsilon_{\mathbf{k}+\mathbf{q}}$ we approximate the equation for $q_\parallel$ (Eq. (D.6)) by $v_{\mathrm{F}}q_\parallel = -(\omega+\omega')$. This gives us for equation (D.5) the approximation

$$\Sigma''^{\mathrm{R}}(\mathbf{k}_{\mathrm{F}},\omega) \approx -\frac{UU_{\mathrm{sp}}}{4}\frac{1}{2v_{\mathrm{F}}}\int\frac{\mathrm{d}^{d-1}q_\perp}{(2\pi)^{d-1}}\int\frac{\mathrm{d}\omega'}{\pi}\frac{T}{\omega'}\chi''_{\mathrm{sp}}\left(q_\perp, q_\parallel = -\frac{\omega+\omega'}{v_{\mathrm{F}}};\omega'\right). \tag{D.10}$$

The dependence of χ''_{sp} on ω' through $q_\parallel = -(\omega+\omega')/v_{\mathrm{F}}$ may be neglected because $q_\parallel$ appears only in the combination $(\xi^{-2}+q_\perp^2+q_\parallel^2)$ and in the regime $\omega_{\mathrm{SF}} \ll T$ we have $\omega'/v_{\mathrm{F}} < \omega_{\mathrm{SF}}/v_{\mathrm{F}} \sim D\xi^{-2}/v_{\mathrm{F}} \ll \xi^{-1}$. The latter inequality is generically satisfied when $\xi^{-1} \ll 1$. Using

$$T\int\frac{\mathrm{d}\omega'}{\pi}\frac{1}{\omega'}\chi''_{\mathrm{sp}}\left(q_\perp, q_\parallel = -\frac{\omega}{v_{\mathrm{F}}};\omega'\right) = T\chi'_{\mathrm{sp}}\left(q_\perp, q_\parallel = -\frac{\omega}{v_{\mathrm{F}}}; iq_{\mathrm{n}} = 0\right) \tag{D.11}$$

$$= \frac{2}{U_{\mathrm{sp}}\xi_0^2}\frac{T}{\xi^{-2}+q_\perp^2+\left(\frac{\omega}{v_{\mathrm{F}}}\right)^2} \tag{D.12}$$

the above equation (Eq. (D.10)) for $\Sigma''^{\mathrm{R}}(\mathbf{k}_{\mathrm{F}},\omega)$ reduces precisely to the classical contribution found using imaginary-time formalism (Eq. (55)). As we saw in Section 5.1.1, when the condition $\xi > \xi_{\mathrm{th}}$ is satisfied, then this contribution is dominant and leads to $\lim_{T\to 0}\Sigma''^{\mathrm{R}}(\mathbf{k}_{\mathrm{F}},0) \to \infty$.

Appendix E

Expanded Discussion of Other Approaches

This appendix expands in Section 8 to discuss in detail various theories, explaining the advantages and disadvantages of each in the context of the sets of constraints described in Appendices A.2 and A.3.

E.1. Paramagnon Theories. — In standard Paramagnon theories [32, 46], the spin and charge fluctuations are computed by RPA, using either bare or dressed Green's functions. Then the fluctuations are fed back in the self-energy. In fact there is a whole variety of paramagnon

theories. They are largely phenomenological. The reader is referred to reference [46] for a review. We concentrate our discussion on recent versions [52] of the so-called Berk-Schrieffer formula [92]. In this approach, infinite subsets of diagrams are summed and bare propagators are used in the calculation of both the susceptibilities and the self-energy, the latter being given by

$$\Sigma_\sigma^{\mathrm{BS}}(k) = Un_{-\sigma} + \frac{U}{4}\frac{T}{N}\sum_q \left[\left(3U\chi_{\mathrm{sp}}^{\mathrm{RPA}}(q) - 2U\chi_0(q)\right) + U\chi_{\mathrm{ch}}^{\mathrm{RPA}}(q)\right] G_\sigma^0(k+q). \tag{E.1}$$

The RPA spin and charge susceptibilities have been defined in equations (A.17, A.18). Comparing with our self-energy formula (Eq. (46)), it is clear that here there is no vertex correction. In addition, the factor of three in front of the spin susceptibility in equation (E.1) is supposed to take into account the presence of both longitudinal and transverse spin waves and the subtracted term is to avoid double-counting the term of order U^2.

We can now see the advantages and disadvantages of this approach. First, note that the susceptibilities entering the Berk-Schrieffer formula are the RPA ones. As we saw in Appendix A, these fail to satisfy both the local Pauli principle and the Mermin-Wagner theorem. Hence, spurious phase transitions will influence the self-energy in uncontrollable ways. The collective modes do however satisfy conservation laws since they are obtained with bare vertices and Green's functions containing a constant self-energy. The f-sum rule (Eqs. (A.22)) then is satisfied without renormalization of the distribution function $n_{\mathbf{k}}$ because the zeroth order self-energy is constant. This is all in agreement with the definition of a conserving approximation for the collective modes.

The high-energy asymptotics of the self-energy sets in at the correct energy scale $k_{\mathrm{n}} > W$ in this approach, but the second term of the large-frequency asymptotics is incorrect. Indeed, at large values of ik_{n},

$$\lim_{ik_{\mathrm{n}}\to\infty} \Sigma_\sigma^{\mathrm{BS}}(k) = Un_{-\sigma} + \frac{U}{4ik_{\mathrm{n}}}\frac{T}{N}\sum_q \left[3U\chi_{\mathrm{sp}}^{\mathrm{RPA}}(q) + U\chi_{\mathrm{ch}}^{\mathrm{RPA}}(q) - 2U\chi_0(q)\right] + \cdots \tag{E.2}$$

and the sums can be evaluated as follows using the fluctuation-dissipation theorem

$$\frac{T}{N}\sum_q \chi_{\mathrm{sp}}^{\mathrm{RPA}}(q) = 2\langle n_\uparrow n_\uparrow\rangle - 2\langle n_\uparrow n_\downarrow\rangle \tag{E.3}$$

$$\frac{T}{N}\sum_q \chi_{\mathrm{ch}}^{\mathrm{RPA}}(q) = 2\langle n_\uparrow n_\uparrow\rangle + 2\langle n_\uparrow n_\downarrow\rangle - n^2 \tag{E.4}$$

$$\frac{T}{N}\sum_q \chi_0(q) = n - \frac{n^2}{2}\cdot \tag{E.5}$$

The correlators on the right-hand side take their RPA value so they do not satisfy the Pauli principle, *i.e.* $\langle n_\uparrow n_\uparrow\rangle \neq \langle n_\uparrow\rangle$. Taking these results together we have

$$\lim_{ik_{\mathrm{n}}\to\infty} \Sigma_\sigma^{\mathrm{BS}}(k) = Un_{-\sigma} + \frac{U^2}{ik_{\mathrm{n}}}\left[2\langle n_\uparrow n_\uparrow\rangle - \langle n_\uparrow n_\downarrow\rangle - \frac{n}{2}\right] + \cdots \tag{E.6}$$

This does not gives the correct asymptotic behavior (Eq. (68)) even if the Pauli principle $\langle n_\uparrow n_\uparrow\rangle = \langle n_\uparrow\rangle$ were satisfied, because $\langle n_\uparrow n_\downarrow\rangle$ depends on the interaction U.

The Paramagnon self-energy (Eq. (E.1)) also does not satisfy the consistency requirement (Eq. (45)) between self-energy and collective modes imposed by the equations of motion. To see this we first note that

$$\lim_{\tau\to 0^-} \frac{T}{N}\sum_k \Sigma_\sigma^{\mathrm{BS}}(k)\, G_\sigma^{(0)}(k)\, \mathrm{e}^{-ik_\mathrm{n}\tau}$$
$$= Un_{-\sigma}^2 - \frac{U^2}{8}\frac{T}{N}\sum_{\mathbf{q}} \left[3\chi_{\mathrm{sp}}^{\mathrm{RPA}}(q) + \chi_{\mathrm{ch}}^{\mathrm{RPA}}(q) - 2\chi_0(q)\right]\chi_0(q)\,. \tag{E.7}$$

Using this expression in the sum-rule (Eq. (45)) which relates one and two-particle correlators and expanding both sides of this sum-rule in powers of U, one finds that it is satisfied only up to order U^2. On the other hand, if one replaces $3\chi_{\mathrm{sp}} - 2\chi_0$ in equation (E.1) by χ_{sp}, the sum-rule (Eq. (45)) is satisfied to all orders in U. In our opinion, the problem of enforcing rotational invariance in approximate theories is highly non-trivial and cannot be solved simply by adding factor of 3 in front of χ_{sp} and then subtracting $2\chi_0$ to avoid double counting. For more detailed discussions see reference [50] and the comments at the end of Section 3.2.2.

Luttinger's theorem is trivially satisfied if the occupation number is calculated with the initial constant self-energy since it gets absorbed in the chemical potential. If the occupation number is calculated with the Green's function that contains the Berk-Schrieffer self-energy then Luttinger's theorem is in general violated. It is advisable to use a new chemical potential.

E.2. Conserving Approximations (FLEX). — In the conserving approximation schemes [26], one takes any physically motivated subset of skeleton diagrams to define a Luttinger-Ward functional Φ. Skeleton diagrams contain fully dressed Green's functions and no self-energy insertions. This functional is functionally differentiated to generate a self-energy that is then calculated self-consistently since it appears implicitly in the Green's functions used in the original set of diagrams. A further functional differentiation allows one to calculate the irreducible vertices necessary to obtain the collective modes in a way that preserves Ward identities. If one uses for the free energy the formula

$$\ln Z = \mathrm{Tr}\left[\ln(-G)\right] + \mathrm{Tr}\left(\Sigma G\right) - \Phi \tag{E.8}$$

then one obtains thermodynamic consistency in the sense that thermodynamic quantities obtained by derivatives of the free energy are identical to quantities computed directly from the single-particle Green's function. For example, particle number can be obtained either from a trace of the Green's function or from a chemical potential derivative of the free energy. In this scheme, Luttinger's theorem is satisfied as long as perturbation theory converges since then any initial guess for the Luttinger-Ward functional will satisfy Luttinger's theorem.

FLEX refers to a particular choice of diagrams for Φ. This choice leads to the following self-consistent expression for the self-energy

$$\Sigma_\sigma^{\mathrm{BS}}(k) = Un_{-\sigma} + \frac{U}{4}\frac{T}{N}\sum_q \left[\left(3U\tilde{\chi}_{\mathrm{sp}}^{\mathrm{RPA}}(q) - 2U\tilde{\chi}_0(q)\right) + U\tilde{\chi}_{\mathrm{ch}}^{\mathrm{RPA}}(q)\right] G_\sigma(k+q). \tag{E.9}$$

This expression for the self-energy does not contain vertex corrections, despite the fact that, contrary to the electron-phonon case, Migdal's theorem does not apply here. We have explained in detail in Section 6.2 why this may lead to qualitatively wrong results, such as the absence of precursors of antiferromagnetic bands and of the pseudogap in $A(\mathbf{k}_{\mathrm{F}}, \omega)$ in two dimensions.

Another drawback of this approach is that it does not satisfy the Pauli principle in any form, either local or through crossing symmetry [93]. Indeed, one would need to include

all exchange diagrams to satisfy it. In practice this is never done. In the same way that there is nothing to constrain the value of $\langle n_\uparrow n_\uparrow \rangle$ obtained by the fluctuation-dissipation theorem to be equal to $\langle n_\uparrow \rangle$, there is nothing to explicitly constrain the value of $\langle n_\uparrow n_\downarrow \rangle$. Nevertheless, the Mermin-Wagner theorem is believed to be satisfied in FLEX because the feedback through the self-energy tends to prevent the divergence of fluctuations in low dimension [38,94]. Physically however, this seems to be an artificial way of satisfying the Mermin-Wagner theorem since this theorem should be valid even in localized spin systems where single-particle properties are negligibly influenced by thermal fluctuations. We also point out that the proof of the Mermin-Wagner theorem in $n \to \infty$ models implies that the finite temperature phase transition in two dimensions is not simply removed by thermal fluctuations, but that it is replaced by a crossover to the renormalized classical regime with exponentially growing susceptibility. The fact that the conserving susceptibility in FLEX does not show such behavior [38] means that FLEX is actually inconsistent with the generic phase space arguments responsible for the absence of finite-temperature phase transition in two dimensions. The case of one dimension also suggests that collective modes by themselves should suffice to guarantee the Mermin-Wagner theorem without feedback on single-particle properties. Indeed, in one dimension one shows by diagrammatic methods (parquet summation or renormalization) that the zero-temperature phase transition is prohibited at the two-particle level even *without* self-energy effects [8].

Although, the second-order diagram is included correctly in FLEX, it does not have the correct coefficient in the $1/ik_{\mathrm{n}}$ expansion of the self-energy. More importantly, the high-frequency behavior sets-in too late to give the Hubbard bands, as we have explained in Section 6.2. We have also seen a case where FLEX, as judged from comparisons with Monte Carlo simulations (Fig. 1a of Ref. [30]), does not reproduce the results of second-order perturbation theory even when it is a good low-energy approximation.

One of the inconsistencies of conserving approximations that is seldom realized, is that the self-energy is inconsistent with the collective modes. In other words, the consistency formula (Eq. (44)) is not satisfied in the following sense. The explicit calculation of ΣG leads to an estimate of $U \langle n_\uparrow n_\downarrow \rangle$ that differs from the one obtained by applying the fluctuation-dissipation theorem to the *conserving* spin and charge susceptibilities.

E.3. Pseudo-Potential Parquet Approach. — In the parquet approach, one enforces complete antisymmetry of the four point function by writing down fully crossing-symmetric equations for these. There are three irreducible vertices, namely one for the particle-particle channel, and one for each of the two particle-hole channels. They obey the so-called parquet equations [95]. The Green's functions are dressed by a self-energy which itself contains the four point function. In this way, self-consistency between one-particle and two-particle quantities is built-in. Solutions are possible for the one-impurity problem [96] and in one-dimension [8]. However, to solve the parquet equations in higher dimension with presently available computing power is impossible. Bickers *et al.* [25,53] have formulated the parquet equations as a systematic improvement over FLEX and have devised a way to do practical calculations by introducing so-called pseudo-potentials. Since the main computational difficulty is in keeping the full momentum and frequency dependence of the four point functions entering the calculation of the self-energy, this is where the various fluctuations channels are approximated by RPA-like forms (Eq. (A.25)) but with fully dressed propagators and an effective interaction (pseudo-potential) instead of U. A different strategy is under development [94]. The criticism of the present section applies only to the current pseudo-potential parquet approach [25, 53].

It can be seen that one drawback of this approach at the physical level is that the use of constant effective interactions with dressed single-particle propagators means that the fluctuations used in the calculation of the self-energy do not satisfy conservation laws, as we just

demonstrated in Section A.3. Furthermore, the pseudopotentials are determined by asking that the susceptibilities extracted from the four-point functions in the parquet equations match the corresponding RPA-pseudo-potential susceptibility at only one wave vector and frequency. The choice of this matching point is arbitrary: should the match be done for the typical, the average, or the maximal value of the susceptibility in the Brillouin zone?

As we have seen in Section 6, even if the expression for the self-energy in this approach explicitly has the second-order perturbation theory diagram in it, this is not sufficient to ensure that the correct high frequency asymptotic behavior starts at the appropriate frequency scale $ik_{\mathrm{n}} \sim W$. Nevertheless, in many cases the results of the calculations performed with this approach are not so different from second-order perturbation theory, as can be seen from Figure 1 of reference [30].

Going rapidly through the rest of our list of properties, we see that the consistency requirement $\Sigma_\sigma\left(1,\overline{1}\right) G_\sigma\left(\overline{1},1^+\right) = U\left\langle n_\uparrow n_\downarrow\right\rangle$ is at least approximately built-in by construction. Concerning the local-moment sum-rule and the Mermin-Wagner theorem, it has been shown that the so-called "basic" parquet equations should have the same critical behavior as the leading term in the $1/N$ expansion [97], and hence should satisfy the Mermin-Wagner theorem [94]. The pseudo-potentials should not affect the self-consistency necessary to satisfy the Mermin-Wagner theorem but the fact that they are matched at a single point might introduce difficulties, especially if the wave-vector at which χ_{sp} becomes unstable is unknown from the start. As far as the Pauli principle is concerned, it should be at least approximately satisfied both locally and in momentum space. Nothing however in the approach enforces conservation laws.

E.4. Present Approach. — The role of the above sum-rules in our approach has been discussed in detail in the main text. Here we will discuss only a few additional points.

If we concentrate on the $\mathbf{q} = 0$ properties, our spin and charge correlations behave as a special case of the "local Fermi liquid" defined in reference [98]. A "local Fermi liquid" is a description of $\mathbf{q} = 0$ properties that applies when the self-energy, and consequently irreducible vertices, depend only on frequency, not on momentum. In a local Fermi liquid there are only two Landau parameters, which in our case are $F_0^{\mathrm{a}} = -U_{\mathrm{sp}}\chi_0\left(0^+,0\right)/2$ and $F_0^{\mathrm{s}} = U_{\mathrm{ch}}\chi_0\left(0^+,0\right)/2$. Unitarity and the forward scattering sum rule, if valid, imply that there is no ferromagnetism in the repulsive case [98], as we have found. One can check explicitly that the forward scattering sum rule is satisfied to within about 15% in our usual Monte Carlo parameter range. However, as discussed in Appendix A.4, the forward scattering sum-rule refers only to wave vectors on the Fermi surface, not to the local version of the Pauli principle. Furthermore, the validity of this sum rule has been questioned [84]. The effective mass at this level of approximation is the bare one, as in a transitionally invariant *local* Fermi liquid [98]. Recall however that our microscopic calculations are not phenomenological: they explicitly give a value for the Landau parameters. Also, our results extend well beyond the $\mathbf{q} = 0$ quantities usually considered in Fermi liquid theory.

The quasi-particle weight Z calculated with $\Sigma_\sigma^{(1)}$ can differ substantially from the initial one. This means that if we were to calculate the susceptibility with the corresponding frequency and momentum dependent irreducible vertices $\Gamma^{(1)}$ there would be sizeable compensation between vertices and self-energy because our calculations with $\Sigma_\sigma^{(0)}$ $(Z = 1)$ and constant renormalized vertices already gave excellent agreement with Monte Carlo simulations.

Finally, consider the high-frequency asymptotics. Since we use bare propagators, the high-frequency asymptotics comes in at the appropriate frequency scale, namely $ik_{\mathrm{n}} \sim W$ and the Hubbard bands do exist in our theory. However, the coefficient of proportionality in front of the asymptotic form $1/ik_{\mathrm{n}}$ is incorrect. Using equations (46, A.14, A.15) we can write

the high-frequency asymptotics in the following form

$$\lim_{ik_n\to\infty}\Sigma_\sigma(\mathbf{k},ik_n) = Un_{-\sigma}+\frac{U}{ik_n}\left[\left(\frac{U_{sp}+U_{ch}}{2}\right)\langle n^2_{-\sigma}\rangle - U_{ch}n^2_{-\sigma}+\left(\frac{U_{sp}-U_{ch}}{2}\right)\langle n_\uparrow n_\downarrow\rangle\right]+\cdots \quad \text{(E.10)}$$

This form is useful to understand what is necessary to obtain the quantitatively correct high-frequency behavior. Indeed, one would recover the exact result (Eq. (68)), if one were to take into account that: i) the irreducible vertices become equal to the bare one U at high-frequencies; ii) the local Pauli principle $\langle \hat{n}^2_{-\sigma}\rangle = n_{-\sigma}$ is satisfied. Contrary to most other approaches, our theory does satisfy the local Pauli principle (Eq. (A.12)) exactly. However, since our irreducible vertices are constant and tuned to describe the low energy physics, we violate the first of the above requirements. It is thus clear that for a correct quantitative description of both the low energy physics and the Hubbard bands one needs to work with frequency-dependent irreducible vertices.

