Mott Transition in the $d=\infty$ Hubbard Model at Zero Temperature

X. Y. Zhang, M. J. Rozenberg, and G. Kotliar
Serin Physics Laboratory, Rutgers University, Piscataway, New Jersey 08854
(Received 3 August 1992)

Using a combination of perturbation theory and quantum Monte Carlo, we elucidate the behavior of the single-particle Green's function and the local spin-spin correlation function near the Mott transition in the infinite dimensional Hubbard model at half filling. The model has three fixed points: the Fermi liquid phase qualitatively described by the Brinkman-Rice picture, the insulating phase resembling Hubbard's solution, and an unstable Mott point which connects the two.

PACS numbers: 71.30.+h

In his pioneering work on the metal-insulator transition (MIT) [1], Mott envisioned that in transition metals, as the Coulomb interaction among the charged carriers increases, the system will undergo a first-order transition from a metal to an insulator. The first serious attempt using many-body theory to produce this effect is due to Hubbard [2]. He based his calculation on the atomic limit which naturally leads to a two-band picture, the lower and upper Hubbard bands separated by the interaction U. He concluded that the MIT happens at $U_c \approx D$ the bandwidth, and a gap opens gradually as a function of U. Although this treatment provides a good insulating solution for large U, it fails to capture correctly the low-energy physics in the metallic side: the Fermi liquid quasiparticles are absent [3]. Brinkman and Rice (BR) [4] attacked the problem from the opposite limit by using a Gutzwiller variational wave function, and found the MIT at a much higher U_c. The Gutzwiller wave function gives a good Fermi liquid description for the metallic side, but misses the insulating side completely and lacks the high-energy excitations which are the precursors, in the metal, of the upper and lower Hubbard bands of the insulating solution.

Until now, it is still not clear what is the right picture of the transition: How is Hubbard's solution related to BR's? Does the gap open continuously or does it jump at the transition? Behind these questions lies the far more interesting issue, How does Fermi liquid theory break down when the interactions become strong?

Recently Metzner and Vollhardt [5] recognized a simple but nontrivial limit of the Hubbard model: large dimensionality. In that limit, the Gutzwiller approximation used by BR becomes exact, and a major assumption made in Hubbard III, the self-energy being site diagonal, also becomes exact. In fact, a set of self-consistent equations that describe the paramagnetic phase of the Hubbard model for $d \to \infty$ [6,7] is reminiscent of the Hubbard III treatment. Therefore, it is natural that a solution in the $d=\infty$ limit should provide a bridge between the Hubbard III and the BR treatments and a clear picture of the Mott transition.

Recently, quantum Monte Carlo (QMC) simulations were used for solving these equations by three groups independently [8], and all found metallic solution for small U and insulating state for large U, thus confirming that a transition indeed exists. But, due to the limitations of QMC, the picture of the transition is still unclear.

Our strategy here is to combine QMC with a more conventional tool: the perturbation expansion. QMC, being essentially exact, is used as a benchmark to select among the various perturbative schemes the one that works. Perturbation is then used in the low-temperature region that QMC fails to reach, to obtain important information like the density of states. The result is a complete numerical solution of the self-consistent equations at half filling and $T=0$. The metallic side exhibits aspects of BR's solution, particularly, the divergence of the effective mass of quasiparticles as $U \to U_c$, and the insulating side is very similar to that of the Hubbard III. The emergence of a three band picture close to the transition, the two well separated upper and lower Hubbard bands and a central narrow quasiparticle band at the Fermi level, provides a bridge between the two approaches. The continuous narrowing of the quasiparticle peak drives the MIT. Right above the transition, a full gap is already opened, as predicted by Mott for a different reason. This is valid as long as the density of states (DOS) is bounded. We focus on this type of DOS, because we believe it provides the correct scenario to understand the physics in finite dimensions.