References

[1] Hubbard J., *Proc. Roy. Soc.* **A276** (1963) 238; Gutzwiller M., *Phys. Rev. Lett.* **10** (1963) 159.
[2] Kanamori J., *Prog. Theor. Phys.* **30** (1963) 275.
[3] For a recent and complete review of electrons interacting in one-dimension see Voit J., *Rep. Prog. Phys.* **58** (1995) 977.
[4] Haldane F.D.M., *Phys. Rev. Lett.* **45** (1980) 1358; *J. Phys. C* **14** (1981) 2585; *Phys. Lett.* **81A** (1981) 153; *Phys. Rev. Lett.* **47** (1981) 1840.
[5] Dzyaloshinski I.E. and Larkin A.I., *Sov. Phys. JETP* **38** (1974) 202.
[6] Emery V.J., in "Highly Conducting One-Dimensional Solids", J.T. Devreese, R.P. Evrard and V.E. van Doren, Eds. (Plenum 1979) p. 247.
[7] Bourbonnais C. and Caron L.G., *Physica* **B143** (1986) 451; *Europhys. Lett.* **5** (1988) 209; C.Bourbonnais C., Ph.D thesis, Université de Sherbrooke (1985) unpublished.
[8] For reviews, see Solyom J., *Adv. Phys.* **28** (1979) 201; Voit J., *Rep. Prog. Phys.* **58** (1995) 977; and [47].
[9] Frahm H. and Korepin V.E., *Phys. Rev. B* **42** (1990) 10553; Frahm H. and Korepin V.E., *Phys. Rev. B* **43** (1991) 5653.
[10] For an introduction to conformal methods, see Conformal Field Theory, P. DiFrancesco, P. Mathieu and D. Sénéchal, Eds. (Springer Verlag, New York, 1996).
[11] Georges A., Kotliar G., Krauth W. and Rozenberg M.J., *Rev. Mod. Phys.* **68** (1996) 13.
[12] For reviews of slave-boson approaches applied to the Hubbard model, see for example, Kotliar G., in Correlated Electrons Systems, Jerusalem Winter School for Theoretical Physics, Vol. 9, V.J. Emery, Ed. (World Scientific, Singapore, 1993) p. 118 and Kotliar G., in Strongly Interacting Fermions and High T_c Superconductivity, Les Houches, Session LVI, 1991, B. Douçot and J. Zinn-Justin, Eds. (Elsevier, Amsterdam, 1995) p. 197.
[13] Yoshioka D., *J. Phys. Soc. Japan* **58** (1989) 32; *ibid.* (1989) 1516; Jayaprakash C., Krishnamurthy H.R. and Sarker S., *Phys. Rev. B* **40** (1989) 2610.
[14] Bickers N.E., *Rev. Mod. Phys.* **59** (1987) 845.
[15] Boies D., Jackson F. and Tremblay A.-M.S., *Int. J. Mod. Phys.* **9** (1995) 1001.
[16] Boies D., Bourbonnais C. and Tremblay A.-M.S., *Phys. Rev. Lett.* **74** (1995) 968.
[17] Dagotto E., *Rev. Mod. Phys.* **66** (1994) 763.
[18] Chakravarty S., Halperin B.I. and Nelson D.R., *Phys. Rev. B* **39** (1989) 2344.

[19] Sachdev S., Chubukov A.V. and Sokol A., *Phys. Rev. B* **51** (1995) 14874.
[20] Chubukov A.V., Sachdev S. and Ye J., *Phys. Rev. B* **49** (1994) 11919.
[21] Hertz J.A., *Phys. Rev. B* **14** (1976) 1165.
[22] Millis A.J., *Phys. Rev. B* **48** (1993) 7183.
[23] For a review, see Moriya T., Spin Fluctuations in Itinerant Electron Magnetism (Springer-Verlag, 1985).
[24] Bickers N.E. and Scalapino D.J., *Annals of Physics* **193** (1989) 206.
[25] Bickers N.E., *Int. J. of Mod. Phys. B* **5** (1991) 253.
[26] Baym G., *Phys. Rev. B* **127** (1962) 1391.
[27] Kadanoff L.P. and Baym G., Quantum Statistical Mechanics (Benjamin, Menlo Park, 1962).
[28] Luttinger J.M. and Ward J.C., *Phys. Rev.* **118** (1960) 1417; Luttinger J.M., *Phys. Rev.* **119** (1960) 1153.
[29] Vilk Y.M., Chen L. and Tremblay A.-M.S., *Phys. Rev. B* **49** (1994) 13267; Vilk Y.M., Chen L. and Tremblay A.-M.S., *Physica C* **235-240** (1994) 2235.
[30] Vilk Y.M. and Tremblay A.-M.S., *Europhys. Lett.* **33** (1996) 159; Vilk Y.M. and Tremblay A.-M.S., *J. Phys. Chem. Solids* **56** (1995) 1769.
[31] For a review, see Singwi K.S. and Tosi M.P., in "Solid State Physics", H. Ehrenreich, F. Seitz and D. Turnbull, Eds. (Academic, New York, 1981) p. 177; Ichimaru S., *Rev. Mod. Phys.* **54** (1982) 1017.
[32] For a review of paragnon theories, see Enz C.P., A Course on Many-Body Theory Applied to Solid-State Physics (World Scientific, Singapore, 1992).
[33] Daré A.-M., Vilk Y.M. and Tremblay A.-M.S., *Phys. Rev. B* **53** (1996) 14236.
[34] Veilleux A., Daré A.-M., Chen L., Vilk Y.M. and Tremblay A.-M.S., *Phys. Rev. B* **52** (1995) 16255.
[35] Kampf A.P. and Schrieffer J.R., *Phys. Rev. B* **41** (1990) 6399; *Phys. Rev. B* **42** (1990) 7967.
[36] Vilk Y.M., *Phys. Rev. B* **55** (1997) 3870.
[37] Langer M., Schmalian J., Grabowski S. and Bennemann K.H., *Phys. Rev. Lett.* **75** (1995) 4508.
[38] Deisz J.J., Hess D.W. and Serene J.W., *Phys. Rev. Lett.* **76** (1996) 1312.
[39] Marshall D.S. *et al.*, *Phys. Rev. Lett.* **76** (1996) 4841; Loeser A.G. *et al.*, *Science* **273** (1996) 325.
[40] Ding H. *et al.*, *Nature* **382** (1996) 51.
[41] Chubukov A.V., *Phys. Rev. B* **52** (1995) R3847.
[42] Martin P.C. and Schwinger J., *Phys. Rev.* **115** (1959) 1342.
[43] $\Sigma_\sigma^{(0)} = \Gamma_{\sigma-\sigma} G_{-\sigma}(\overline{1}, \overline{1}^+) + \Gamma_{\sigma\sigma} G_\sigma(\overline{1}, \overline{1}^+)$. Note that at equilibrium, this initial guess for the self-energy has a Hartree-Fock from but with U replaced by $U_{\rm ch}$, clearly only a low-frequency approximation to the self-energy.
[44] Chen L., Bourbonnais C., Li T. and Tremblay A.-M.S., *Phys. Rev. Lett.* **66** (1991) 369.
[45] Daré A.-M., Chen L. and Tremblay A.-M.S., *Phys. Rev. B* **49** (1994) 4106.
[46] Stamp P.C.E., *J. Phys. F: Met. Phys.* **15** (1985) 1829.
[47] Bourbonnais C. and Caron L.G., *Int. J. Mod. Phys. B* **5** (1991) 1033; Bourbonnais C. and Caron L.G., *Europhys. Lett.* **5** (1988) 209; *Physica B* **143** (1986) 451; Bourbonnais C., in Strongly Interacting Fermions and High $T_{\rm c}$ Superconductivity, Les Houches, Session LVI, 1991, B. Douçot and J. Zinn-Justin, Eds. (Elsevier, Amsterdam, 1995) p. 307.
[48] Independent estimates of the size of contact vertex corrections by Sharifzadeh Amin M.H. and Stamp P.C.E., *Phys. Rev. Lett.* **77** (1996) 3017 give the same order of magnitude as those we find.

[49] Menge B. and Müller-Hartmann E., *Z. Phys. B. Cond. Matt.* **82** (1991) 237.
[50] Macedo C.A. and Coutino-Filho M.D., *Phys. Rev. B* **43** (1991) 13515.
[51] Hotta T. and Fujimoto S., *Phys. Rev. B* **54** (1996) 5381, have checked within perturbation theory that this decrease of charge fluctuations as one approaches half-filling at fixed temperature can indeed be explained only if vertex corrections are included.
[52] Bulut N., Scalapino D.J and White S.R., *Phys. Rev. B* **47** (1993) 2742.
[53] Bickers N.E. and White S.R., *Phys. Rev. B* **43** (1991) 8044.
[54] Galán J., Vergés J.A. and Martín-Rodero A., *Phys. Rev. B* **48** (1993) 13654.
[55] White S.R., *Phys. Rev. B* **46** (1992) 5679; M. Vekić and White S.R., *Phys. Rev. B* **46** (1992) 5679; Vekić M. and White S.R., *Phys. Rev. B* **47** (1993) 1160.
[56] White S.R., Scalapino D.J., Sugar R.L., Loh Jr. E.Y., Gubernatis J.E. and Scalettar R.T., *Phys. Rev. B* **40** (1989) 506.
[57] Moreo A., Scalapino D.J., Sugar R.L., White S.R. and Bickers N.E., *Phys. Rev. B* **41** (1990) 2313.
[58] A similar analysis of Monte Carlo data has been used earlier for the negative U Hubbard model by Randeria M., Trivedi N., Moreo A. and Scalettar R.T., *Phys. Rev. Lett.* **69** (1992) 2001.
[59] Tasaki H., *Phys. Rev. Lett.* **75** (1996) 4678.
[60] Müller-Hartmann E., *Int. J. Mod. Phys.* **12** (1989) 2169.
[61] Amadon J.C. and Hirsch J.E., *Phys. Rev. B* **54** (1996) 6364.
[62] Frésard R. and Wölfle P., *J. Phys.: Cond. Matt.* **4** (1992) 3625; Möller B., Doll K. and Frésard R., *J. Phys.: Cond. Matt.* **5** (1993) 4847. For the phase diagram predicted by the Gutzwiller variational approach, see Dzierzawa M. and Frésard R., *Z. Phys. B: Cond. Matt.* **91** (1993) 245.
[63] Freericks K. and Jarrell M., *Phys. Rev. Lett.* **74** (1995) 186; Tahvildar-Zadeh A.N., Freericks J.K. and Jarrell M., *Phys. Rev. B* **55** (1997) 942.
[64] Furukawa N. and Imada M., *J. Phys. Soc. Jpn* **61** (1992) 3331.
[65] Creffield C.E., Klepfish E.G., Pike E.R. and Sarkar S., *Phys. Rev. Lett.* **75** (1995) 517.
[66] Chakravarty S., Halperin B.I. and Nelson D.R., *Phys. Rev. B* **39** (1989) 2344.
[67] Prudnikov A.P., Brichkov Yu.A. and Marichev O.I., Integrals and Sums (Science, Moscow, 1981) (In russian).
[68] Preuss R., Hanke W. and von der Linden W., *Phys. Rev. Lett.* **75** (1995) 1344.
[69] Schmalian J., Langer M., Grabowski S. and Bennemann K.H., *Phys. Rev. B* **54** (1996) 4336.
[70] Matsumoto H. and Mancini F., *Phys. Rev. B* **55** (1997) 2095. In this work, the crossover to the renormalized classical regime is not clearly identified with the appearance of antiferromagnetic precursor bands. Nevertheless, the new features that appear at low temperature, in addition to Hubbard bands, are associated with antiferromagnetic correlations. We thank H. Matsumoto for correspondance on this subject.
[71] Nozières P. and Schmitt-Rink S., *J. Low Temp. Phys.* **59** (1985) 195.
[72] Sa de Melo C.A.R., Randeria M. and Engelbrecht J.R., *Phys. Rev. Lett.* **71** (1993) 3202.
[73] Trivedi N. and Randeria M., *Phys. Rev. Lett.* **75** (1995) 312.
[74] Tarenko R., Tarenko E. and Malek J., *J. Phys. F* **18** (1988) L87; Müller-Hartmann E., *Z. Phys. B* **76** (1989) 211; Schweitzer H. and Czycholl G., *Z. Phys. B* **83** (1991) 93.
[75] McKenzie R.H. and Scarratt D., *Phys. Rev. B* **54** (1996) R12709.
[76] Schrieffer J.R., *J. Low Temp. Phys.* **99** (1995) 397.
[77] Blaizot J.P. and Iancu E., *Phys. Rev. Lett.* **76** (1996) 3080. The problem of doing correctly the many-body problem below T_{X} is similar to the one encountered in hot three-dimensional quantum electrodynamic plasmas. There, the photons play the role of

fluctuations at their critical point. The divergences are logarithmic in frequency. Furthermore, since the photons can be taken into account order by order in a systematic manner, an infinite resummation is possible. In our case, the spin propagator is not really a different physical entity and, as we saw, it is not clear how to implement a systematic diagrammatic approach. For more details, see Blaizot J.-P. and Iancu E., *Phys. Rev. D* **55** (1997) 973.

[78] Moreo A., Haas S., Sandvik A.W. and Dagotto E., *Phys. Rev. B* **51** (1995) 12045.

[79] In-plane infrared conductivity measurements in $La_{2-x}Sr_xCuO_4$ suggest a pseudogap both above and below optimal doping, Startseva T., Timusk T., Puchkov A.V., Basov D.N., Mook H.A., Kimura T. and Kishio K., cond-mat/9706145.

[80] Emery V.J. and Kivelson S.A., *Phys. Rev. Lett.* **74** (1995) 3253.

[81] We thank D. Sénéchal and S. Pairault for discussion on this point.

[82] Kalashnikov O.K. and Fradkin E.S., *Phys. Stat. Sol. (b)* **59** (1972) 9; Nolting W., *Z. Phys.* **255** (1972) 25; White S.R., *Phys. Rev. B* **44** (1991) 4670.

[83] Abrikosov A.A., Gorkov L.P. and Dzyaloshinski I.E., Methods of quantum field theory in statistical physics (Prentice-Hall, Englewood Cliffs, 1963) Sect. 19.4.

[84] Chitov G.Y. and Senechal D., cond-mat/9705037, *Phys. Rev. B* (in press).

[85] Abrikosov A.A., Gorkov L.P. and Dzyaloshinski I.E., Methods of quantum field theory in statistical physics (Prentice-Hall, Englewood Cliffs, 1963) Eq. (21.26).

[86] Millis A.J., *Phys. Rev. B* **45** (1992) 13047.

[87] Once the expression for the self-energy has been written in the form equation (68), it can be interpreted as the leading self-energy of an electron interacting with a bosonic excitation. Numerous studies of this kind of problem have appeared, including some where the bosonic excitation is softening. For a recent example, see Crisan M. and Tataru L., *J. Superconductivity* **8** (1995) 341; Crisan M. and Tataru L., *Phys. Rev. B* **54** (1996) 3597.

[88] Baym G. and Pethick C., Landau Fermi Liquid Theory, Concepts and Applications (Wiley, New York, 1991). For a microscopic calculation, see: Amit D.J., Kane J.W. and Wagner H., *Phys. Rev. Lett.* **19** (1967) 425 and *Phys. Rev.* **175** (1968) 313.

[89] Stamp P.C.E., *J. Phys. I France* **3** (1993) 625 Appendix A.

[90] Hodges C., Smith H. and Wilkins J.W., *Phys. Rev.* **4** (1971) 302.

[91] Ruvalds J. and Virosztek A., *Phys. Rev. B* **42** (1991) 4064.

[92] Berk N.F. and Schrieffer J.R., *Phys. Rev. Lett.* **17** (1966) 433.

[93] Stamp P.C.E., DPhil Thesis, University of Sussex (1983).

[94] Bickers N.E., private communication.

[95] de Dominicis C. and Martin P.C., *J. Math. Phys.* **5** (1964) 14.

[96] Roulet B., Gavoret F. and Nozières P., *Phys Rev.* **178** (1969) 1072.

[97] Bickers N.E. and Scalapino D.J., *Phys. Rev. B* **46** (1992) 8050. The proof that parquet equations are equivalent to the leading order of the $1/N$ expansion is valid when there is a mode diverging in a single channel [94]. It is known that in one dimension parquet equations are much more general than the $1/N$ expansion because they contain interference terms between channels that are absent in the classical $1/N$ expansion. These interference terms are the cause of the vanishing of the transition temperature in one dimension.

[98] Engelbrecht J.R. and Bedell K.S., *Phys. Rev. Lett.* **74** (1995) 4265.

[99] Bulut N., Scalapino D.J. and White S.R., *Phys. Rev. B* **50** (1994) 9623.

[100] Jackson F., MSc thesis, Université de Sherbrooke (unpublished).

Eur. Phys. J. B **32**, 429–436 (2003)
DOI: 10.1140/epjb/e2003-00121-8

THE EUROPEAN PHYSICAL JOURNAL B

Self-energy-functional approach to systems of correlated electrons

M. Potthoff[a]

Lehrstuhl Festkörpertheorie, Institut für Physik, Humboldt-Universität zu Berlin, 10115 Berlin, Germany

Received 13 January 2003 / Received in final form 5 March 2003
Published online 24 April 2003 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2003

Abstract. The grand potential of a system of interacting electrons is considered as a stationary point of a self-energy functional. It is shown that a rigorous evaluation of the functional is possible for self-energies that are representable within a certain reference system. The variational scheme allows to construct new non-perturbative and thermodynamically consistent approximations. Numerical results illustrate the practicability of the method.

PACS. 71.10.-w Theories and models of many-electron systems – 71.15.-m Methods of electronic structure calculations – 74.20.-z Theories and models of superconducting state – 75.10.-b General theory and models of magnetic ordering – 71.30.+h Metal-insulator transitions and other electronic transitions

1 Introduction

Systems of strongly correlated electrons continue to represent a central subject of current research. Different interesting correlation phenomena, such as high-temperature superconductivity [1], Mott metal-insulator transitions [2] or itinerant ferromagnetism [3], are far from being finally clarified. Progress in this field crucially depends on the development of new theoretical methods as even highly idealized model systems pose notoriously difficult problems. There are only a few general approaches which are able to access the equilibrium thermodynamics as well as excitation properties of an extended system of correlated electrons.

General methods can be based on the Green's-function formalism of Luttinger and Ward [4] and Baym and Kadanoff: [5] Here the grand potential Ω is expressed in terms of the time- or frequency-dependent one-electron Green's function $\mathbf{G}$. The functional $\Omega[\mathbf{G}]$ can be shown to be stationary at the physical $\mathbf{G}$. In principle, this is an exact variational approach which provides information not only on static equilibrium but also on dynamic excitation properties. The functional dependence $\Omega[\mathbf{G}]$, however, is generally not known explicitly as it must be constructed by summation of an infinite series of renormalized skeleton diagrams. In the standard approximation the exact but unknown functional is replaced by an explicitly known but approximate one which is based on an incomplete summation of the diagram series. This leads to the well-known perturbational ("conserving") theories [5]. Higher-order approaches as the fluctuation-exchange approximation [6] are mainly applied to discrete lattice models while for continuum systems, *e.g.* for the inhomogeneous electron gas, one has to be content with lowest-order theories as the GW method [7–9].

A second type of general methods is based on density-functional (DF) approaches [10,11]. Normally these aim at the inhomogeneous electron gas but can also be applied to Hubbard-type lattice models [12]. Compared with the Green's-function formalism, there is a conceptually similar situation for DF approaches: In the latter the ground-state energy E (or the grand potential Ω) [13] is given as a functional of the (static) density $\mathbf{n}$. The variational principle associated with the functional $E[\mathbf{n}]$ is rigorous but cannot be evaluated as $E[\mathbf{n}]$ is generally unknown. In the standard local-density approximation (LDA) the (unknown part of the) functional is replaced by an explicitly known but approximate functional which is taken from the homogeneous system. For systems with weakly varying density the LDA should be justified. Information on excitation properties is contained in dynamic response functions which are in principle accessible via time-dependent DF theory [14] where the action A is considered as a functional of the time-dependent density $\mathbf{n}$. Again, the exact but unknown functional $A[\mathbf{n}]$ is approximated to make it explicit and the variational principle is exploited afterwards.

The method proposed here rests on a variational principle which uses the electron self-energy $\mathbf{\Sigma}$ as the basic dynamic variable. A new functional $\Omega[\mathbf{\Sigma}]$ is constructed which can be shown to be stationary at the physical self-energy. The main result is that the variational principle can be exploited without any approximation of the

[a] *Present address*: Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany
e-mail: potthoff@physik.uni-wuerzburg.de

functional dependence. Namely, a *rigorous* evaluation of the functional $\Omega[\mathbf{\Sigma}]$ is possible on a certain subspace of trial self-energies. Trial self-energies must be representable within an exactly solvable reference system sharing the same interaction with the original system.

This result has important consequences as it opens a route for constructing a novel class of approximations. Although the self-energy essentially contains the same information as the Green's function or the (time-dependent) density, the new approach is conceptually contrary to the Green's-function approach and to the DF approach as there is no approximation to be tolerated for the central functional. Instead of approximating the functional itself, it is considered on a restricted domain. The self-energy-functional approach is completely general and yields approximations which are non-perturbative, thermodynamically consistent and systematic. Opposed to numerical techniques directly applied to systems of finite size, the self-energy-functional approach provides a variational or self-consistent embedding of finite systems and thus yields results in the thermodynamical limit. Such techniques are needed to construct phase diagrams from standard correlated lattice models. A potentially fruitful field of application are systems with competing types of order resulting from spin, charge or orbital correlations as it is typical *e.g.* for numerous transition-metal oxides [1–3].

In the present paper the approach is introduced and a number of general aspects are discussed in detail (Sect. 2). To demonstrate its usefulness, two applications will be considered for the single-band Hubbard model: In Section 3 it is shown that the dynamical mean-field theory (DMFT) [15] can be recovered within the self-energy-functional approach, namely by choosing a decoupled set of impurity Anderson models as a reference system. The DMFT generally requires the treatment of a quantum-impurity problem with an infinite number of degrees of freedom ($n_\mathrm{s} = \infty$). In Section 4 a new approximation is discussed which is based on an impurity model with a *finite* number of degrees of freedom only and which approaches the DMFT for $n_\mathrm{s} = \infty$. The method is closely related to the exact-diagonalization approach (ED) [16,17]. Opposed to the ED, however, thermodynamical consistency is guaranteed at any stage of the approximation. New approaches beyond the mean-field level will be discussed elsewhere. The conclusions and an outlook are given in Section 5.

2 Self-energy-functional approach

Consider a general Hamiltonian $H = H_0(\mathbf{t}) + H_1(\mathbf{U})$ with one-particle ("hopping") parameters $\mathbf{t}$ and two-particle interaction parameters $\mathbf{U}$:

$$H = \sum_{\alpha\beta} t_{\alpha\beta} c_\alpha^\dagger c_\beta + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} U_{\alpha\beta\delta\gamma} c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta \,. \qquad (1)$$

Here $\alpha, \beta, \ldots$ refer to an orthonormal and complete set of one-particle basis states. We are interested in the equilibrium thermodynamics and in elementary one-particle excitations of the system for temperature T and chemical potential μ. This is described by the one-particle Green's function $G_{\alpha\beta}(\mathrm{i}\omega) = \langle\langle c_\alpha ; c_\beta^\dagger \rangle\rangle$ of the imaginary fermionic Matsubara frequencies $\mathrm{i}\omega = \mathrm{i}(2n+1)\pi T$ with integer n [18]. The Green's function can be calculated from the self-energy $\Sigma_{\alpha\beta}(\mathrm{i}\omega)$ *via* the Dyson equation. Using a matrix notation, this reads as $\mathbf{G} = \mathbf{G}_0 + \mathbf{G}_0 \mathbf{\Sigma} \mathbf{G}$ where $\mathbf{G}_0 = 1/(\mathrm{i}\omega + \mu - \mathbf{t})$ is the "free" Green's function. The self-energy is given by $\mathbf{\Sigma} = \mathbf{\Sigma}[\mathbf{G}] = T^{-1} \delta \Phi[\mathbf{G}] / \delta \mathbf{G}$, where $\Phi[\mathbf{G}]$ is the so-called Luttinger-Ward functional [4,5]. This allows to derive the Green's function from a variational principle: One has $\delta \Omega[\mathbf{G}] / \delta \mathbf{G} = 0$ where $\Omega[\mathbf{G}] = \Phi[\mathbf{G}] + \mathrm{Tr} \ln(-\mathbf{G}) - \mathrm{Tr}((\mathbf{G}_0^{-1} - \mathbf{G}^{-1})\mathbf{G})$ and using the notation $\mathrm{Tr}\, \mathbf{A} = T \sum_{\omega,\alpha} A_{\alpha\alpha}(\mathrm{i}\omega)$. In general, however, the functional $\Phi[\mathbf{G}]$ is not known explicitly which prevents an evaluation of $\Omega[\mathbf{G}]$ for a given $\mathbf{G}$. So-called conserving approximations [5] provide an explicit but approximate functional $\Phi_\mathrm{pert.}[\mathbf{G}] \approx \Phi[\mathbf{G}]$. However, these are weak-coupling approaches where a certain subclass of Φ diagrams is summed up.

Here a different but still rigorous variational principle is proposed which is based on a functional $\mathbf{G} = \mathbf{G}[\mathbf{\Sigma}]$ defined as the inverse of $\mathbf{\Sigma} = \mathbf{\Sigma}[\mathbf{G}]$. We can assume the latter to be invertible (locally) provided that the system is not at a critical point for a phase transition (see Appendix A). Consider then:

$$\Omega_\mathbf{t}[\mathbf{\Sigma}] \equiv \mathrm{Tr} \ln(-(\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1}) + F[\mathbf{\Sigma}] \qquad (2)$$

where $F[\mathbf{\Sigma}] \equiv \Phi[\mathbf{G}[\mathbf{\Sigma}]] - \mathrm{Tr}(\mathbf{\Sigma}\, \mathbf{G}[\mathbf{\Sigma}])$ is the Legendre transform of $\Phi[\mathbf{G}]$. The subscript $\mathbf{t}$ indicates the explicit $\mathbf{t}$ dependence of Ω due to the free Green's function $\mathbf{G}_0$. Using $T^{-1} \delta F[\mathbf{\Sigma}] / \delta \mathbf{\Sigma} = \mathbf{G}[\mathbf{\Sigma}]$, one finds that

$$\delta \Omega_\mathbf{t}[\mathbf{\Sigma}] / \delta \mathbf{\Sigma} = 0 \;\Leftrightarrow\; \mathbf{G}[\mathbf{\Sigma}] = (\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1} \,. \qquad (3)$$

Thus $\Omega_\mathbf{t}[\mathbf{\Sigma}]$ is stationary at the exact (physical) self-energy and its value is the exact grand potential of the system. Again, the problem is that the functional $\Omega_\mathbf{t}[\mathbf{\Sigma}]$ is in general not known explicitly.

As the domain of the self-energy functional $\Omega_\mathbf{t}[\mathbf{\Sigma}]$ we define the class of all $\mathbf{t}'$ representable self-energies. $\mathbf{\Sigma}$ is termed $\mathbf{t}'$ representable, if there is a set of hopping parameters $\mathbf{t}'$ such that $\mathbf{\Sigma}$ is the exact self-energy of the model $H_0(\mathbf{t}') + H_1(\mathbf{U})$. This implies that any self-energy in the domain of $\Omega_\mathbf{t}[\mathbf{\Sigma}]$ can be parameterized as $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{t}')$. The interaction parameters $\mathbf{U}$ are taken to be fixed. Suppose we are interested in the model $H = H_0(\mathbf{t}) + H_1(\mathbf{U})$. Then the function $\Omega_\mathbf{t}(\mathbf{t}') \equiv \Omega_\mathbf{t}[\mathbf{\Sigma}(\mathbf{t}')]$ is stationary at $\mathbf{t}' = \mathbf{t}$. Thus $\partial \Omega_\mathbf{t}(\mathbf{t}') / \partial \mathbf{t}' = 0$ at $\mathbf{t}' = \mathbf{t}$.

It is important to note that $F[\mathbf{\Sigma}]$ is universal: The functional dependence is the same for any $\mathbf{t}$, *i.e.* it remains unchanged for an arbitrary reference system H' with the same interaction but modified hopping parameters: $H' = H_0(\mathbf{t}') + H_1(\mathbf{U})$. $F[\mathbf{\Sigma}]$ is universal as it is the Legendre transform of $\Phi[\mathbf{G}]$ which in turn is universal because it can be constructed formally as the sum of all closed, irreducible, and renormalized skeleton diagrams which, apart from $\mathbf{G}$, include the vertices $\mathbf{U}$ only.

Consequently, one has:

$$\Omega_{\mathbf{t}'}[\mathbf{\Sigma}] = \mathrm{Tr}\ln(-(\mathbf{G}_0'^{-1} - \mathbf{\Sigma})^{-1}) + F[\mathbf{\Sigma}]\,, \qquad (4)$$

for the reference system H' with $\mathbf{G}_0'^{-1} = \mathrm{i}\omega + \mu - \mathbf{t}'$.

The self-energy functionals $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$, equation (2), and $\Omega_{\mathbf{t}'}[\mathbf{\Sigma}]$, equation (4), are different if $\mathbf{t} \neq \mathbf{t}'$, *i.e.* if the original and the reference system are different. $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$ is stationary at $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{t})$ while $\Omega_{\mathbf{t}'}[\mathbf{\Sigma}]$ is stationary at $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{t}')$. The important point is that both functionals share the same (unknown but universal) part $F[\mathbf{\Sigma}]$ and that both functionals are defined globally. Therefore, combining equations (2) and (4), $F[\mathbf{\Sigma}]$ can be eliminated:

$$\Omega_{\mathbf{t}}[\mathbf{\Sigma}] = \Omega_{\mathbf{t}'}[\mathbf{\Sigma}] + \mathrm{Tr}\ln(-(\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1}) - \mathrm{Tr}\ln(-(\mathbf{G}_0'^{-1} - \mathbf{\Sigma})^{-1})\,. \qquad (5)$$

This form of the self-energy functional $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$ is still exact.

In the next step the functional has to be evaluated to search for a stationary point. Evaluating $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$ for self-energies taken from the reference system, *i.e.* for self-energies parameterized as $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{t}')$, one obtains:

$$\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')] = \Omega' + \mathrm{Tr}\ln(-(\mathbf{G}_0^{-1} - \mathbf{\Sigma}(\mathbf{t}'))^{-1}) - \mathrm{Tr}\ln(-\mathbf{G}')\,. \qquad (6)$$

Here it has been used that $\Omega_{\mathbf{t}'}[\mathbf{\Sigma}(\mathbf{t}')] = \Omega'$, the exact grand potential of the reference system H', and $(\mathbf{G}_0'^{-1} - \mathbf{\Sigma}(\mathbf{t}'))^{-1} = \mathbf{G}'$, the exact Green's function of H'. Suppose that the reference system H' is much simpler than the original system H so that it can be solved exactly for any $\mathbf{t}'$ belonging to a certain subspace of the entire space of hopping parameters. The resulting equation (6) is remarkable, as it shows that the functional $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$ can be evaluated rigorously for trial self-energies $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{t}')$ taken from the reference system H'.

This is the main result. Contrary to previous approaches (*e.g.* conserving theories, LDA), there is no need to approximate the functional dependence in a fundamental variational principle. Approximations are constructed by searching for a stationary point of $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$ on a *restricted* set of trial self-energies $\mathbf{\Sigma}(\mathbf{t}')$.

The stationary point is determined by the Euler equation: $\partial\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')]/\partial\mathbf{t}' = 0$. Calculating the derivative,

$$T\sum_{\omega}\sum_{\alpha\beta}\left(\frac{1}{\mathbf{G}_0^{-1} - \mathbf{\Sigma}(\mathbf{t}')} - \mathbf{G}'\right)_{\beta\alpha}\frac{\partial\Sigma_{\alpha\beta}(\mathbf{t}')}{\partial\mathbf{t}'} = 0. \qquad (7)$$

Note that the equation involves, apart from $\mathbf{G}_0$, quantities of the reference system H' only. The linear response of the self-energy of H' due to a change of the hopping $\mathbf{t}'$ can be calculated along the lines of reference [5]. It turns out that $\partial\mathbf{\Sigma}(\mathbf{t}')/\partial\mathbf{t}'$ is given by a two-particle Green's function of H'. Since $\mathbf{G}' = \mathbf{G}[\mathbf{\Sigma}(\mathbf{t}')]$, the exact self-energy of the system H is determined by the condition that the bracket in (7) be zero. Hence, one can consider equation (7) to be obtained from the *exact* equation that determines the "vector" $\mathbf{\Sigma}$ in the self-energy space through *projection* onto the hypersurface of $\mathbf{t}'$ representable trial self-energies $\mathbf{\Sigma}(\mathbf{t}')$ by taking the scalar product with vectors $\partial\mathbf{\Sigma}(\mathbf{t}')/\partial\mathbf{t}'$ tangential to the hypersurface.

An analysis of the second derivative $\partial^2\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')]/\partial t'_{\alpha\beta}\partial t'_{\gamma\delta}$ shows that a stationary point is not an extremum point in general. This feature is shared with the time-dependent DF approach [14], the Green's-function approach [5] and also with a recently considered variant [19]. Only in the static DF theory there is a convex (density) functional [10,11,13]. Nevertheless, the proposed self-energy-functional approach is systematic: For any sequence of reference systems H' including more and more degrees of freedom and converging to the original system H there is, from the variational principle, a corresponding sequence of grand potentials which must converge to the exact $\Omega = \Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t})]$ as the subspace of trial self-energies increases and eventually includes the exact self-energy $\mathbf{\Sigma}(\mathbf{t})$.

It is also possible to build the theory on the Green's function $\mathbf{G}$ instead of the self-energy $\mathbf{\Sigma}$ as the basic variable. One may start *e.g.* from the Green's-function functional $\Omega_{\mathbf{t}}[\mathbf{G}]$ of reference [19]. This can be split into a universal part and a part depending on the hopping explicitly. Introducing the concept of the reference system and using arguments similar to those that have lead to equations (4, 5), and (6), one can show that an exact evaluation of $\Omega_{\mathbf{t}}[\mathbf{G}]$ is possible on a certain subspace of $\mathbf{t}'$ representable Green's functions. Here it is preferred to use the self-energy as the basic variable, however, as it is advantageous to approximate an irreducible rather than a reducible quantity.

3 Relation to the DMFT

Given an original model H, what could a suitable reference system H' look like? Consider, for example, H to be the Hubbard model [20] which is shown in Figure 1a schematically: A filled dot represents a correlated site i with on-site Hubbard interaction U, and a line connecting two sites i and j represents the nearest-neighbor hopping $t_{i,j}$. The number of sites is $L \mapsto \infty$. Figure 1c shows a conceivable reference system H'. H' is obtained from H (Fig. 1a) by (i) adding to each correlated site i a number of $n_\mathrm{s} - 1$ uncorrelated ("bath") sites $k = 2, ..., n_\mathrm{s}$ (open dots) which are disconnected from the rest of the system, by (ii) switching off the hopping $t_{i,j}$ between the correlated sites and (iii) switching on a hopping $V_{i,k}$ to the bath sites. After step (i) the Hamiltonian Figure 1b (in the figure $n_\mathrm{s} = 5$) has an enlarged Hilbert space but the same self-energy. It is important to note that steps (i) - (iii) leave the interaction part unchanged and thus preserve the functional dependence $F[\mathbf{\Sigma}]$. Actually, the system H' is a set of L decoupled single-impurity Anderson models (SIAM) [21] with n_s sites each. Compared to H, the problem posed by H' is strongly simplified. This is achieved at the cost of restricting the set of trial self-energies. In particular, as the correlated sites are decoupled in H', the trial self-energies are local: $\Sigma_{ij}(\mathrm{i}\omega, \mathbf{t}') \propto \delta_{ij}$. One has to consider H' for arbitrary one-particle parameters, namely the on-site energies of the correlated ("c") and of the bath sites ("a"), $\epsilon_i^{(c)}$ and $\epsilon_{i,k}^{(a)}$, respectively, and the hopping ("hybridization") $V_{i,k}$

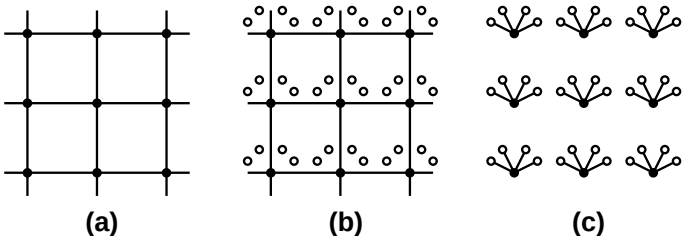

Fig. 1. Schematic representation of the Hubbard model H (a), an equivalent model (b), and a possible reference system H' (c). See text for discussion.

between them and take these as variational parameters in the principle $\delta\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')] = 0$.

Let us discuss the case $n_{\mathrm{s}} \mapsto \infty$. For a homogeneous phase of the (translationally invariant) original system, $\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')]$ will be stationary at a homogeneous set of variational parameters: $\mathbf{t}' = \{\epsilon_i^{(c)}, \epsilon_{i,k}^{(a)}, V_{i,k}\} = \{\epsilon^{(c)}, \epsilon_k^{(a)}, V_k\}$. Consequently, it is sufficient to consider one SIAM only. As the different equivalent SIAM's are spatially decoupled, not only the self-energy but also its linear response is local: $\partial\Sigma_{ij}(\mathbf{t}')/\partial\mathbf{t}' \propto \delta_{ij}$. To solve the Euler equation (7), it is thus sufficient to fulfill the "locally projected" equation

$$\left(\frac{1}{\mathbf{G}_0^{-1}(\mathrm{i}\omega) - \mathbf{\Sigma}(\mathrm{i}\omega)}\right)_{ii} = G'_{ii}(\mathrm{i}\omega)\,. \tag{8}$$

This is just the self-consistency equation of the DMFT [15]: the SIAM parameters have to be found such that the on-site ("impurity") Green's function at a correlated site i coincides with the on-site Green's function of the Hubbard model which is calculated from $\mathbf{G}_0$ and the ("impurity") self-energy of H' by means of the Dyson equation. Therefore, one can state that the DMFT (as an approximation for any finite-dimensional system or as the exact theory in infinite dimensions) is recovered as a stationary point of $\Omega_{\mathbf{t}}[\mathbf{\Sigma}]$ when restricting the search to local self-energies representable by a SIAM.

Within the DMFT the computation of the self-energy requires an iterative procedure: $\mathbf{\Sigma}_{\mathrm{old}} \mapsto \mathbf{\Sigma}_{\mathrm{new}}$. Here it turns out that this corresponds to a certain (discrete) path on the hypersurface of SIAM trial self-energies. Convergence of the iteration ($\mathbf{\Sigma}_{\mathrm{old}} = \mathbf{\Sigma}_{\mathrm{new}}$), however, is by no means guaranteed physically but depends on the contracting properties of the map $\mathbf{\Sigma}_{\mathrm{old}} \mapsto \mathbf{\Sigma}_{\mathrm{new}}$. The self-energy-functional approach offers an alternative as instead of solving equation (8) one may calculate $\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')]$ by equation (6) and determine the stationary point. Hence, the DMFT can also be obtained by a more *direct* computation avoiding any iterations – similar (in this respect) to the random-dispersion approximation [22]. Note that in case of more than a single stationary point there is also an equally direct access to metastable phases.

For any inhomogeneous situation, equation (8) represents a system of self-consistency equations to fix the parameters of non-equivalent impurity models labeled by the site index i. The models can be solved independently but are coupled indirectly due to the matrix inversion in (8). This exactly recovers the DMFT generalized to systems with reduced translational symmetry [23,24].

4 A consistent ED method

A brief discussion of two limiting cases of the Hubbard model may be instructive. Consider the band limit with $U = 0$ first. Here $H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma}$ describes a system of non-interacting electrons. This case is exceptional as obviously the functional $F[\mathbf{\Sigma}] \equiv 0$ and therefore $\Omega_{\mathbf{t}}[\mathbf{\Sigma}] = \mathrm{Tr}\ln(-(\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1})$. Any valid reference system H' must have the same (*i.e.* a vanishing) interaction part as H, and thus $\mathbf{\Sigma}(\mathbf{t}') \equiv 0$ and $\Omega_{\mathbf{t}}[\mathbf{\Sigma}(\mathbf{t}')] = \mathrm{Tr}\ln(-\mathbf{G}_0^{-1})$, the exact grand potential for non-interacting electrons.

The atomic limit, $H = \sum_{i\sigma}(t_0 c_{i\sigma}^\dagger c_{i\sigma} + (U/2) n_{i\sigma} n_{i-\sigma})$ is more interesting as $\Phi[\mathbf{G}]$ and $F[\mathbf{\Sigma}]$ cannot be constructed explicitly. Within the self-energy-functional approach one has to compute $\mathbf{\Sigma}(\mathbf{t}')$, $\mathbf{G}'$, and Ω' for a suitable reference system H' and to insert into equation (6) for optimization. The only meaningful choice for the reference system is $H' = H$ in this case. Obviously, this yields the exact solution. Generally, whenever the original model H is exactly solvable, the choice $H' = H$ will do.

For a non-trivial model H, the choice $H' = H$ is useless for a practical computation. Any simplified reference system, however, yields a consistent approximation. The case of the DMFT discussed in Section 3 is an illustrative example. On the other hand, in the context of the DMFT actually both, H and H', are highly non-trivial models, and further approximations or large-scale numerics are needed to treat the reference system H'. More simple approximations for the Hubbard model which are still consistent are generated by considering reference systems with a *finite* number of degrees of freedom. The reference system of Figure 1c with $n_{\mathrm{s}} < \infty$ is an interesting example which shall be discussed in the following. For small n_{s} one can easily obtain numerical results as a complete diagonalization of H' is feasible.

Consider the Hubbard model

$$H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \frac{U}{2} \sum_{i\sigma} n_{i\sigma} n_{i-\sigma} \tag{9}$$

at temperature $T = 0$ and chemical potential $\mu = U/2$. For symmetric conditions this implies half-filling. The Hamiltonian of the reference system is given by $H' = \sum_i H'(i)$ with

$$H'(i) = \sum_\sigma \epsilon_i^{(c)} c_{i\sigma}^\dagger c_{i\sigma} + \frac{U}{2} \sum_\sigma n_{i\sigma} n_{i-\sigma} + \sum_{\sigma,k=2}^{n_{\mathrm{s}}} \epsilon_{i,k}^{(a)} a_{ik\sigma}^\dagger a_{ik\sigma} + \sum_{\sigma,k} \left(V_{i,k} c_{i\sigma}^\dagger a_{ik\sigma} + \mathrm{h.c.}\right)\,. \tag{10}$$

For the sake of simplicity we consider a homogeneous paramagnetic phase and the most simple case $n_{\mathrm{s}} = 2$ where one is left with three independent variational parameters only,

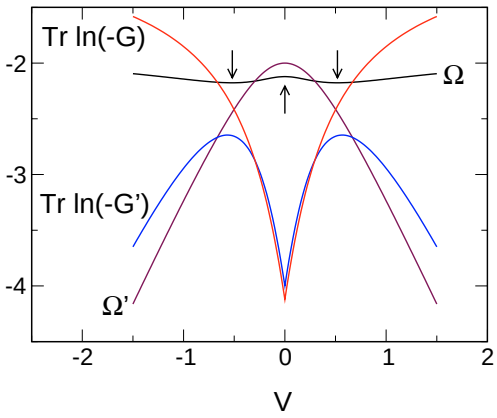

Fig. 2. Grand potential Ω (per lattice site) and the different contributions (per lattice site) according to equation (6): Ω', $\mathrm{Tr}\ln(-(\mathbf{G}_0^{-1} - \mathbf{\Sigma}(\mathbf{t}'))^{-1})$, and $\mathrm{Tr}\ln(-\mathbf{G}')$ for $U = 4$, $T = 0$, and $\mu = U/2$ (half-filling) as functions of V. Stationary points (arrows) at $V = \pm 0.519$ (metal) and at $V = 0$ (insulator). $\epsilon_c = 0$ and $\epsilon_a = 2$.