The central object in the self-consistent scheme [6] is a quantity G_0 which plays the role of the effective field in magnetic systems. G_0 is defined in an effective local action S obtained by integrating out all the degrees of freedom except for a single site 0,

$$S[G_0] = -\int\int d\tau\, d\tau' c_{0\sigma}^{\dagger} G_0^{-1} c_{0\sigma} + \int d\tau\, U(n_\uparrow - \tfrac{1}{2})(n_\downarrow - \tfrac{1}{2}) .$$

The self-consistent equations for the Weiss field G_0 are written in terms of an impurity self-energy $\Sigma_{\rm imp}(G_0) = \langle c^{\dagger} c \rangle_{S(G_0)}^{-1} - G_0^{-1}$ and the lattice density of states $\rho(\epsilon) = \sum_k \delta(\epsilon_k - \epsilon)$,

$$[G_0^{-1} - \Sigma_{\rm imp}(G_0)]^{-1} = \int \frac{\rho(\epsilon)\, d\epsilon}{i\omega + \mu - \epsilon - \Sigma_{\rm imp}(i\omega)} . \qquad (1)$$

The impurity self-energy evaluated at the self-consistent G_0 [6] gives the self-energy of the Hubbard model in infinite dimensions.

We use a semicircle density of states as in Hubbard III, $\rho(\epsilon) = (2/\pi D)[1 - (\epsilon/D)^2]^{1/2}$, which corresponds to a

Bethe lattice in infinite d, where Hubbard's hopping parameter $t=D/2$ [9]. The set of self-consistent equations then becomes

$$G_0^{-1}=i\omega_n+\mu-t^2G(i\omega_n),\quad G=\langle c^\dagger c\rangle_{S(G_0)}, \qquad (2)$$

$G(\omega_n)$ being the *local* Green's function of the Hubbard model. The spin index has been removed because paramagnetic phase is assumed. Throughout the calculation the bandwidth D is set to unity.

Methodology.—The key part in solving the above equation is to obtain the self-energy Σ given G_0. This is equivalent to solving the Anderson impurity model with an arbitrary hybridization function [6]. Although QMC is exact, it has severe limitations: The data collected are on the imaginary axis and the analytic continuation cannot be always reliably carried out; a more essential difficulty is that the method does not allow one to reach the arbitrarily low temperatures. Strictly speaking, there should be no MIT at any finite temperatures but only a sharp crossover at a value of the interaction U_c. U_c is determined by the onset of the divergence in the Matsubara self-energy at low frequencies.

The perturbative approaches to the impurity problem fall into four categories: (a) fully self-consistent perturbation in G_{ii} [10]; (b) noncrossing approximation (NCA) [11]; (c) analytic expansion in U as in Yamada and Yosida [12]; (YY); and (d) a fourth approach proposed by Metzner [13], who systematized the expansion in the kinetic energy. It is discussed in [6] that (a) does not produce the correct high-energy features, while it is well known that the NCA becomes inaccurate at low temperatures. We will show below that approach (c) can successfully reproduce the QMC data, and argue why it does so.

First, it is crucial to demonstrate that the perturbative calculation is able to reproduce all the features observed in QMC at a temperature lower than the lowest energy scale of the problem. Only then applying the perturbative calculation at $T=0$ becomes meaningful. Near the MIT, the low-energy scale is set by ϵ_f^*: the width of the quasiparticle peak obtained from the zero temperature perturbative calculation. Figure 1 shows U_c as a function of temperature and ϵ_f^* as a function of U. It is evident from this figure that the low-energy physics is only observed when $T<e_f^*$. Figure 2 shows a comparison of QMC with perturbation calculations using (c) for values of U on the metallic side and the insulating side at a $T=1/32D$. With almost point by point fit on both sides, *a posteriori* one can rationalize the success of the perturbative calculation as follows: (1) The transition happens at an intermediate value of U, around which it is known that perturbation to second-order captures all the important features of the Anderson impurity model as was shown by YY; (2) the essential ingredient that drives the transition is not in the self-energy calculation, but rather in the process of enforcing self-consistency. Higher-order corrections to Σ will only change the exact value of U_c, but not the nature of the transition. This will become clear after the subsequent analytical discussion of the Mott transition. (3) From the pioneer work of YY and Zlatic, Horvatic, and Sokcevic [14] we know that the impurity problem is analytic in U. Therefore it is natural to treat the analytic parts of the problem using a power series expansion while the nonanalytic aspects of the problem are addressed using a self-consistent scheme. (4) It can produce the atomic solution exactly (see below). Therefore this approximation becomes at least an interpolation between the two extreme limits (free case $U=0$, and atomic $t=0$). The latter is purely ac-