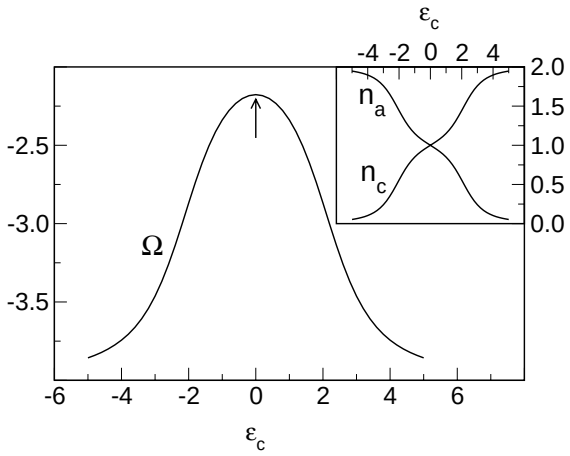

Fig. 3. Ω as a function of ϵ_c for $V = 0.519$ (metal) and $\epsilon_\mathrm{a} = 2$. $U = 4$. Inset: impurity- and bath-orbital filling, n_c and n_a, as functions of ϵ_c.

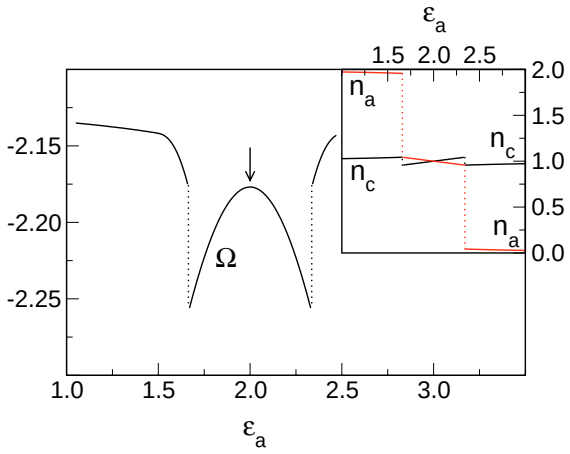

Fig. 4. Ω and the impurity- and bath-orbital filling, n_c and n_a, as functions of ϵ_a for $V = 0.519$ and $\epsilon_\mathrm{c} = 0$. $U = 4$.

namely $V \equiv V_{i,k=2}$, $\epsilon_a \equiv \epsilon^{(a)}_{i,k=2}$, and $\epsilon_c \equiv \epsilon^{(c)}_i$. The computation of the different contributions to the grand potential, equation (6), is straightforward: Diagonalization of H' yields the ground-state energy E'_0 and $\Omega' = E'_0 - \mu\langle N'\rangle$ as well as the excitation energies, the ground state and the excited states. The Green's function $\mathbf{G}'$ and the free Green's function $\mathbf{G}'_0$ can be computed from their respective Lehmann representations. The self-energy of the reference system is obtained as $\mathbf{\Sigma}(\mathbf{t}') = \mathbf{G}'^{-1}_0 - \mathbf{G}'^{-1}$. Since the self-energy is local, as in the DMFT, the lattice structure enters via the free ($U = 0$) density of states only. Therefore, the $\mathbf{k}$-sum which appears in the first trace in equation (6) can be performed conveniently by a one-dimensional density-of-states integration. A semi-elliptical free density of states with the band width $W = 4$ is used for the calculations. This sets the energy scale for the results discussed below.

Figure 2 shows the grand potential Ω and the three different contributions as functions of V. The interaction is kept fixed at $U = W = 4$. The remaining variational parameters are set to $\epsilon_c = 0$ and $\epsilon_a = U/2 = 2$, as required by particle-hole symmetry. Each of the three contributions strongly depends on V and none of them has a stationary point at a finite $V \neq 0$. Two of them show a singular behavior at $V = 0$. Contrary, the resulting Ω is regular for any V and shows a much weaker V dependence. There are three stationary points which are indicated by the arrows. The maximum at $V = 0$ corresponds to an insulating phase since $\Sigma(\omega)$ for $n_\mathrm{s} = 2$ and $V = 0$ is the Hubbard-I self-energy which implies a vanishing spectral density $-(1/\pi)\mathrm{Im}\mathbf{G}(\omega + \mathrm{i}0^+)$ at $\omega = 0$. The minima at $V = \pm 0.519$ correspond to a metallic phase. Ω as well as the different contributions are symmetric functions of V. As $\Sigma(V, \omega) = \Sigma(-V, \omega)$, however, this symmetry is trivial and does not yield an additional physical phase (see also Appendix A). Due to the lower Ω the metallic phase is stable as compared to the insulating one.

The minimum at $V = 0.519$ is actually a saddle point if the entire space of variational parameters is considered. This is demonstrated by Figure 3 which shows Ω as a function of ϵ_c for fixed $V = 0.519$ and $\epsilon_a = 2$. While $\Omega(V)$ is at a minimum for $V = 0.519$, $\Omega(\epsilon_c)$ is at a *maximum* for $\epsilon_c = 0$. In the (V, ϵ_c) space one therefore encounters a saddle point. As already noted in Section 2, there is no reason to expect an extremum in general. It is worth mentioning that stationarity at $\epsilon_c = 0$ is consistent with the requirements of particle-hole symmetry. For any $\epsilon_c \neq 0$ the impurity model is asymmetric. This can be seen from the inset where the average occupations of the impurity and of the bath site are plotted as functions of ϵ_c. The total particle number $\langle N'\rangle = \sum_\sigma(\langle c^\dagger_\sigma c_\sigma\rangle + \langle a^\dagger_\sigma a_\sigma\rangle) = n_\mathrm{c} + n_\mathrm{a}$ (i and $k = 2$ fixed) is constant: $\langle N'\rangle = 2$.

With respect to the third variational parameter ϵ_a, the grand potential Ω is at a maximum for $\epsilon_a = 2 = \mu$, see Figure 4. Again, this value is required by particle-hole symmetry. If ϵ_a exceeds a certain critical value (away from the stationary point), the ground-state of the reference system H' no longer lies within $N' = 2$ subspace

but is found in the $N' = 1$ or $N' = 3$ subspace, respectively (see inset of Fig. 4). While Ω' is continuous at the level crossing, the symmetry of the ground state changes. Consequently, there is a discontinuous change of the trial self-energy which implies a discontinuous change of Ω.

Consider now the *original* model at slightly modified parameters, *e.g.* U, μ, or T. Clearly, the stationary point of Ω will be expected then at slightly different values of the variational parameters V, ϵ_c, ϵ_a. This implies that all physical quantities which derive from the thermodynamical potential Ω will be continuous functions of the (original) model parameters in general – irrespective of the fact that the reference system includes a finite number of degrees of freedom only: It is a typical feature of any mean-field approach that results are directly provided for the thermodynamical limit. A discontinuous jump of Ω due to a symmetry change of the ground state of the reference system (see Fig. 4) usually occurs away from stationarity and is thus irrelevant physically. It is conceivable, however, that the stationary point moves to a point of discontinuity as a function a parameter of the original model. In this case the approach would generate an artifact which is a reminiscence of the finiteness of H'.

So far we discussed the case $U = 4 = W$ only. As a function of U the half-filled paramagnetic Hubbard model at $T = 0$ is expected to undergo a transition from a metal to a Mott-Hubbard insulator [2]. This is marked by a divergence of the effective mass or, equivalently, by a vanishing quasi-particle weight $z = 1/(1 - \mathrm{d}\Sigma(\omega)/\mathrm{d}\omega|_{\omega=0})$ as U approaches a critical value U_c from below [15,25]. The result for $z(U)$ as obtained by the use of the two-site reference system ("dynamical impurity approximation", DIA with $n_\mathrm{s} = 2$) is shown in Figure 5. As there are less degrees of freedom contained in H', the approximation should be considered to be *inferior* as compared to the full DMFT the results of which are in Figure 5, too. It is remarkable that the simple $n_\mathrm{s} = 2$-DIA (which requires the diagonalization of a dimer model only) yields $z(U)$ in an almost quantitative agreement with the full DMFT.

The results of the $n_\mathrm{s} = 2$-DIA may also be compared with those of the recently developed "linearized" or "two-site" DMFT [26,28] where the Hubbard model is mapped onto the two-site SIAM (10) by means a strongly simplified self-consistency condition. As compared to the two-site DMFT, the present self-energy-functional approach not only represents a clear conceptual improvement but also improves the actual results for $z(U)$ and for U_c (see Fig. 5 and note that $U_\mathrm{c} = 6$ within the linearized DMFT, $U_\mathrm{c} = 5.8450$ within the DIA for $n_\mathrm{s} = 2$, and $U_\mathrm{c} = 5.84$ and $U_\mathrm{c} = 5.88$ from numerical evaluations [27,29] of the full DMFT).

The self-energy-functional approach with reference system of Figure 1c and small n_s actually represents a new variant of the DMFT-exact-diagonalization method (ED) [16,17]. As compared to previous formulations of the ED, the convergence with respect to n_s appears to be faster: Compare the results for $n_\mathrm{s} = 2$, $n_\mathrm{s} = 4$ and $n_\mathrm{s} = \infty$ (full DMFT) in Figure 5 with those of reference [16]. More important, however, there is no need for a fit procedure in the present approach; any arbitrariness in the method to find the SIAM parameters is avoided completely. Furthermore, consistent results will be obtained for any finite n_s while in the usual ED this can be expected in the DMFT limit $n_\mathrm{s} \mapsto \infty$ only.

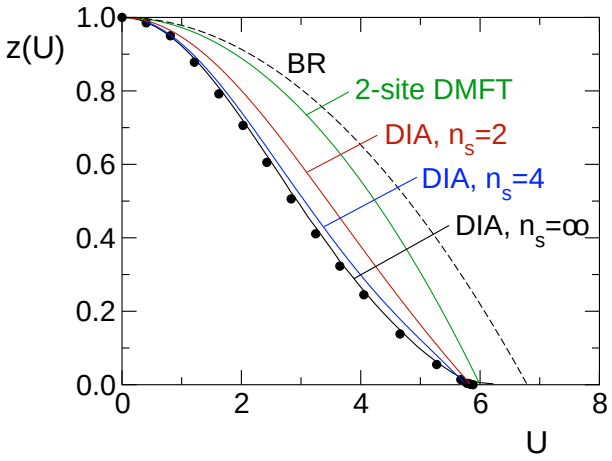

Fig. 5. U dependence of the quasi-particle weight $z = 1/(1 - \mathrm{d}\Sigma(\omega = 0)/\mathrm{d}\omega)$ within different approximations. BR: Brinkman-Rice (Gutzwiller) approach [25]. 2-site DMFT: a non-variational two-site approach [26]. DIA, $n_\mathrm{s} = 2$: self-energy-functional approach with a reference system H' consisting of decoupled two-site impurity models. DIA, $n_\mathrm{s} = 4$: dynamical-impurity approximation with $n_\mathrm{s} = 4$ sites. DIA, $n_\mathrm{s} = \infty$: DMFT limit (circles: NRG [27], line: ED using 8 sites).

5 Conclusions and outlook

Concluding, the proposed self-energy-functional method is a systematic scheme for the construction of new non-perturbative and consistent approximations for extended systems of interacting fermions. For Hubbard-type lattice models with on-site interaction several relations to and generalizations of existing approaches are obtained immediately. The numerical results obtained by considering a rather simple reference system clearly demonstrate the practicability of the theory. Its generality promises that the approach may successfully be applied also in different contexts.

For a Hubbard-type system including $M > 1$ orbitals per site, a consistent DMFT can only be defined when using M baths. There is no such necessity within the self-energy-functional approach. While clearly the optimal local approximation requires M baths, any $M' < M$ will nevertheless lead to a fully consistent approximation. This represents an interesting option for numerical studies of multi-band systems. Non-local trial self-energies can be constructed by grouping the sites into identical clusters of finite size N_s, switching off the inter-cluster hopping and treating the intra-cluster hopping as variational parameters. Each of the N_s sites in a cluster can be coupled to $n_\mathrm{s} - 1$ additional bath sites. The relation of such an approach to cluster extensions of the DMFT [30,31] and to

the cluster perturbation theory [32–34] will be the subject of a forthcoming paper. Extensions and applications of the method to continuous models (inhomogeneous electron gas) and to Bose systems deserve further investigations.

The work is supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 290).

Appendix A: The functional $\mathbf{G}[\mathbf{\Sigma}]$

For the definition of $\mathbf{G}[\mathbf{\Sigma}]$, invertibility of the functional $\mathbf{\Sigma}[\mathbf{G}]$ is required. The *local* invertibility of $\mathbf{\Sigma}[\mathbf{G}]$ is controlled by the Jacobian $\Gamma_{\alpha\beta'\alpha'\beta}(\mathrm{i}\omega,\mathrm{i}\omega') = \delta\Sigma_{\alpha\beta}(\mathrm{i}\omega)/\delta G_{\alpha'\beta'}(\mathrm{i}\omega')$. The two-particle self-energy [5] $\mathbf{\Gamma} = \delta\mathbf{\Sigma}/\delta\mathbf{G}$ can be assumed to be non-singular in general.

For a further analysis we need the following *lemma:* Consider the interaction $\mathbf{U}$ to be fixed. Then two different Green's functions $\mathbf{G}_1$ and $\mathbf{G}_2$ must result from two different sets of one-particle parameters $\mathbf{t}'_1$ and $\mathbf{t}'_2$, respectively. The *proof* is straightforward: Consider the high-frequency expansion of the Green's function $G_{\alpha\beta}(\omega) = \sum_{n=1}^{\infty} M^{(n)}_{\alpha\beta}\omega^{-n}$. The coefficients are given by the moments $M^{(n)}_{\alpha\beta} = \int \mathrm{d}\omega\ \omega^n(-1/\pi)\,\mathrm{Im}\,G_{\alpha\beta}(\omega+\mathrm{i}0^+) = \langle[\mathcal{L}^n c_\alpha, c_\beta]_+\rangle$ with $\mathcal{L}\mathcal{O} \equiv [\mathcal{O}, H]_-$. Using the symmetry $U_{\alpha\beta\gamma\delta} = U_{\beta\alpha\delta\gamma}$ one has:

$$G_{\alpha\beta}(\omega) = \delta_{\alpha\beta}\frac{1}{\omega} + \Big(t'_{\alpha\beta} + \sum_{\gamma\delta}(U_{\alpha\gamma\beta\delta} - U_{\alpha\gamma\delta\beta})\,\langle c^\dagger_\gamma c_\delta\rangle\Big)\frac{1}{\omega^2} + \mathcal{O}(\omega^{-3}). \quad \text{(A.1)}$$

Now $\mathbf{t}'_1 \neq \mathbf{t}'_2$ implies the ω^{-2} coefficients to be different because

$$\langle c^\dagger_\alpha c_\beta\rangle = -\frac{1}{\pi}\int_{-\infty}^{\infty}\mathrm{d}\omega\,\frac{1}{\mathrm{e}^{\omega/T}+1}\,\mathrm{Im}\,G_{\beta\alpha}(\omega+\mathrm{i}0^+). \quad \text{(A.2)}$$

Consequently, we must have $\mathbf{G}_1 \neq \mathbf{G}_2$.

The lemma shows that the relation $\mathbf{t}' \leftrightarrow \mathbf{G}$ is one-to-one. Consequently, we can write $\mathbf{\Sigma}[\mathbf{G}] = \mathbf{\Sigma}[\mathbf{t}'[\mathbf{G}]]$ and $\mathbf{\Gamma} = \delta\mathbf{\Sigma}/\delta\mathbf{G} = \delta\mathbf{\Sigma}/\delta\mathbf{t}'\cdot\delta\mathbf{t}'/\delta\mathbf{G}$ with a non-singular Jacobian $\delta\mathbf{t}'/\delta\mathbf{G}$. Hence, a singular $\mathbf{\Gamma} = \delta\mathbf{\Sigma}/\delta\mathbf{G}$ implies a singular $\delta\mathbf{\Sigma}/\delta\mathbf{t}'$ and *vice versa*. However, $\delta\mathbf{\Sigma}/\delta\mathbf{t}'$ is just the "projector" in the Euler equation (7). We conclude that local non-invertibility of the functional $\mathbf{\Sigma}[\mathbf{G}]$ at $\mathbf{G} = \mathbf{G}(\mathbf{t}')$ is indicated by $\partial\mathbf{\Sigma}[\mathbf{t}']/\partial t'_{\mathbf{n}} = 0$ with $t'_{\mathbf{n}} = \mathbf{t}'\cdot\mathbf{n}$ for a certain "direction" $\mathbf{n}$ in the space of hopping parameters. For such a direction, the Euler equation (7) would be satisfied *trivially*.

Referring to the present numerical results, one can state that generally the projector $\delta\mathbf{\Sigma}/\delta\mathbf{t}'$ is non-singular in fact, as has been expected. There is one exception, however, namely points in the hopping-parameter space where one or more bath sites are *decoupled* from the rest of the system (vanishing hybridization V). Here the one-particle energy of a decoupled bath site can be varied without changing the trial self-energy. Even for this case, however, there are no formal difficulties with the inverse functional $\mathbf{G}[\mathbf{\Sigma}]$: To ensure the local invertibility of $\mathbf{\Sigma}[\mathbf{G}]$, one simply has to restrict the space of variational parameters $\mathbf{t}'$ by excluding the one-particle energies of the decoupled bath sites, *i.e.* one has to focus on the physically relevant parameters. This implies a respective restriction of the space of $\mathbf{t}'$-representable Green's functions $\mathbf{G}(\mathbf{t}')$ and ensures the local invertibility of $\mathbf{\Sigma}[\mathbf{G}]$ on the restricted domain. Similarly, a restriction of the $\mathbf{t}'$ space becomes necessary to ensure the local invertibility of $\mathbf{\Sigma}[\mathbf{G}]$ in case of a system where the self-energy is trivial (as *e.g.* for a model of spinless fermions with nearest-neighbor Coulomb interaction in the limit of infinite spatial dimensions where the self-energy is given by the Hartree term, *cf.* Ref. [35]).

Finally, it should be mentioned that generally the functional $\mathbf{\Sigma}[\mathbf{G}]$ cannot be inverted *globally*. Consider, for example, the Hubbard model on the infinite-dimensional hypercubic lattice with nearest-neighbor hopping t at half-filling. Due to manifest particle-hole symmetry, a sign change of the hopping $t \mapsto -t$ leaves the (local) self-energy invariant but transforms (the non-local elements of) the Green's function $\mathbf{G}$. We conclude that $\mathbf{G}[\mathbf{\Sigma}]$ in general cannot be defined uniquely. Due to this non-uniqueness, and also due to non-linearity, there may be more than a single solution of equation (3). However, this does not cause any problem since for any $\mathbf{\Sigma}$ satisfying (3) we have:

$$\begin{aligned} & \mathbf{G}[\mathbf{\Sigma}] = (\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1} \\ \Rightarrow\quad & \mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1} \\ \Rightarrow\quad & \mathbf{\Sigma} = \mathbf{\Sigma}[\mathbf{G}] \quad \text{and} \quad \mathbf{G} = (\mathbf{G}_0^{-1} - \mathbf{\Sigma})^{-1}. \end{aligned} \quad \text{(A.3)}$$

This means that $\mathbf{\Sigma}$ is given by the (formal) sum of all skeleton diagrams built up by a propagator $\mathbf{G}$ which is constructed via the Dyson equation from the same $\mathbf{\Sigma}$ in turn. Hence, any stationary point should be regarded as a physically meaningful solution. Among different physical solutions (corresponding *e.g.* to different phases) the minimum grand potential selects the stable one.

References

1. J. Orenstein, A.J. Millis, Science **288**, 468 (2000)
2. F. Gebhard, *The Mott Metal-Insulator Transition* (Springer, Berlin, 1997)
3. *Band-ferromagnetism*, edited by K. Baberschke, M. Donath, W. Nolting (Springer, Berlin, 2001)
4. J.M. Luttinger, J.C. Ward, Phys. Rev. **118**, 1417 (1960)
5. G. Baym, L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)
6. N.E. Bickers, D.J. Scalapino, S.R. White, Phys. Rev. Lett. **62**, 961 (1989)
7. G. Strinati, H.J. Mattausch, W. Hanke, Phys. Rev. Lett. **45**, 290 (1980)
8. G. Strinati, H.J. Mattausch, W. Hanke, Phys. Rev. B **25**, 2867 (1982)
9. F. Aryasetiawan, O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998)
10. P. Hohenberg, W. Kohn, Phys. Rev. **136**, 864 (1964)

11. W. Kohn, L.J. Sham, Phys. Rev. **140**, 1133 (1965)
12. K. Schönhammer, O. Gunnarsson, R.M. Noack, Phys. Rev. B **52**, 2504 (1995)
13. N.D. Mermin, Phys. Rev. **137**, A 1441 (1965)
14. E. Runge, E.K.U. Gross, Phys. Rev. Lett. **52**, 997 (1984)
15. A. Georges, G. Kotliar, W. Krauth, M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
16. M. Caffarel, W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
17. Q. Si, M.J. Rozenberg, G. Kotliar, A.E. Ruckenstein, Phys. Rev. Lett. **72**, 2761 (1994)
18. A.A. Abrikosow, L.P. Gorkov, I.E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey, 1964)
19. R. Chitra, G. Kotliar, Phys. Rev. B **63**, 115110 (2001)
20. J. Hubbard, Proc. R. Soc. London A **276**, 238 (1963); M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963); J. Kanamori, Prog. Theor. Phys. (Kyoto) **30**, 275 (1963)
21. P.W. Anderson, Phys. Rev. **124**, 41 (1961)
22. R.M. Noack, F. Gebhard, Phys. Rev. Lett. **82**, 1915 (1999)
23. M. Potthoff, W. Nolting, Phys. Rev. B **59**, 2549 (1999)
24. M. Potthoff, W. Nolting, Phys. Rev. B **60**, 7834 (1999)
25. W.F. Brinkman, T.M. Rice, Phys. Rev. B **2**, 4302 (1970)
26. M. Potthoff, Phys. Rev. B **64**, 165114 (2001)
27. R. Bulla, Phys. Rev. Lett. **83**, 136 (1999)
28. R. Bulla, M. Potthoff, Eur. Phys. J. B **13**, 257 (2000)
29. G. Moeller, Q. Si, G. Kotliar, M. Rozenberg, D.S. Fisher, Phys. Rev. Lett. **74**, 2082 (1995)
30. M.H. Hettler, A.N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, H.R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998)
31. G. Kotliar, S.Y. Savrasov, G. Pálsson, G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001)
32. C. Gros, R. Valenti, Phys. Rev. B **48**, 418 (1993)
33. D. Sénéchal, D. Pérez, M. Pioro-Ladrière, Phys. Rev. Lett. **84**, 522 (2000)
34. D. Sénéchal, D. Pérez, D. Plouffe, Phys. Rev. B **66**, 075129 (2002)
35. E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989)

arXiv:cond-mat/0406671 v1 28 Jun 2004

Non-perturbative construction of the Luttinger-Ward functional

Michael Potthoff
Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Germany

For a system of correlated electrons, the Luttinger-Ward functional provides a link between static thermodynamic quantities on the one hand and single-particle excitations on the other. The functional is useful to derive several general properties of the system as well as for the formulation of thermodynamically consistent approximations. Its original construction, however, is perturbative as it is based on the weak-coupling skeleton-diagram expansion. Here, it is shown that the Luttinger-Ward functional can be derived within a general functional-integral approach. This alternative and non-perturbative approach stresses the fact that the Luttinger-Ward functional is universal for a large class of models.

PACS numbers: 71.10.-w, 71.15.-m

I. INTRODUCTION

For a system of correlated electrons in equilibrium, there are several relations[1,2,3] between static quantities which describe the thermodynamics of the system and dynamic quantities which describe its one-particle excitations. Static quantities are given by the grand potential Ω and its derivatives with respect to temperature T, chemical potential μ etc. The one-electron Green's function $\boldsymbol{G} = \boldsymbol{G}(i\omega_n)$ or the self-energy $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(i\omega_n)$, on the other hand, are dynamic quantities which yield (equivalent) information on an idealized (photoemission or inverse photoemission) excitation process.

The Luttinger-Ward functional $\widehat{\Phi}[\boldsymbol{G}]$ provides a special relation between static and dynamic quantities with several important properties:[4] First, the grand potential is obtained from the Luttinger-Ward functional evaluated at the exact Green's function, $\Phi = \widehat{\Phi}[\boldsymbol{G}]$, via

$$\Omega = \Phi + \mathrm{Tr}\ln \boldsymbol{G} - \mathrm{Tr}\,\boldsymbol{\Sigma}\boldsymbol{G}\,. \tag{1}$$

Second, the functional derivative of the Luttinger-Ward functional,

$$\frac{1}{T}\frac{\delta\widehat{\Phi}[\boldsymbol{G}]}{\delta \boldsymbol{G}} = \widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]\,, \tag{2}$$

defines a functional $\widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]$ which gives the exact self-energy of the system if evaluated at the exact Green's function. The relation $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]$ is independent from the Dyson equation $\boldsymbol{G}^{-1} = \boldsymbol{G}_0^{-1} - \boldsymbol{\Sigma}$. Third, in the non-interacting limit:

$$\widehat{\Phi}[\boldsymbol{G}] \equiv 0 \quad \text{for}\, \boldsymbol{U} = 0\,. \tag{3}$$

Finally, the functional dependence $\widehat{\Phi}[\boldsymbol{G}]$ is completely determined by the interaction part of the Hamiltonian and independent from the one-particle part:

$$\widehat{\Phi}[\boldsymbol{G}] \quad \text{universal}\,. \tag{4}$$

This universality property can also be expressed as follows: Two systems with the same interaction $\boldsymbol{U}$ but different one-particle parameters $\boldsymbol{t}$ (on-site energies and hopping integrals) in the respective Hamiltonian are described by the same Luttinger-Ward functional. Using Eq. (2), this implies that the functional $\widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]$ is universal, too.

If Ref. 4 it is shown by Luttinger and Ward that $\widehat{\Phi}[\boldsymbol{G}]$ can be constructed order by order in diagrammatic weak-coupling perturbation theory. Φ is obtained as the limit of the infinite series of closed diagrams without any self-energy insertions and with all free propagators in a diagram replaced by fully interacting ones (see Fig. 1). Generally, this skeleton-diagram expansion cannot be summed up to get a closed form for $\widehat{\Phi}[\boldsymbol{G}]$. So, unfortunately, the explicit functional dependence $\widehat{\Phi}[\boldsymbol{G}]$ is actually unknown – even for the most simple Hamiltonians such as the Hubbard model.[5] The defining properties, Eqs. (1–4), however, are easily verified.[4]

The Luttinger-Ward functional is useful for several general considerations: With the help of $\widehat{\Phi}[\boldsymbol{G}]$ and the Dyson equation, the grand potential can be considered as a functional of the Green's function $\Omega = \widehat{\Omega}[\boldsymbol{G}]$ or as functional of the self-energy $\Omega = \widehat{\Omega}[\boldsymbol{\Sigma}]$, such that Ω is stationary at the physical $\boldsymbol{G}$ or $\boldsymbol{\Sigma}$.[4,6] This represents a remarkable variational principle which connects static with dynamic physical quantities. The Luttinger-Ward functional is also used in the microscopic derivation of some zero- or low-temperature properties of Fermi liquids as discussed in Refs. 4,7. The derivative of the functional, Eq. (2), shows the self-energy to be gradient field when considered as a functional of the Green's function, $\widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]$. This fact is related to certain symmetry properties of two-particle Green's functions as orig-

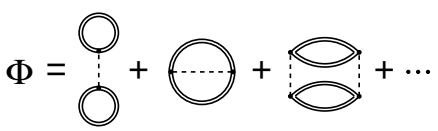

FIG. 1: Classical definition of the Luttinger-Ward functional $\widehat{\Phi}[\boldsymbol{G}]$. Double lines: fully interacting propagator $\boldsymbol{G}$. Dashed lines: interaction $\boldsymbol{U}$.

inally noted by Baym and Kadanoff.[8] Furthermore, the Luttinger-Ward functional is of great importance in the construction of thermodynamically consistent approximations. So-called conserving approximations virtually start from the Luttinger-Ward functional.[6,8] This is essential to prove these approximations to respect a number of macroscopic conservation laws. The Hartree-Fock and the random-phase approximations are well-known examples. These "classical" conserving approximations are essentially limited to the weak-coupling regime. However, the Luttinger-Ward functional can also be used to construct non-perturbative approximations. This was first realized in the context of the dynamical mean-field theory (DMFT) for lattice models of correlated electrons.[9,10,11,12] Here, one exploits the universality of the functional, Eq. (4), to achieve an (approximate) mapping of the original lattice model onto a simpler impurity model with the same interaction part. The fact that $\widehat{\Phi}[\boldsymbol{G}]$ is the same for a large class of systems, has recently been shown[13,14] to be the key feature that allows to construct several non-perturbative and thermodynamically consistent approximations.[15,16] This idea has been termed "self-energy-functional approach" (SFA).

Such general considerations remain valid as long as the Luttinger-Ward functional is well defined. This presupposes that the skeleton-diagram expansion is convergent or at least that formal manipulations of diagrammatic quantities are consistent in themselves and eventually lead to physically meaningful results. Provided that one can assure that no singular point is passed when starting from the non-interacting Fermi gas and increasing the interaction strength, this seems to be plausible. A strict proof that the skeleton-diagram expansion is well-behaved, however, will hardly be possible in most concrete situations. On the contrary, it is well known that the expansion is questionable in a number of cases, e.g. in case of a symmetry-broken state or a state that is not "adiabatically connected" to the non-interacting limit, such as a Mott insulator. The skeleton-diagram expansion may break down even in the absence of any spontaneous symmetry breaking in a (strongly correlated) state that gradually evolves from a metallic Fermi liquid. This has explicitly been shown by Hofstetter and Kehrein[17] for the narrow-band limit of the single-impurity Anderson model (see Refs. 18,19 for a discussion of possible physical consequences). Generally speaking there is no strict argument available that ensures the convergence of the skeleton-diagram expansion in the strong-coupling regime.

The purpose of the present paper is to show that a construction of the Luttinger-Ward functional is possible that does not make use of the skeleton-diagram expansion. The proposed construction is based on a standard functional-integral approach and avoids the formal complications mentioned above. Thereby, one achieves an alternative and in particular non-perturbative route to the general properties of correlated electron systems derived from the functional, to the dynamical mean-field theory as well as to the self-energy-functional approach. It should be stressed that the intended construction of the Luttinger-Ward functional requires more than a simple definition of the quantity Φ (which could trivially be achieved by using Eq. (1): $\Phi \equiv \Omega - \mathrm{Tr}\ln\boldsymbol{G} + \mathrm{Tr}\,\boldsymbol{\Sigma}\boldsymbol{G}$). The task is rather to provide a functional $\widehat{\Phi}[\boldsymbol{G}]$ with the properties Eqs. (1–4).

Previous approaches are either perturbative or cannot prove Eqs. (1–4): A construction of the Luttinger-Ward functional different from the original one[4] has been given by Baym:[6] The existence of $\widehat{\Phi}[\boldsymbol{G}]$ is deduced from a "vanishing curl condition", $\delta\Sigma(1,1')/\delta G(2',2) = \delta\Sigma(2,2')/\delta G(1',1)$, which is derived from an analysis of the functional dependence of $\boldsymbol{G}$ on an arbitrary (time-dependent) external perturbation $\boldsymbol{J}$. However, an independent functional relation $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]$ is required in addition. In Ref. 6 the latter is assumed to be given by the (full or by a truncated) skeleton-diagram expansion, and consequently this approach is perturbative again.

As also shown in Ref. 6, the Green's function in the presence of an external field $\boldsymbol{J}$ can be derived from the grand potential $\widehat{\Omega}'[\boldsymbol{J}]$ as $\widehat{\boldsymbol{G}}(\boldsymbol{J}) = (1/T)\delta\widehat{\Omega}'[\boldsymbol{J}]/\delta\boldsymbol{J}$. Using the inverse functional, $\widehat{\boldsymbol{J}}[\boldsymbol{G}]$, Legendre transformation yields $\widehat{\Omega}[\boldsymbol{G}] = \widehat{\Omega}'[\widehat{\boldsymbol{J}}[\boldsymbol{G}]] - \mathrm{Tr}\,\boldsymbol{G}\,\widehat{\boldsymbol{J}}[\boldsymbol{G}]$. This (non-perturbative) functional and the Dyson equation can be used to define $\widehat{\Phi}[\boldsymbol{G}] \equiv \widehat{\Omega}[\boldsymbol{G}] - \mathrm{Tr}\ln\boldsymbol{G} + \mathrm{Tr}\,(\boldsymbol{G}_0^{-1} - \boldsymbol{G}^{-1})\boldsymbol{G}$. This idea is in the spirit of the effective action approach.[20,21,22] Here, however, the problem is that the universality of $\widehat{\Phi}[\boldsymbol{G}]$, Eq. (4), cannot be proven. The Luttinger-Ward functional constructed in this way explicitly depends on $\boldsymbol{G}_0$ and thus on the one-particle parameters $\boldsymbol{t}$.

The paper is organized as follows: The next section briefly introduces the notations and the quantities of interest. The construction of the Luttinger-Ward functional is described in Sec. III. Sec. IV gives a brief discussion of the properties of the functional and its use within the dynamical mean-field theory and the self-energy-functional approach. The results are summed up in Sec. V.

II. STATIC AND DYNAMIC QUANTITIES

Consider a system of electrons at temperature T and chemical potential μ in thermal equilibrium and let $H = H(\boldsymbol{t},\boldsymbol{U}) = H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$ be its Hamiltonian where

$$\begin{aligned} H_0(\boldsymbol{t}) &= \sum_{\alpha\beta} t_{\alpha\beta}\, c_\alpha^\dagger c_\beta \,, \\ H_1(\boldsymbol{U}) &= \frac{1}{2}\sum_{\alpha\beta\gamma\delta} U_{\alpha\beta\delta\gamma}\, c_\alpha^\dagger c_\beta^\dagger c_\gamma c_\delta \,. \end{aligned} \tag{5}$$

An index α refers to an arbitrary set of quantum numbers characterizing a one-particle basis state. If N is the total particle-number operator, the grand potential of the

system is given by $\Omega_{\boldsymbol{t},\boldsymbol{U}} = -T \ln Z_{\boldsymbol{t},\boldsymbol{U}}$ where

$$Z_{\boldsymbol{t},\boldsymbol{U}} = \mathrm{tr} \exp(-(H(\boldsymbol{t},\boldsymbol{U}) - \mu N)/T) \tag{6}$$

is the partition function. Here and in the following the dependence of all quantities on the one-particle parameters $\boldsymbol{t}$ and the interaction parameters $\boldsymbol{U}$ is made explicit through the subscripts.

Using a matrix notation, the free one-particle Green's function is denoted by $\boldsymbol{G}_{\boldsymbol{t},0}$. Its elements (for fixed μ) are given by:

$$G_{\boldsymbol{t},0,\alpha\beta}(i\omega_n) = \left(\frac{1}{i\omega_n + \mu - \boldsymbol{t}} \right)_{\alpha\beta} . \tag{7}$$

Here $i\omega_n = i(2n+1)\pi T$ is the n-th Matsubara frequency. The fully interacting Green's function is denoted by $\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$. Using Grassmann variables $\xi_\alpha(i\omega_n) = T^{1/2} \int_0^{1/T} d\tau \; e^{i\omega_n \tau} \xi_\alpha(\tau)$ and $\xi^*_\alpha(i\omega_n) = T^{1/2} \int_0^{1/T} d\tau \, e^{-i\omega_n \tau} \xi^*_\alpha(\tau)$, its elements can be written as[3]

$$\begin{aligned} G_{\boldsymbol{t},\boldsymbol{U},\alpha\beta}(i\omega_n) &= -\langle \xi_\alpha(i\omega_n) \xi^*_\beta(i\omega_n) \rangle_{\boldsymbol{t},\boldsymbol{U}} \\ &= \frac{-1}{Z_{\boldsymbol{t},\boldsymbol{U}}} \int D\xi D\xi^* \xi_\alpha(i\omega_n) \xi^*_\beta(i\omega_n) \exp\left(A_{\boldsymbol{t},\boldsymbol{U},\xi\xi^*}\right) \end{aligned} \tag{8}$$

where

$$\begin{aligned} A_{\boldsymbol{t},\boldsymbol{U},\xi\xi^*} &= \sum_{n,\alpha\beta} \xi^*_\alpha(i\omega_n)((i\omega_n + \mu)\delta_{\alpha\beta} - t_{\alpha\beta}) \xi_\beta(i\omega_n) \\ &- \frac{1}{2} \sum_{\alpha\beta\gamma\delta} U_{\alpha\beta\delta\gamma} \int_0^{1/T} d\tau \; \xi^*_\alpha(\tau) \xi^*_\beta(\tau) \xi_\gamma(\tau) \xi_\delta(\tau) \end{aligned} \tag{9}$$

is the action. Finally, the self-energy is defined as

$$\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}} = \boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{G}^{-1}_{\boldsymbol{t},\boldsymbol{U}} . \tag{10}$$

The goal is to construct a functional $\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ (where $\boldsymbol{G}$ is considered as a free variable) which vanishes in the non-interacting case, $\widehat{\Phi}_0[\boldsymbol{G}] = 0$ [Eq. (3)], which is universal, i.e. independent of $\boldsymbol{t}$ [Eq. (4)], which yields $\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}} - \mathrm{Tr} \ln \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} + \mathrm{Tr}(\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}} \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}})$ if evaluated at the exact Green's function $\boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$ [Eq. (1)], and the derivative of which is a functional $\widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}]$ with $\widehat{\boldsymbol{\Sigma}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}] = \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}$ [Eq. (2)]. (Here the notation $\mathrm{Tr}\boldsymbol{A} \equiv T \sum_n \sum_\alpha e^{i\omega_n 0^+} A_{\alpha\alpha}(i\omega_n)$ is used. 0^+ is a positive infinitesimal. Functionals $\widehat{A} = \widehat{A}[\cdots]$ are indicated by a hat and should be distinguished clearly from physical quantities A.)

For the classical construction of $\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ via the skeleton-diagram expansion (Fig. 1), these properties are easily verified: The universality of the functional [Eq. (4)] is obvious as any diagram depends on $\boldsymbol{U}$ and on $\boldsymbol{G}$ *only*; there is no explicit dependence on the free Green's function $\boldsymbol{G}_{\boldsymbol{t},0}$, i.e. no explicit dependence on $\boldsymbol{t}$. Since there is no zeroth-order diagram, $\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ trivially vanishes for $\boldsymbol{U} = 0$ [Eq. (3)]. The functional derivative of $\widehat{\Phi}[\boldsymbol{G}]$ with respect to $\boldsymbol{G}$ corresponds to the removal of a propagator from each of the Φ diagrams. Taking care of topological factors,[1,4] one ends up with the skeleton-diagram expansion for the self-energy, i.e. one gets Eq. (2). Using Eq. (2), the Dyson equation (10), and $\Phi_{\boldsymbol{t},\boldsymbol{U}} \equiv \widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}]$, the μ derivative of the l.h.s and of the r.h.s of Eq. (1) are equal for any fixed interaction strength $\boldsymbol{U}$ and temperature T. Namely, $(\partial/\partial\mu)(\Phi_{\boldsymbol{t},\boldsymbol{U}} + \mathrm{Tr} \ln \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} - \mathrm{Tr}\, \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}} \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}) = \mathrm{Tr}\, \boldsymbol{G}^{-1}_{\boldsymbol{t},\boldsymbol{U}} (\partial \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}/\partial\mu) - \mathrm{Tr}\, \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} (\partial \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}/\partial\mu) = -\mathrm{Tr} \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} = -\langle N \rangle_{\boldsymbol{t},\boldsymbol{U}} = \partial \Omega_{\boldsymbol{t},\boldsymbol{U}}/\partial\mu$. Integration over μ then yields Eq. (1). (Note that Eq. (1) holds trivially for $\mu \to -\infty$, i.e. for $\langle N \rangle_{\boldsymbol{t},\boldsymbol{U}} \to 0$ since $\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}} = 0$ and $\Phi_{\boldsymbol{t},\boldsymbol{U}} = 0$ in this limit). An equivalent derivation of Eq. (1) can be given by a coupling-constant integration.[4]

III. LUTTINGER-WARD FUNCTIONAL

The starting point is the standard functional-integral representation of the partition function as given in Ref. 3, for example: Define the functional

$$\widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}] = -T \ln \widehat{Z}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}] \tag{11}$$

with

$$\widehat{Z}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}] = \int D\xi D\xi^* \exp\left(\widehat{A}_{\boldsymbol{U},\xi\xi^*}[\boldsymbol{G}_0^{-1}] \right) \tag{12}$$

and

$$\begin{aligned} \widehat{A}_{\boldsymbol{U},\xi\xi^*}[\boldsymbol{G}_0^{-1}] &= \sum_{n,\alpha\beta} \xi^*_\alpha(i\omega_n) G^{-1}_{0,\alpha\beta}(i\omega_n) \xi_\beta(i\omega_n) \\ &- \frac{1}{2} \sum_{\alpha\beta\gamma\delta} U_{\alpha\beta\delta\gamma} \int_0^{1/T} d\tau \; \xi^*_\alpha(\tau) \xi^*_\beta(\tau) \xi_\gamma(\tau) \xi_\delta(\tau) . \end{aligned} \tag{13}$$

$\widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]$ parametrically depends on $\boldsymbol{U}$. $\boldsymbol{G}_0^{-1}$ is considered as a free variable. At the (matrix inverse of the) exact free Green's function, $\boldsymbol{G}_0^{-1} = \boldsymbol{G}^{-1}_{\boldsymbol{t},0}$, the functional yields the exact grand potential,

$$\widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}^{-1}_{\boldsymbol{t},0}] = \Omega_{\boldsymbol{t},\boldsymbol{U}} , \tag{14}$$

of the system with Hamiltonian $H = H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$. Its derivative defines a functional $\widehat{\mathcal{G}}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]$,

$$\frac{1}{T} \frac{\delta \widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]}{\delta \boldsymbol{G}_0^{-1}} = -\frac{1}{\widehat{Z}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]} \frac{\delta \widehat{Z}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]}{\delta \boldsymbol{G}_0^{-1}} \equiv -\widehat{\mathcal{G}}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}] , \tag{15}$$

with the property

$$\widehat{\mathcal{G}}_{\boldsymbol{U}}[\boldsymbol{G}^{-1}_{\boldsymbol{t},0}] = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} \tag{16}$$

which is easily verified using Eq. (8).

The strategy to be pursued is the following: $\widehat{\mathcal{G}}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]$ is a universal ($\boldsymbol{t}$ independent) functional and can be used

to construct a universal relation $\boldsymbol{G} = \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ between the one-particle Green's function and the self-energy independent from the Dyson equation. Using the universal functionals $\widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]$ and $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$, a universal functional $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is defined the derivative of which essentially yields $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$. The Luttinger-Ward functional can then be obtained by Legendre transformation and is universal by construction.

To start with, consider the equation

$$\widehat{\mathcal{G}}_{\boldsymbol{U}}[\boldsymbol{G}^{-1} + \boldsymbol{\Sigma}] = \boldsymbol{G} \,. \tag{17}$$

This is a relation between the variables $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$ which, for a given $\boldsymbol{\Sigma}$, may be solved for $\boldsymbol{G}$. This defines a functional $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$, i.e.

$$\widehat{\mathcal{G}}_{\boldsymbol{U}}[\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]^{-1} + \boldsymbol{\Sigma}] = \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \,. \tag{18}$$

For a given self-energy $\boldsymbol{\Sigma}$, the Green's function $\boldsymbol{G} = \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is defined to be the solution of Eq. (17). From the Dyson equation (10) and Eq. (16) it is obvious that the relation (17) is satisfied for $\boldsymbol{G}$ and $\boldsymbol{\Sigma}$ being the exact Green's function and the exact self-energy, $\boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}$, of a system with the interaction $\boldsymbol{U}$ and some set of one-particle parameters $\boldsymbol{t}$ ($H = H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$). Hence,

$$\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}] = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} \,. \tag{19}$$

A brief discussion of the existence and the uniqueness of possible solutions of the relation (17) is given in Appendix A.

With the help of the functionals $\widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}]$ and $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$, a functional $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ can be defined as:

$$\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] = \widehat{\Omega}_{\boldsymbol{U}}[\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]^{-1} + \boldsymbol{\Sigma}] - \operatorname{Tr}\ln \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \,. \tag{20}$$

Using Eq. (15) one finds:

$$\begin{aligned} \frac{1}{T}\frac{\delta \widehat{F}_U[\boldsymbol{\Sigma}]}{\delta \boldsymbol{\Sigma}} \;&=\; -\widehat{\mathcal{G}}_{\boldsymbol{U}}[\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]^{-1} + \boldsymbol{\Sigma}] \cdot \left(\frac{\delta \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]^{-1}}{\delta \boldsymbol{\Sigma}} + \mathbf{1} \right) \\ &-\; \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]^{-1} \cdot \frac{\delta \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]}{\delta \boldsymbol{\Sigma}} \,, \end{aligned} \tag{21}$$

and, using Eq. (18),

$$\frac{1}{T}\frac{\delta \widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]}{\delta \boldsymbol{\Sigma}} = -\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \,. \tag{22}$$

So $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ can be considered as the gradient of the (scalar) self-energy functional $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$. Therewith, the Legendre transform of $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ can be constructed:

$$\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}] = \widehat{F}_{\boldsymbol{U}}[\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}]] + \operatorname{Tr}(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}]\; \boldsymbol{G}) \,. \tag{23}$$

Here $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}]$ is the inverse of the functional $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$. The functional can be assumed to be invertible (locally) provided that the system is not at a critical point for a phase transition (see also Ref. 13). Eq. (23) defines the Luttinger-Ward functional.

IV. DISCUSSION

A. Properties of the Luttinger-Ward functional

The properties of the Luttinger-Ward functional, Eqs. (1–4), can be verified easily: Eqs. (10), (14), (19) and (20) imply

$$\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}} - \operatorname{Tr}\ln \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} \,, \tag{24}$$

and with $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}] = \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}$ the evaluation of the Luttinger-Ward functional at $\boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$ yields

$$\Phi_{\boldsymbol{t},\boldsymbol{U}} \equiv \widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}} - \operatorname{Tr}\ln \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}} + \operatorname{Tr}(\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}) \,, \tag{25}$$

i.e. Eq. (1). From Eqs. (22) and (23), one immediately has:

$$\frac{1}{T}\frac{\delta \widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]}{\delta \boldsymbol{G}} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}] \,, \tag{26}$$

i.e. Eq. (2). In the limit $\boldsymbol{U} = 0$, the functionals $\widehat{\boldsymbol{G}}_{\boldsymbol{U}=0}[\boldsymbol{\Sigma}]$ and $\widehat{F}_{\boldsymbol{U}=0}[\boldsymbol{\Sigma}]$ are ill-defined (the domain of the functionals shrinks to a single point, $\boldsymbol{\Sigma} = 0$, see Appendix A). However, from Eq. (25), one directly has $\Phi_{\boldsymbol{U}=0}[\boldsymbol{G}_{\boldsymbol{t},0}] = 0$ for any $\boldsymbol{t}$ [see Eq. (3)] since $\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{0}} = 0$ and $\Omega_{\boldsymbol{t},\boldsymbol{0}} = \operatorname{Tr}\ln \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{0}}$ (a proof for the latter can be found in Ref. 4). Finally, the universality of $\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$, Eq. (4) is obvious as the definition (23) of the Luttinger-Ward functional involves the universal ($\boldsymbol{t}$ independent) functionals $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}]$ only.

B. Variational principle

Using the Legendre transform of the Luttinger-Ward functional $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$, one may define

$$\widehat{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}] = \operatorname{Tr}\ln \frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}} + \widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \,. \tag{27}$$

The functional derivative is easily calculated:

$$\frac{1}{T}\frac{\delta \widehat{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}]}{\delta \boldsymbol{\Sigma}} = \frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}} - \widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \,. \tag{28}$$

The equation

$$\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] = \frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}} \tag{29}$$

is a (highly non-linear) conditional equation for the self-energy of the system $H = H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$: Eqs. (10) and (19) show that it is satisfied by the exact self-energy $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}$. Note that the l.h.s of (29) is independent of $\boldsymbol{t}$ but depends on $\boldsymbol{U}$ (universality of $\widehat{\boldsymbol{G}}[\boldsymbol{\Sigma}]$), while the r.h.s is independent of $\boldsymbol{U}$ but depends on $\boldsymbol{t}$ via $\boldsymbol{G}_{\boldsymbol{t},0}^{-1}$. The obvious problem of finding a solution of Eq. (29) is that

there is no closed form for the functional $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$. Solving Eq. (29) is equivalent to a search for the stationary point of the grand potential as a functional of the self-energy:

$$\frac{\delta\widehat{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}]}{\delta\boldsymbol{\Sigma}} = 0 \; . \tag{30}$$

Similarly, one can also construct a variational principle using the Green's function as the basic variable, $\delta\widehat{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}]/\delta\boldsymbol{G} = 0$.

C. Dynamical mean-field theory

The dynamical mean-field theory[9,10,11,12] basically applies to lattice models of correlated electrons with on-site interactions such as the Hubbard model,[5] for example. The DMFT aims at an approximate solution of Eq. (29) and is based on two ingredients:

(i) It is important to note that the Luttinger-Ward functional $\widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ is the same for the lattice (e.g. Hubbard) model and for an impurity model (single-impurity Anderson model). Actually a (decoupled) set of impurity models has to be considered – one impurity model with the according local interaction at each site of the original lattice. This ensures that the interaction ($\boldsymbol{U}$) term is the same as in the lattice model. (In case of translational symmetry the a priori different impurity models can be assumed to be equivalent). As $\boldsymbol{U}$ is the same in the lattice and in the impurity model, the Luttinger-Ward functional, as well as $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$, is the same.

(ii) Let the lattice model be characterized by one-particle parameters $\boldsymbol{t}$ and the impurity model by parameters $\boldsymbol{t}'$. The fundamental equation (29) for the lattice model would then be solved by the exact self-energy $\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}$. As an ansatz for an approximate solution $\boldsymbol{\Sigma}$ of Eq. (29), the self-energy is assumed to be local within the DMFT and to be representable as the exact self-energy of the impurity model for some parameters $\boldsymbol{t}'$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}} \; . \tag{31}$$

The universality of the Luttinger-Ward functional (i) and the local approximation for the self-energy (ii) are sufficient to derive the DMFT: Inserting the ansatz (31) into Eq. (29) yields a conditional equation for the one-particle parameters of the impurity model $\boldsymbol{t}'$. The l.h.s becomes $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}] = \boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}$, i.e. the exact Green's function of the impurity model, while the r.h.s reads $(\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}})^{-1}$. The resulting equation for the parameters $\boldsymbol{t}'$ can be fulfilled only locally, i.e. by equating the local elements of the respective Green's functions at the impurity and the original site respectively:

$$(\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}})_{\mathrm{loc}} = \left(\frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}}\right)_{\mathrm{loc}} \; . \tag{32}$$

This is the so-called self-consistency equation of the DMFT.[12]

This consideration can be seen as an independent and, in particular, non-perturbative re-derivation of the DMFT which supplements known approaches such as the cavity method.[12]

D. Self-energy-functional approach

The universality of the Luttinger-Ward functional or of its Legendre transform $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is central to the recently developed self-energy-functional approach.[13,14] The SFA is a general variational scheme which includes the DMFT as a special limit. The idea is to take as an ansatz for the self-energy of a model $H = H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$ the exact self-energy $\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}$ of a so-called reference system $H' = H_0(\boldsymbol{t}') + H_1(\boldsymbol{U})$ that shares with the original model the same interaction part. The parameters $\boldsymbol{t}'$ of the one-particle part are considered as variational parameters to search for the stationary point of the grand potential as a functional of the self-energy. This means to insert the ansatz $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}$ into the general expression (27) and to solve the Euler equation $\partial\widehat{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]/\partial\boldsymbol{t}' = 0$, i.e.:

$$\frac{\partial}{\partial\boldsymbol{t}'}\left(\mathrm{Tr}\ln\frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}} + \widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]\right) = 0 \tag{33}$$

for $\boldsymbol{t}'$. If the search for the optimum set of one-particle parameters $\boldsymbol{t}'$ was unrestricted, the approach would be exact in principle as the Euler equation would then be equivalent with the Euler equation (29) of the general variational principle Eq. (30).

A restriction of the space of variational parameters becomes necessary to evaluate the quantity $\widehat{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]$ which, in general, is impossible as a closed form for the functional $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is not known. With a proper restriction, however, the reference system H' can be made accessible to an exact (numerical) solution which allows to derive the exact grand potential and the exact Green's function of the system H'. Therewith, making use of the universality of $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ and using Eqs. (23) and (25) for the reference system,

$$\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}] = \Omega_{\boldsymbol{t}',\boldsymbol{U}} - \mathrm{Tr}\ln\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}} \; . \tag{34}$$

Note that this implies that an exact evaluation of $\widehat{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is possible for self-energies of a exactly solvable reference system with the same interaction part as the original one. Using this result in Eq. (33), one obtains:

$$\frac{\partial}{\partial\boldsymbol{t}'}\left(\Omega_{\boldsymbol{t}',\boldsymbol{U}} + \mathrm{Tr}\ln\frac{1}{\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}} - \mathrm{Tr}\ln\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}\right) = 0 \; , \tag{35}$$

which can be evaluated to fix $\boldsymbol{t}'$ and therewith the optimal self-energy and grand potential (see Refs. 13,14,15,16 for details and concrete examples).

E. Luttinger's theorem

Finally, the role of the Luttinger-Ward functional in the derivation of general properties of correlated electron systems shall be discussed. As an important example, the Luttinger theorem[4] is considered. For a translationally invariant system, the theorem states that in the limit $T \to 0$ the average particle number is equal to the volume enclosed by the Fermi surface in $\boldsymbol{k}$ space:

$$\langle N \rangle = V_{\rm FS} \,. \tag{36}$$

The Fermi surface is defined by the set of $\boldsymbol{k}$ points in the first Brillouin zone that satisfies $\mu - \eta_{\boldsymbol{k}} = 0$ where $\eta_{\boldsymbol{k}}$ are the eigenvalues of the matrix $\boldsymbol{t} + \boldsymbol{\Sigma}(\omega)$ at vanishing excitation energy $\omega = 0$. Hence, to formulate the Luttinger theorem, one obviously has to presuppose that there is a Fermi surface at all, i.e. that $\boldsymbol{\Sigma}(\omega = 0)$ is Hermitian.[25] The original proof of the theorem[4] is perturbative as it makes use of the skeleton-diagram expansion. A non-perturbative proof, based on topological considerations, was proposed recently[23] and is based on the assumption that the system is a Fermi liquid.

To discuss the Luttinger theorem in the present context, consider the following shift transformation of the Green's function

$$\boldsymbol{S}^{(z)}\boldsymbol{G}(i\omega_n) = \boldsymbol{G}^{(z)}(i\omega_n) = \boldsymbol{G}(i\omega_n + iz) \tag{37}$$

with $z = 2\pi k T$ and k integer (z is a bosonic Matsubara frequency). $\boldsymbol{S}^{(z)}$ is a linear and unitary transformation. The shift transformation leaves the functional integral Eq. (11) unchanged:

$$\widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{S}^{(z)}\boldsymbol{G}_0^{-1}] = \widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{G}_0^{-1}] \,. \tag{38}$$

To verify this invariance, one has to note that the shift of the Matsubara frequencies in $\boldsymbol{G}_0^{-1}$ by z can be transformed into a shift $\omega_n \to \omega_n - z$ in the Grassmann numbers:

$$\xi_\alpha(i\omega_n) \to \xi_\alpha(i\omega_n - iz) \,. \tag{39}$$

In imaginary-time representation this shift is equivalent with the multiplication of a phase:

$$\xi_\alpha(\tau) \to e^{-iz\tau}\xi_\alpha(\tau) \,, \qquad \xi^*_\alpha(\tau) \to e^{iz\tau}\xi^*_\alpha(\tau) \,. \tag{40}$$

This, however, leaves the functional integral unchanged as the transformation Eq. (39) or Eq. (40) is linear and the Jacobian is unity. Note that antiperiodic boundary conditions $\xi_\alpha(\tau = 1/T) = -\xi_\alpha(\tau = 0)$ are respected for a bosonic shift frequency z.

Denoting $\Omega_{\boldsymbol{t},\boldsymbol{U}}(z) \equiv \widehat{\Omega}_{\boldsymbol{U}}[\boldsymbol{S}^{(z)}\boldsymbol{G}_{\boldsymbol{t},0}^{-1}]$, Eq. (38) states that $\Omega_{\boldsymbol{t},\boldsymbol{U}}(z) = \Omega_{\boldsymbol{t},\boldsymbol{U}}(0)$. Following the steps in the construction of the Luttinger-Ward functional in Sec. III, one easily verifies that this implies $\Phi_{\boldsymbol{t},\boldsymbol{U}}(z) = \Phi_{\boldsymbol{t},\boldsymbol{U}}(0)$ where $\Phi_{\boldsymbol{t},\boldsymbol{U}}(z) \equiv \widehat{\Phi}_{\boldsymbol{U}}[\boldsymbol{S}^{(z)}\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}]$. For the Legendre transform, one has $F_{\boldsymbol{t},\boldsymbol{U}}(z) = F_{\boldsymbol{t},\boldsymbol{U}}(0)$ where $F_{\boldsymbol{t},\boldsymbol{U}}(z) \equiv \widehat{F}_{\boldsymbol{U}}[\boldsymbol{S}^{(z)}\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}]$. Now, in the limit $T \to 0$, z becomes a continuous variable. Hence,

$$\frac{d}{dz} \lim_{T\to 0} F_{\boldsymbol{t},\boldsymbol{U}}(z) = 0 \,. \tag{41}$$

If the limit and the derivative can be interchanged,

$$\frac{d}{dz} \lim_{T\to 0} F_{\boldsymbol{t},\boldsymbol{U}}(z) = \lim_{T\to 0} \frac{d}{dz} F_{\boldsymbol{t},\boldsymbol{U}}(z) \,, \tag{42}$$

Eqs. (24) and (41) imply

$$-\lim_{T\to 0} \frac{d\Omega_{\boldsymbol{t},\boldsymbol{U}}(z)}{dz} = -\lim_{T\to 0} \frac{d\mathrm{Tr}\ln \boldsymbol{S}^{(z)}\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}}{dz} \,. \tag{43}$$

The z dependence of the grand potential is the same as its μ dependence, and thus $-(d/dz)\Omega_{\boldsymbol{t},\boldsymbol{U}}(z = 0) = -\partial\Omega_{\boldsymbol{t},\boldsymbol{U}}/\partial\mu = \langle N \rangle$. The evaluation of the r.h.s in Eq. (43) is straightforward and can be found in Ref. 4, for example. It turns out that at $z = 0$ the r.h.s is just the Fermi-surface volume $V_{\rm FS}$.

Consequently, the non-perturbative construction of the Luttinger-Ward functional allows to reduce the proof of the Luttinger theorem to the proof of Eq. (42). This, however, requires certain assumptions on the regularity of the $T \to 0$ limit which are non-trivial generally.

V. SUMMARY

To summarize, the present paper has shown that the Luttinger-Ward functional can be constructed within the framework of functional integrals under fairly general assumptions. In particular, there no need for an adiabatic connection to the non-interacting limit and no expansion in the interaction strength as was required in the original approach of Luttinger and Ward.[4] The construction merely assumes the very existence of the functional integral over Grassmann fields, i.e. the existence of the Trotter limit, for the representation of the partition function.

It is well known that the Luttinger-Ward functional can be employed for different purposes, some of which have been discussed here: The functional is used to derive some general properties of correlated electron systems, such as the Luttinger theorem. It allows to formulate a variational principle involving a thermodynamical potential as a functional of the Green's function or the self-energy and thereby provides a unique and thermodynamically meaningful link between static and dynamic quantities which is helpful for interpretations and for the construction of approximations. An independent derivation of the dynamical mean-field is possible using the special properties of the Luttinger-Ward functional and the universality of the functional in particular. The latter is of central importance in the context of the self-energy-functional approach which is a general framework to construct thermodynamically consistent approximations.

Referring to the standard definition of the Luttinger-Ward functional that is based on the weak-coupling skeleton-diagram expansion, the above-mentioned and any further considerations based on the functional and its unique properties meet with criticism when applied to strongly correlated, non-Fermi liquid or symmetry-broken states. This is exactly the point where the presented non-perturbative construction of the Luttinger-Ward functional is useful.

Acknowledgments

The author acknowledges helpful discussions with F.F. Assaad, M. Bechmann, W. Hanke and G. Kotliar. The work is supported by the Deutsche Forschungsgemeinschaft (Forschergruppe 538).

APPENDIX A

As the relation (17) is highly non-linear, the existence and the uniqueness of possible solutions have to be discussed:

Take $\boldsymbol{U}$ to be fixed and assume that the self-energy given is the exact self-energy of a system $H = H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$ with some hopping parameters $\boldsymbol{t}$. So the self-energy $\boldsymbol{\Sigma}$ is assumed to be given from the space $\mathcal{S}_{\boldsymbol{U}}$ of $\boldsymbol{t}$ representable self-energies $\mathcal{S}_{\boldsymbol{U}} \equiv \{\boldsymbol{\Sigma} \,|\, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}\,,\, \boldsymbol{t}\,\text{arbitrary}\}$ ($\boldsymbol{U}$ fixed). With the help of Eq. (16) it is then obvious that the exact Green's function of this system, $\boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$, solves Eq. (17) as the Dyson equation (10) shows that $\boldsymbol{G}^{-1}_{\boldsymbol{t},\boldsymbol{U}} + \boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}$ is the exact free Green's function of this system. Concluding, one has $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}] = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$, and thus the existence of a solution is guaranteed on the space $\mathcal{S}_{\boldsymbol{U}}$. Note that it is very convenient to consider $\mathcal{S}_{\boldsymbol{U}}$ as the domain of the functional $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ since this ensures the correct analytical and causal properties of the variable $\boldsymbol{\Sigma}$.

Under the functional $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ the space $\mathcal{S}_{\boldsymbol{U}}$ is mapped onto the space $\mathcal{G}_{\boldsymbol{U}}$ of $\boldsymbol{t}$ representable Green's functions $\mathcal{G}_{\boldsymbol{U}} \equiv \{\boldsymbol{G} \,|\, \boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}\,,\, \boldsymbol{t}\,\text{arbitrary}\}$ ($\boldsymbol{U}$ fixed). Generally, the map $\widehat{\boldsymbol{G}}_{\boldsymbol{U}} : \mathcal{S}_{\boldsymbol{U}} \to \mathcal{G}_{\boldsymbol{U}}$ is not unique.[13] Hence, the uniqueness of the functional $\widehat{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ must be enforced by a proper restriction of the range $\mathcal{G}_{\boldsymbol{U}}$, i.e. of the solution set of Eq. (17). The considerations in Secs. III and IV, however, are unaffected and hold for any choice of the range, see also the related discussion in Ref. 13.

[1] A. A. Abrikosow, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey, 1964).
[2] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
[3] J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, Redwood City, 1988).
[4] J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).
[5] J. Hubbard, Proc. R. Soc. London A **276**, 238 (1963).
[6] G. Baym, Phys. Rev. **127**, 1391 (1962).
[7] J. M. Luttinger, Phys. Rev. **119**, 1153 (1960).
[8] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
[9] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
[10] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
[11] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992).
[12] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[13] M. Potthoff, Euro. Phys. J. B **32**, 429 (2003).
[14] M. Potthoff, Euro. Phys. J. B **36**, 335 (2003).
[15] M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003).
[16] C. Dahnken, M. Aichhorn, W. Hanke, E. Arrigoni, and M. Potthoff, preprint cond-mat **0309407** (2003).
[17] W. Hofstetter and S. Kehrein, Phys. Rev. B **59**, R12732 (1999).
[18] S. Kehrein, Phys. Rev. Lett. **81**, 3912 (1998).
[19] G. Kotliar, Euro. Phys. J. B **11**, 27 (1999).
[20] S. Y. Savrasov and G. Kotliar, preprint cond-mat **0106308** (2001).
[21] S. Y. Savrasov and G. Kotliar, preprint cond-mat **0308053** (2003).
[22] A. Georges, preprint cond-mat **0403123** (2004).
[23] M. Oshikawa, Phys. Rev. Lett. **84**, 3370 (2000).
[24] J. K. Freericks and V. Zlatić, Rev. Mod. Phys. **75**, 1333 (2003).
[25] For systems without Fermi surface, there is no Luttinger theorem. A nice example is given by the Falicov-Kimball model in infinite dimensions, see Ref. 24.

arXiv:cond-mat/0503715 v1 30 Mar 2005

Dynamical Variational Principles for Strongly Correlated Electron Systems

Michael Potthoff

Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany

Abstract. The self-energy-functional approach (SFA) is discussed in the context of different variational principles for strongly correlated electron systems. Formal analogies between static and dynamical variational approaches, different types of approximation strategies and the relations to density-functional and dynamical mean-field theory are emphasized. The discussion elucidates the strengths of the SFA in the construction of new non-perturbative approximations but also the limitations of the approach and thereby opens up future perspectives.

Variational approaches have a long and successful tradition in the theory of condensed-matter systems as they offer a smart, controlled and systematic way to treat the problem of electron-electron interaction. A well-known variational approach is Hartree-Fock (HF) theory. It is based on the Rayleigh-Ritz principle and provides a practicable and consistent mean-field description of an interacting electron system. As quantum fluctuations are neglected completely, HF theory must be classified as a *static* mean-field approximation. This may be contrasted with dynamical mean-field theory (DMFT) [1,2] which includes temporal fluctuations in the mean-field picture. The DMFT, however, cannot be derived from the Ritz principle. It must be constructed from some *dynamical* variational principle which involves a dynamical (i.e. time- or frequency-dependent) quantity as the basic object. Dynamical variational principles have already been suggested in the sixties [3,4] but, compared to the Ritz principle, were employed with rather limited success only. This brings up the following questions: What are the similarities and the differences between different variational principles and approximation strategies? How can the DMFT be considered as an approximation within a variational concept? Can dynamical variational principles be used for constructing practicable and non-perturbative approximations different from the DMFT? An attempt to answer these questions straightforwardly leads to the self-energy-functional approach (SFA) [5] suggested recently. The purpose of this paper is to discuss different variational approaches and to place the SFA into this context with the objective to explore possible future developments.

1 Variational Principles and Approximation Strategies

Consider a many-electron system in the volume V, at temperature T and with chemical potential μ. It is characterized by a Hamiltonian $H_{\boldsymbol{t},\boldsymbol{U}} =$

$H_0(\boldsymbol{t}) + H_1(\boldsymbol{U})$ consisting of a one-particle and an interaction term H_0 and H_1, respectively, which depend on the "model parameters" $\boldsymbol{t}$ and $\boldsymbol{U}$ (a matrix notation is used). In second-quantized form,

$$H_{\boldsymbol{t},\boldsymbol{U}} = \sum_{\alpha\beta} t_{\alpha\beta}\, c^\dagger_\alpha c_\beta + \frac{1}{2} \sum_{\alpha\beta\gamma\delta} U_{\alpha\beta\delta\gamma}\, c^\dagger_\alpha c^\dagger_\beta c_\gamma c_\delta \,, \tag{1}$$

where an index (e.g. α) refers to the states of a one-particle basis.

The characteristic of a variational approach is a certain physical quantity $\boldsymbol{X}$ to be varied, as e.g. the statistical operator, the electron density, the (local) Green's function, the self-energy etc. Clearly, at equilibrium this quantity will depend on the model parameters: $\boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}$ (and on V, T and μ).

In a variational approach, the quantity is considered to be a variable. The first task is to express a thermodynamical potential Ω (the grand potential, for example) as a function(al) of this variable: $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X}]$. As this functional is characteristic for the model system (1), it will depend on the model parameters. At the equilibrium or "physical" value, i.e. at $\boldsymbol{X} = \boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}$, one must have $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$ where $\Omega_{\boldsymbol{t},\boldsymbol{U}} = -T \ln \mathrm{tr} \exp(-(H_0(\boldsymbol{t}) + H_1(\boldsymbol{U}) - \mu N)/T)$.

Furthermore, the functional $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X}]$ should be constructed such that it becomes stationary at the physical value: $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X} = \boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}] = 0$. This *variational principle* determines $\boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}$ once the functional is known. Note that the domain of the functional must be specified in addition since in most cases $\boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}$ must satisfy some constraint or sum rule or normalization condition.

Even if the functional is known, however, it is usually impossible to evaluate $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X}]$ for a given $\boldsymbol{X}$, and one has to resort to approximations. One may distinguish between three types of approximation strategies:

In a *type-I approximation* one derives the Euler equation $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X}]/\delta\boldsymbol{X} = 0$ first and then chooses (a physically motivated) simplification of the equation afterwards to render the determination of $\boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}$ possible. This is the most general but worst type, as normally the approximated Euler equation no longer derives from some approximate functional. This may result in thermodynamical inconsistencies.

A *type-II approximation* modifies the form of the functional dependence, $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\cdots] \to \widetilde{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\cdots]$, to get a simpler one that allows for a solution of the resulting Euler equation $\delta\widetilde{\Omega}_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{X}]/\delta\boldsymbol{X} = 0$. This type is more particular but yields a thermodynamical potential consistent with $\boldsymbol{X}_{\boldsymbol{t},\boldsymbol{U}}$. Generally, however, it is not easy to find a sensible approximation of a functional form.

Finally, in a *type-III approximation* one restricts the domain of the functional. The restriction comes in addition to those conditions that are physically necessary anyway (e.g. normalizations) and requires a precise definition of the domain. This type is most specific and, from a conceptual point of view, should be preferred as compared to type-I or type-II approximations as the exact functional form is retained. In addition to conceptual clarity and thermodynamical consistency, type-III approximations are truely systematic since improvements can be obtained by an according extension of the domain.

Note that any type-III approximation can always be understood as a type-II one (and type-II approximations as type-I) but not vice versa.

2 Various Variational Approaches

In the following different variational principles and possible approximations shall be discussed according to this scheme, starting with Hartree-Fock and density-functional theory as illuminating examples.

2.1 Ritz Variational Approach

In the Ritz variational approach the ground-state energy is considered as a functional of the quantum state $|\Psi\rangle$. There is a generalization of the Ritz principle to arbitrary temperatures by Mermin [6]. Here the basic variable is the statistical operator ρ characterizing the system's (mixed) state, and the grand potential as a functional of ρ reads:

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho] = \mathrm{tr}(\rho(H_{\boldsymbol{t},\boldsymbol{U}} - \mu N + T\ln\rho)) \,. \tag{2}$$

Following the classical calculation of Gibbs, it can easily be shown [6] that on the set of normalized and positive definite operators, stationarity of the functional (2) is achieved for $\rho = \rho_{\boldsymbol{t},\boldsymbol{U}} = e^{-(H_{\boldsymbol{t},\boldsymbol{U}}-\mu N)/T}/\mathrm{tr}(e^{-(H_{\boldsymbol{t},\boldsymbol{U}}-\mu N)/T})$. One also has $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$. An additional feature of the functional (2) consists in the fact that $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho] \geq \Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho_{\boldsymbol{t},\boldsymbol{U}}]$ for any ρ. This "upper-bound property" is extremely helpful but specific to the Ritz principle.

For a many-electron system and an arbitrary ρ, the computation of the trace in Eq. (2) is an exponentially hard problem. A nice type-III approximation is the HF approach: Here the variational search is restricted to the subclass of *disentangled statistical operators*, i.e. statistical operators corresponding to independent-electron states. This can be made precise by introducing the important concept of a reference system:

A reference system is a system with a different (microscopic) Hamiltonian $H_{\boldsymbol{t}',\boldsymbol{U}'}$ ($\boldsymbol{t}' \neq \boldsymbol{t}$, $\boldsymbol{U}' \neq \boldsymbol{U}$) but with a macroscopic state characterized by the same values of the thermodynamic state variables as the original system (1): $V' = V$, $T' = T$ and $\mu' = \mu$. The sole purpose of the reference system is to specify the domain of the functional (2): Trial statistical operators are taken from the reference system, $\rho = \rho_{\boldsymbol{t}',\boldsymbol{U}'}$, and are varied by varying the parameters $\boldsymbol{t}'$ and $\boldsymbol{U}'$ within a certain subspace. Hence, the choice of the reference system (and the parameter subspace) defines the approximation.

The HF approximation is given by the choice $\boldsymbol{U}' = 0$ and $\boldsymbol{t}'$ arbitrary, i.e. by trial states $\rho_{\boldsymbol{t}',0} = e^{-(H_{\boldsymbol{t}',0}-\mu N)/T}/Z_{\boldsymbol{t}',0}$. Inserting into Eq. (2) yields

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho_{\boldsymbol{t}',0}] = \Omega_{\boldsymbol{t}',0} + \mathrm{tr}(\rho_{\boldsymbol{t}',0}(H_0(\boldsymbol{t}) + H_1(\boldsymbol{U}) - H_0(\boldsymbol{t}'))) \,. \tag{3}$$

The remaining trace can be computed easily using Wick's theorem as $\rho_{\boldsymbol{t}',0}$ derives from a non-interacting Hamiltonian. The variational parameters $\boldsymbol{t}'$ are

fixed by the conditions $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho_{\boldsymbol{t}',0}]/\delta\boldsymbol{t}' = 0$. These are exactly the well-known HF equations as can be seen by some straightforward manipulations.

One learns that type-III approximations can be constructed conveniently by the *concept of a reference system*. On the one hand, the reference system should comprise a large space of parameters $\boldsymbol{t}'$ and $\boldsymbol{U}'$ to generate a powerful approximation. On the other hand, the parameter space must be restricted strongly to keep the calculations feasible.

2.2 Density-Functional Approach

For a many-electron system the statistical operator or, at $T = 0$, the ground-state wave function actually is an object that is by far too complex. The *relevant* information is much more efficiently stored in integral quantities, such as the electron density. This is the starting point of density-functional theory (DFT) [7,8,9]. The density, i.e. the quantum-statistical *average* of the one-particle density operator $n(\boldsymbol{r}) = \mathrm{tr}(\rho\hat{n}(\boldsymbol{r}))$, represents the basic variable. Normally DFT aims at the inhomogeneous electron gas at $T = 0$ but can also be applied to discrete lattice models [10] and finite temperatures [6].

The grand potential $\Omega_{\boldsymbol{t},\boldsymbol{U}}$ obviously depends on the model parameters. Due to the Hohenberg-Kohn theorem [7], however, it can also be considered as a functional of the density $\boldsymbol{n}$ which is stationary at the physical density: $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}] = 0$ for $\boldsymbol{n} = \boldsymbol{n}_{\boldsymbol{t},\boldsymbol{U}}$. Furthermore, if evaluated at $\boldsymbol{n} = \boldsymbol{n}_{\boldsymbol{t},\boldsymbol{U}}$, it yields the exact grand potential: $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$. Keeping the notations introduced above, $\boldsymbol{n}$ is a matrix with $n_{\alpha\beta} = \mathrm{tr}(\rho\, c_\alpha^\dagger c_\beta)$, and the functional reads (cf. Refs. [6,10]):

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}] = \mathrm{tr}(\boldsymbol{t}\,\boldsymbol{n}) + F_{\boldsymbol{U}}[\boldsymbol{n}]\,. \tag{4}$$

Here the trace refers to the one-particle orbitals α, and $F_{\boldsymbol{U}}[\boldsymbol{n}]$ is a *universal* functional, i.e. it depends on the interaction parameters only. Using the Kohn-Sham idea [8,10], the resulting Euler equation has the form of a one-particle Schrödinger equation.

The variational principle $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}] = 0$ is rigorous but cannot be evaluated as $F_{\boldsymbol{U}}[\boldsymbol{n}]$ is generally unknown (after separating the Hartree and a kinetic-energy term, the remaining exchange-correlation functional is not explicit). Due to the universality of $F_{\boldsymbol{U}}[\boldsymbol{n}]$, however, the density-functional for a reference system with modified one-particle parameters $\boldsymbol{t}'$ reads $\Omega_{\boldsymbol{t}',\boldsymbol{U}}[\boldsymbol{n}] = \mathrm{tr}(\boldsymbol{t}'\,\boldsymbol{n}) + F_{\boldsymbol{U}}[\boldsymbol{n}]$, and thus $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}_{\boldsymbol{t}',\boldsymbol{U}}] = \Omega_{\boldsymbol{t}',\boldsymbol{U}} + \mathrm{tr}((\boldsymbol{t}-\boldsymbol{t}')\boldsymbol{n}_{\boldsymbol{t}',\boldsymbol{U}})$ which can be exploited for a type-III approximation. Choosing as a reference system $H_{\boldsymbol{t}',\boldsymbol{U}}$ the homogeneous electron gas, however, turns out to be too restrictive, as this implies a spatially constant density. The *local* density approximation (LDA) [7,8], on the other hand, has proven to be very successful [9]. At least for systems with weakly varying density it is well justified. The LDA, however, is no longer a type-III approximation but a type-II one as the *form* of the (exchange-correlation part of the) functional $F_{\boldsymbol{U}}[\boldsymbol{n}]$ is approximated to have a local dependence on the density only.

As the proof of the Hohenberg-Kohn theorem is based on the Ritz principle [7], the upper-bound property is transferred to the exact functional (4), i.e. $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}] \geq \Omega_{\boldsymbol{t},\boldsymbol{U}}$ for any $\boldsymbol{n}$, but is lost within the LDA due to the type-II character of the approximation.

2.3 Time-Dependent DFT

The weak point of the DFT consists in its inability to describe excitations (see, however, Ref. [11]). This is due to the fact that the Hohenberg-Kohn variational principle is built on the static electron density. Information on excitation properties is contained in dynamic response functions which are accessible in principle via time-dependent density-functional theory (TD-DFT) [12]. In TD-DFT one considers a situation with a time-dependent Hamiltonian and focuses on the time-dependent density $n(\boldsymbol{r},t) = \langle\Psi(t)|\hat{n}(\boldsymbol{r})|\Psi(t)\rangle$ resulting from a solution $|\Psi(t)\rangle$ of the time-dependent Schrödinger equation as the basic variable. Here the action $A = \int dt\langle\Psi(t)|i\partial/\partial t - H(t)|\Psi(t)\rangle$ can be understood as a functional of $n(\boldsymbol{r},t)$,

$$A_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}] = -\int_{t_0}^{t_1} dt\, \mathrm{tr}(\boldsymbol{n}(t)\boldsymbol{t}(t)) + B_{\boldsymbol{U}}[\boldsymbol{n}]\,, \tag{5}$$

where $\boldsymbol{t}(t)$ are explicitly time-dependent one-particle parameters. Contrary to usual DFT, the variational principle $\delta A_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{n}] = 0$ does not derive from the Ritz principle, and consequently there is no upper-bound property available. Type-II approximations can be constructed by approximating the universal but unknown part $B_{\boldsymbol{U}}[\boldsymbol{n}]$ of the functional to make it explicit. Far from equilibrium, however, there is no general recipe.

2.4 Dynamical Variational Principle

In the linear-response regime close to equilibrium, excitations are described by Green's functions. The one-electron Green's function $G_{\alpha\beta}(\omega) = \langle\langle c_\alpha; c_\beta^\dagger\rangle\rangle_\omega$ is the basic quantity in the dynamical variational approach of Luttinger, Ward, Baym and Kadanoff [3,4]. Employing a coupling-constant integration [3], the grand potential can be understood as a functional of $\boldsymbol{G}$:

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}] = \mathrm{Tr}\ln\boldsymbol{G} - \mathrm{Tr}((\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{G}^{-1})\boldsymbol{G}) + \Phi_{\boldsymbol{U}}[\boldsymbol{G}]\,, \tag{6}$$

where $\mathrm{Tr} = T\sum_n e^{i\omega_n 0^+}\mathrm{tr}$ and $\omega_n = (2n+1)\pi T$ are fermion Matsubara frequencies. Furthermore, $\boldsymbol{G}_{\boldsymbol{t},0} = (\omega + \mu - \boldsymbol{t})^{-1}$ is the $\boldsymbol{U} = 0$ (free) Green's function and $\Phi_{\boldsymbol{U}}[\boldsymbol{G}]$ the (universal) Luttinger-Ward (LW) functional defined as the sum of all dressed closed skeleton diagrams [3]. By construction, $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$. In arbitrary order in perturbation theory one has $\delta\Phi_{\boldsymbol{U}}[\boldsymbol{G}]/\delta\boldsymbol{G} = T\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]$. Therewith, the Euler equation $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}]/\delta\boldsymbol{G} = 0$ is given by $\boldsymbol{G}^{-1} - \boldsymbol{G}_{\boldsymbol{t},0}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}] = 0$ which is Dyson's equation. This shows that $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}]$ is stationary at the physical Green's function $\boldsymbol{G} = \boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$.

The LW functional is formally given by a diagrammatic sum that cannot be carried out in practice. A self-evident type-II strategy is to sum up a suitable subclass of diagrams to obtain an approximate but explicit expression for $\Phi_{\boldsymbol{U}}[\boldsymbol{G}]$. In this way the HF approximation can be recovered but there are also new approximations like the fluctuation-exchange approximation [13,14]. These "conserving approximations", however, are necessarily restricted to the weak-coupling regime.

A type-III approximation, on the other hand, would be non-perturbative by construction. Consider a reference system with modified one-particle parameters: $H_{\boldsymbol{t}',\boldsymbol{U}} = H_0(\boldsymbol{t}') + H_1(\boldsymbol{U})$. This defines the domain of the functional (6) to consist of Green's functions $\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}$ with arbitrary $\boldsymbol{t}'$. The interaction is kept fixed ($\boldsymbol{U}' = \boldsymbol{U}$). To evaluate the functional (6) at $\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}$ requires the evaluation of $\Phi_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}]$, in particular. Due to the universality of $\Phi_{\boldsymbol{U}}[\cdots]$ (no $\boldsymbol{t}$ dependence) and due to the choice $\boldsymbol{U}' = \boldsymbol{U}$, one has $\Phi_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}] = \Omega_{\boldsymbol{t}',\boldsymbol{U}} - \mathrm{Tr}\ln \boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}} + \mathrm{Tr}((\boldsymbol{G}^{-1}_{\boldsymbol{t}',0} - \boldsymbol{G}^{-1}_{\boldsymbol{t}',\boldsymbol{U}})\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}})$. Thus,

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}] = \Omega_{\boldsymbol{t}',\boldsymbol{U}} - \mathrm{Tr}(\boldsymbol{G}^{-1}_{\boldsymbol{t},0}\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}) + \mathrm{Tr}(\boldsymbol{G}^{-1}_{\boldsymbol{t}',0}\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}})\,. \tag{7}$$

Hence, on any domain specified by a suitable subspace of one-particle parameters $\boldsymbol{t}'$ which renders the solution of the reference system possible (for fixed $\boldsymbol{U}$), the functional (6) can be evaluated exactly. A possible (but oversimplified) example is the choice $\boldsymbol{t}' = 0$. It reduces the reference model to the atomic limit where the computation of $\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}$ and $\Omega_{\boldsymbol{t}',\boldsymbol{U}}$ in (7) is easy. Cluster approximations represent straightforward generalizations of this example.

Unfortunately, this type-III approach for Eq. (6) yields nothing new: Since $\mathrm{Tr}(\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{G}^{-1}_{\boldsymbol{t}',0})\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}} = \mathrm{tr}(\boldsymbol{t} - \boldsymbol{t}')\boldsymbol{n}_{\boldsymbol{t}',\boldsymbol{U}}$ with the one-electron density of the reference system $\boldsymbol{n}_{\boldsymbol{t}',\boldsymbol{U}} = T\sum_n e^{i\omega_n 0^+}\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}(i\omega_n)$, one gets $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho_{\boldsymbol{t}',\boldsymbol{U}}]$ with $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\rho]$ given by Eq. (2), i.e. the same as in the Ritz variational approach. Interestingly, this implies that upper bounds for the grand potential can be obtained, i.e. $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}] \geq \Omega_{\boldsymbol{t},\boldsymbol{U}}$ for arbitrary $\boldsymbol{t}'$.

2.5 Dynamical Mean-Field Approach

Equipped with these insights, one can address the question of *deriving the DMFT from a variational principle*. Originally, the DMFT was introduced as the exact theory of lattice models with local (Hubbard-type) interactions in infinite spatial dimensions $D = \infty$ [2]. Later on, it was recognized [15,16] that in $D = \infty$ the lattice model $H_{\boldsymbol{t},\boldsymbol{U}}$ can be self-consistently mapped onto an impurity model $H_{\boldsymbol{t}',\boldsymbol{U}}$ with the same interaction $\boldsymbol{U}$. Using this self-consistent mapping procedure as an approximation ("dynamical mean-field approximation"), one can treat lattice models for any finite D.

Instead of considering Dyson's equation in the form, $\boldsymbol{G} = (\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}])^{-1}$ (with $\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]) = (1/T)\,\delta\Phi_{\boldsymbol{U}}[\boldsymbol{G}])/\delta\boldsymbol{G}$), which is solved by the exact $\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}$, the DMFT considers the following simplified equation between *local* quantities at lattice site i:

$$(\boldsymbol{G})_{ii} = (\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}])^{-1}_{ii}\,. \tag{8}$$

Here $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}]$ is the derivative of the LW functional but with *local* vertices only as it is the case for an impurity model. Clearly, this is a type-I approximation. Eq. (8) is often called the DMFT self-consistency condition. This is because its solution is achieved by an iterative procedure in practice: Starting with a guess for $\boldsymbol{\Sigma}$, one computes the local lattice Green's function as $(\boldsymbol{G})_{ii} = (\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma})^{-1}_{ii}$ at first. This is not yet a solution of Eq. (8) since in general $\boldsymbol{\Sigma} \neq \widetilde{\boldsymbol{\Sigma}}_U[\boldsymbol{G}]$ for this $\boldsymbol{G}$. For the necessary update of $\boldsymbol{\Sigma}$ define $(\boldsymbol{G}_{\boldsymbol{t}',0})_{ii} = 1/(1/(\boldsymbol{G})_{ii} + (\boldsymbol{\Sigma})_{ii})$. Assuming that $\boldsymbol{G}_{\boldsymbol{t}',0}$ can be understood as the free impurity Green's function of an impurity model $H_{\boldsymbol{t}',\boldsymbol{U}}$ for some $\boldsymbol{t}'$, the (numerical) solution of the impurity problem yields a new $\boldsymbol{\Sigma} = \widetilde{\boldsymbol{\Sigma}}_U[(\boldsymbol{G})_{ii}] = \widetilde{\boldsymbol{\Sigma}}_U[\boldsymbol{G}]$. Iteration of this cycle until self-consistency yields a solution $\boldsymbol{G}$ of Eq. (8). Note that the resulting DMFT self-energy is local.

Here, the question is whether Eq. (8) can be interpreted as an Euler equation of some variational principle. Starting with the functional (6), one can try a type-II approximation by replacing $\Phi_{\boldsymbol{U}}[\boldsymbol{G}]$ with the LW functional of the impurity model $\widetilde{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$. This implies $(1/T)\delta\widetilde{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]/\delta\boldsymbol{G} = \widetilde{\boldsymbol{\Sigma}}_U[\boldsymbol{G}]$, and the resulting Euler equation reads: $\boldsymbol{G}^{-1} = \boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}]$. This equation is easily seen to be equivalent with Eq. (8) since $\widetilde{\boldsymbol{\Sigma}}_U[(\boldsymbol{G})_{ii}] = \widetilde{\boldsymbol{\Sigma}}_U[\boldsymbol{G}]$ by definition. Hence, DMFT can be understood as a type-II approximation.

Another functional has been suggested recently [17]:

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}] = \mathrm{Tr}\ln\frac{1}{\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]} - \mathrm{Tr}(\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]\boldsymbol{G}) + \Phi_{\boldsymbol{U}}[\boldsymbol{G}]\,. \tag{9}$$

Clearly, $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$, and furthermore the corresponding Euler equation, $\big((\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}])^{-1} - \boldsymbol{G}\big)\cdot(\delta\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]/\delta\boldsymbol{G}) = 0$, is equivalent with Dyson's equation, $\boldsymbol{G} = (\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}])^{-1}$ (assuming local invertibility of the functional $\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]$). The functional (9) therefore yields a valid variational principle. As a type-II approximation, one may replace $\Phi_{\boldsymbol{U}}[\boldsymbol{G}] \to \widetilde{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ and $\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}] \to \widetilde{\boldsymbol{\Sigma}}_U[\boldsymbol{G}] = (1/T)\delta\widetilde{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]/\delta\boldsymbol{G}$ in the functional (9). The resulting Euler equation is equivalent with the DMFT self-consistency equation (8) which implies that DMFT can also be understood as a type-II approximation to the functional (9).

Attempts to prove that a stationary point of the type-II approximated functionals (6) or (9) must be an extremum have failed [17]. Furthermore, while (as shown above) a type-III approximation to the principle based on Eq. (6) with $\boldsymbol{U}$ fixed yields upper bounds to the grand potential, the DMFT cannot be obtained as a type-III approximation starting from Eqs. (6) or (9): Choosing an impurity model $H_{\boldsymbol{t}',\boldsymbol{U}}$ as a reference system to generate trial Green's functions and to define a restricted domain of the functional (6) or (9), respectively, concurrently means that the optimal Green's function will be local. This is obviously a very poor approximation for the Green's function of a lattice model and differs from the DMFT result. The discussion shows that the question whether or not the DMFT grand potential is an upper bound to the true grand potential is still open.

2.6 Functionals of the Local Green's Function

It is also possible [18,19] to focus on the local Green's function $\boldsymbol{G}^{(\mathrm{loc})} = (\boldsymbol{G})_{ii}$ (instead of the full $\boldsymbol{G}$) and to set up a variational principle of the form

$$\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}^{(\mathrm{loc})}] = 0\,. \tag{10}$$

A functional which is stationary at the physical $\boldsymbol{G}^{(\mathrm{loc})} = \boldsymbol{G}^{(\mathrm{loc})}_{\boldsymbol{t},\boldsymbol{U}}$ and which yields $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{G}^{(\mathrm{loc})}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$ can be constructed order by order in the interaction strength [18]. Unfortunately, the diagrammatic formalism is much more cumbersome as compared to the construction of the LW functional. As is shown in Ref. [19], the dynamical mean-field approximation is equivalent with a simple ($\boldsymbol{U}$-independent) approximation to the kinetic-energy part of the functional. So the DMFT appears as a type-II approximation again.

2.7 Self-Energy-Functional Approach

The motivation to characterize the DMFT as a type-III approximation is the following: If it is possible to recover the DMFT merely by restricting the domain of the functional corresponding to an exact variational principle, different choices of the domain will place the DMFT in a systematic series of different and possibly new approximations which, as the DMFT, are all non-perturbative and thermodynamically consistent.

For this purpose it is helpful to focus on the self-energy. Within the self-energy-functional approach (SFA) [5], the self-energy functional

$$\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}] = \mathrm{Tr}\ln\frac{1}{\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma}} + F_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \tag{11}$$

is considered. Here, $F_{\boldsymbol{U}}[\boldsymbol{\Sigma}] = \Phi_{\boldsymbol{U}}[\boldsymbol{G}[\boldsymbol{\Sigma}]] - \mathrm{Tr}(\boldsymbol{\Sigma}\,\boldsymbol{G}_{\boldsymbol{U}}[\boldsymbol{\Sigma}])$ is the Legendre transform of the Luttinger-Ward functional which is well defined provided that the functional $\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]$ is invertible locally. $F_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is universal (independent of $\boldsymbol{t}$) by construction and $-(1/T)\delta F_{\boldsymbol{U}}[\boldsymbol{\Sigma}]/\delta\boldsymbol{\Sigma} = \boldsymbol{G}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ which is the inverse of the functional $\boldsymbol{\Sigma}_{\boldsymbol{U}}[\boldsymbol{G}]$. Obviously, $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t},\boldsymbol{U}}] = \Omega_{\boldsymbol{t},\boldsymbol{U}}$. The Euler equation $\delta\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}]/\delta\boldsymbol{\Sigma} = 0$ is given by $(\boldsymbol{G}^{-1}_{\boldsymbol{t},0} - \boldsymbol{\Sigma})^{-1} = \boldsymbol{G}[\boldsymbol{\Sigma}]$ and equivalent with Dyson's equation.

To construct a type-III approximation, a reference system $H_{\boldsymbol{t}',\boldsymbol{U}} = H_0(\boldsymbol{t}') + H_1(\boldsymbol{U})$ with unchanged interaction part is considered. The one-particle parameters $\boldsymbol{t}'$ are taken such that the different "correlated" sites (non-zero on-site interaction) are decoupled. Instead, $\boldsymbol{t}'$ shall include an arbitrary hopping to "bath" sites (zero on-site interaction) with arbitrary one-particle energies. In case of the Hubbard model on a lattice with L sites, the corresponding reference system constructed in this way is a set of L decoupled single-impurity Anderson models (which in case of translational symmetry are equivalent).

Trial self-energies $\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}$ are local by construction. The Euler equation resulting from this type-III approach reads $\partial\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]/\partial\boldsymbol{t}' = 0$, i.e.:

$$\left((\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}})^{-1} - \boldsymbol{G}_{\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]\right) \cdot \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}}{\partial \mathbf{t}'} = 0\,. \tag{12}$$

Now let $\boldsymbol{t}'$ (the bath parameters) be such that $\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}$ solves the DMFT self-consistency condition (8). Since $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{U}}[\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}}] = \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}$, one has $(\boldsymbol{G}_{\boldsymbol{t}',\boldsymbol{U}})_{ii} = (\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}})_{ii}^{-1}$. Hence, this $\boldsymbol{t}'$ solves Eq. (12). (Note that $\partial\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}/\partial\mathbf{t}'$ is local). So *by a restriction of the domain of the self-energy functional* (11) to local self-energies, *the DMFT is characterized as a type-III approximation.*

Interestingly, a type-II approximation does not work: A replacement of the form $F_{\boldsymbol{U}}[\boldsymbol{\Sigma}] \to \widetilde{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ in Eq. (11) yields the Euler equation $\widetilde{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] = (\boldsymbol{G}_{\boldsymbol{t},0}^{-1} - \boldsymbol{\Sigma})^{-1}$ where $\widetilde{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}] = -(1/T)\delta\widetilde{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]/\delta\boldsymbol{\Sigma}$. If this was equivalent with the DMFT self-consistency condition, a local self-energy would be a solution. This would imply, however, that $\widetilde{\boldsymbol{G}}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ is non-local for a local $\boldsymbol{\Sigma}$. Hence, $\widetilde{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$ cannot be the Legendre transform of $\widetilde{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ where $\widetilde{\Phi}_{\boldsymbol{U}}[\boldsymbol{G}]$ (as above) is the Luttinger-Ward functional with vertices restricted to a single site. An alternative choice for $\widetilde{F}_{\boldsymbol{U}}[\boldsymbol{\Sigma}]$, however, does not suggest itself.

One can conclude that a functional of the self-energy is necessary and sufficient to obtain the DMFT as a type-III approximation while a functional of the Green's function is necessary and sufficient to get the DMFT as a type-II approximation. The decisive point is that rather a local self-energy can be tolerated as an approximation than a local Green's function.

3 New Approximations

The immediate return of these considerations is a number of non-perturbative and thermodynamically consistent type-III approximations as shown in Fig. 1. These differ from the DMFT by a different restriction of the domain for the self-energy functional (11), i.e. by a different reference system with a different subspace of variational parameters $\boldsymbol{t}'$ but the same interaction ($\boldsymbol{U}' = \boldsymbol{U}$). The evaluation of a type-III approximation requires the repeated computation of the grand potential and the Green's function or self-energy of the reference system to get $\Omega_{\boldsymbol{t},\boldsymbol{U}}[\boldsymbol{\Sigma}_{\boldsymbol{t}',\boldsymbol{U}}]$ which must be optimized with respect to $\boldsymbol{t}'$ subsequently.

The class of possible reference systems is essentially spanned by two parameters, namely $n_{\mathrm{s}} - 1$ which is the number of additional bath sites per correlated site and N_{c} which is the number of correlated sites in a cluster that is decoupled from the rest of the correlated sites (Fig. 1). The DMFT is obtained for $N_{\mathrm{c}} = 1$ and $n_{\mathrm{s}} = \infty$ since a continuous bath ($n_{\mathrm{s}} = \infty$) is necessary to represent an arbitrary local free Green's function.

The choice $N_{\mathrm{c}} = 1$ but $n_{\mathrm{s}} < \infty$ yields new approximations ("dynamical impurity approximations", DIA) which are inferior as compared to the full DMFT but allow for much simpler and faster calculations when n_{s} is small.

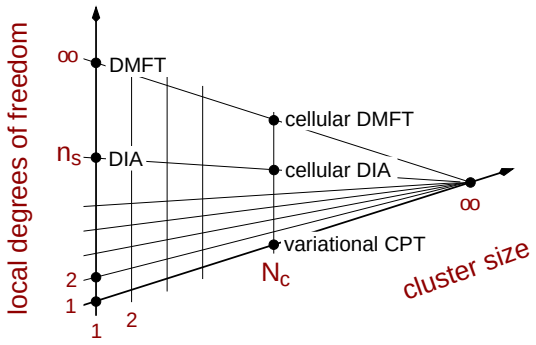

Fig. 1. Different possible approximations within the SFA (see text).

The most simple but non-trivial approximation ($n_s = 2$-DIA) has been shown [5,20] to already cover the essence of the DMFT scenario [1] for the Mott metal-insulator transition in the Hubbard model. At the critical point for $T = 0$ the calculations can be done even analytically [20], and with increasing n_s the grand potential, static quantities and the entire phase diagram rapidly converge to the full DMFT results [5,21]. The DIA is similar but superior as compared to the exact-diagonalization approach [1]. Even for small n_s the approach is thermodynamically consistent and, off half-filling, respects the Luttinger sum rule, for example. The DIA has also been employed successfully to study the influence of phonons on metal-insulator transitions in the Holstein-Hubbard model [22,23].

Nothing new is obtained for $n_s = \infty$ and $N_c > 1$: Here the SFA recovers the cellular DMFT [24]. (Note that the dynamical cluster approximation [25] is a type-II approximation). More interesting is the case $n_s = 1$ and $N_c > 1$ which turns out [26] to represent a variational generalization of the cluster-perturbation theory [27]. This V-CPT is well suited to describe phases with spontaneously broken symmetry and has been employed to study one-particle excitations and antiferromagnetic order in the $D = 2$ and $D = 1$ Hubbard model at half-filling [28] and charge ordering in the extended Hubbard model [29]. A further application concerns antiferromagnetism in quarter-filled ladder systems [30]. An impressing example of the power of the V-CPT approach has been given recently in a study of the competition between antiferromagnetism and d-wave superconductivity in the hole- and electron-doped Hubbard model [31]. The question of phase separation is addressed in Ref. [32].

Summing up, the SFA is able to unify different cluster theories and local approximations within a single and consistent framework which offers a large flexibility in the use of bath sites, ficticious fields, boundary conditions and particle reservoirs [26]. The formalism provides a controlled compromise between the demands for a non-perturbative and systematic theory working in the thermodynamic limit on the one hand, and the limited computational capabilities to diagonalize finite-size systems on the other.

4 Open Problems

The self-energy-functional approach allows to construct a series of consistent approximations which improve systematically as $N_c \to \infty$. It is by no means clear, however, whether bath sites $n_s > 1$ help to speed up the convergence with respect to N_c and whether a cluster extension of DMFT or the V-CPT is more efficient. This can be clarified only empirically by considering different lattice models in different dimensions. As a few bath sites have turned out to be sufficient for reproducing the essential mean-field ($N_c = 1$) physics in a number of studies of the single-band Hubbard model, further applications of the DIA are worthwhile to explore e.g. the mean-field phase diagrams of more complex (multi-orbital) models. Furthermore, one may also envisage the application of a simplified DMFT where a single (but continuous, $n_s = \infty$) bath is optimized for a multi-orbital model. This might be well justified for not too low temperatures.

On the technical side, there are two main future tasks: The full diagonalization and the Lanczos method which have been used so far, should be supplemented by a "reference system solver" based on stochastic techniques to improve the scaling of the numerical effort with the system size. Secondly, it would be advantageous to have an iterative technique at hand that directly yields a solution of the SFA Euler equation without the need for numerical differentiation. First results using full diagonalization [21] are encouraging.

On the conceptual side, the question for the possibility to give strict upper bounds to the grand potential is still open. Probably, a positive answer requires to establish a link to the Ritz variational principle. On the other hand, no example is known yet where the SFA grand potential at a stationary point is *lower* than the exact one.

There are different directions into which the formalism may be extended. As the coherent-potential approximation for the disorder Anderson model has the same (mean-field) status as the DMFT for the Hubbard model, it suggests itself that a self-energy-functional approach can also be constructed for systems with disorder (and interaction). First applications [33] demonstrate that such a theory [34] is feasible. A challenge consists in the extension of the theory to include two-particle Green's functions in a generalized variational principle. Here the recently proposed functional-integral formulation of the SFA [35] can be helpful. Two-particle correlation functions are interesting by themselves and may furthermore facilitate an even greater flexibility in the choice of reference systems. At the same time such an approach could provide a conceptual clear way to treat models with non-local interactions. Currently, this problem is circumvented by a more pragmatic decoupling procedure [29].

The author would like to thank M. Aichhorn, E. Arrigoni, F.F. Assaad, M. Balzer, R. Bulla, C. Dahnken, W. Hanke, A. Millis, and W. Nolting for discussions. Support by the Deutsche Forschungsgemeinschaft within the Sonderforschungsbereich 410 and the Forschergruppe 538 is acknowledged.

References

1. A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
2. W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
3. J. M. Luttinger and J. C. Ward, Phys. Rev. **118**, 1417 (1960).
4. G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
5. M. Potthoff, Euro. Phys. J. B **32**, 429 (2003).
6. N. D. Mermin, Phys. Rev. **137**, A 1441 (1965).
7. P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864 (1964).
8. W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).
9. R. M. Dreizler and E. K. U. Gross, *Density Functional Theory*, Springer, Berlin (1990).
10. K. Schönhammer, O. Gunnarsson, and R. M. Noack, Phys. Rev. B **52**, 2504 (1995).
11. L. J. Sham and W. Kohn, Phys. Rev. **145**, 561 (1966).
12. E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
13. N. E. Bickers, D. J. Scalapino, and S. R. White, Phys. Rev. Lett. **62**, 961 (1989).
14. N. E. Bickers and D. J. Scalapino, Ann. Phys. (N.Y.) **193**, 206 (1989).
15. A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
16. M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992).
17. R. Chitra and G. Kotliar, Phys. Rev. B **63**, 115110 (2001).
18. R. Chitra and G. Kotliar, Phys. Rev. B **62**, 12715 (2000).
19. A. Georges, cond-mat/0403123.
20. M. Potthoff, Euro. Phys. J. B **36**, 335 (2003).
21. K. Pozgajcic, cond-mat/0407172.
22. W. Koller, D. Meyer, Y. Ono, and A. C. Hewson, **66**, 559 (2004).
23. W. Koller, D. Meyer, A. C. Hewson, and Y. Ono, cond-mat/0406241.
24. G. Kotliar, S. Y. Savrasov, G. Pálsson, and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001).
25. M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke and H. R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998).
26. M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003).
27. D. Sénéchal, D. Pérez, and M. Pioro-Ladrière, Phys. Rev. Lett. **84**, 522 (2000).
28. C. Dahnken, M. Aichhorn, W. Hanke, E. Arrigoni, and M. Potthoff, Phys. Rev. B **70**, 245110 (2004).
29. M. Aichhorn, H.G. Evertz, W. von der Linden, and M. Potthoff, Phys. Rev. B **70**, 235107 (2004).
30. M. Aichhorn, E. Ya. Sherman, and H. G. Evertz, cond-mat/0409162.
31. D. Sénéchal, P.-L. Lavertu, M.-A. Marois, and A.-M.S. Tremblay, cond-mat/0410162.
32. M. Aichhorn and E. Arrigoni, cond-mat/0502047.
33. M. Balzer and M. Potthoff, Physica B, in press.
34. M. Potthoff and M. Balzer, to be published.
35. M. Potthoff, cond-mat/0406671.