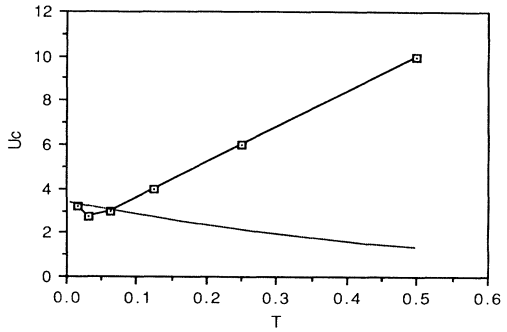

FIG. 1. The critical U_c as a function of temperature dividing the metallic and insulating regions. The phase boundary is a crossover. A true transition occurs only at $T=0$. U vs the width of the quasiparticle peak at $T=0$ is also shown (dotted line). The width sets the low-temperature scale of the problem. The transition is driven by the interaction U instead of the temperature only below this scale.

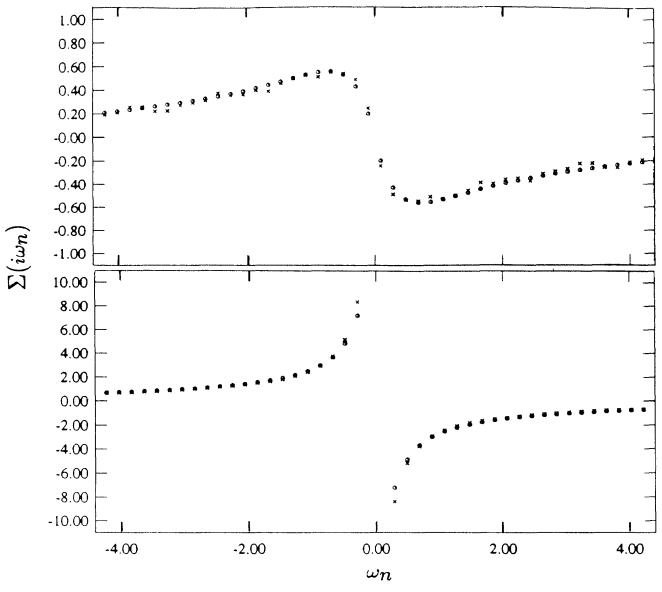

FIG. 2. Comparison of Matsubara self-energy obtained from quantum Monte Carlo (crosses) to that from perturbation calculation (dots) for $U=2$ on the metallic side and $U=3.6$ on the insulating side at a $T=1/64$. Note that on the insulating side the growing deviation at low energies is purely due to the ω^{-1} divergency, where the relative error remains the same but the absolute error becomes bigger.

cidental, and unfortunately does not apply to the asymmetric case.

We calculate the self-energy to second order following [12] and [6]. The perturbation expansion is generated by $H_I = U(n_\uparrow - \frac{1}{2})(n_\downarrow - \frac{1}{2})$. To second order, only the bubble diagram survives, $\Sigma(t) = U^2 G_0(t)^2 G_0(-t)$. It is worth noting here that the same expression for the self-energy holds both in the zero temperature and in the Matsubara formalism. The zero-temperature causal Green's function is then used to obtain the density of states and other quantities.

Mott transition.—An observable that signals the MIT is the density of states (DOS) at the Fermi surface. Figure 3 shows the zero-temperature DOS as a function of interaction U. As pointed out by Müller-Hartmann [10], the height of the peak at zero frequency is unrenormalized by the interactions in a theory with a k independent self-energy, as long as one is in the Fermi liquid regime. Therefore the value of the one-particle spectral function jumps discontinuously to zero at the Mott transition. Right after the transition, the structure of a full gap is already in place.

Since $\mathrm{Im}G_0^{-1} = -(D^2/4)\mathrm{Im}G$, the width $\Delta \equiv \epsilon_f^*$ of the quasiparticle at the Fermi level is also the width of the coherent hybridization $\Delta(\omega)$ in the Anderson model analogy. As U approaches U_c, this scale vanishes in the following way: G_0 develops a singularity at an energy $\omega_0 = \sqrt{\Delta D}$, very near the real axis ($\sim\Delta^2$), which then leads to a similar singularity in the self-energy. The Mott transition happens when the position of the pole collapses to zero, and Σ becomes divergent.

A different view of the transition is obtained by measuring the local spin-spin correlator using QMC. On the metallic side, $\langle m(\tau)m(0)\rangle \approx e^{-\Delta\tau}$. Long-range order sets in as $\Delta \to 0$ as shown in Fig. 4(a). A spin mode whose energy is independent of momentum (local paramagnon) is softening as $U \to U_c$.

We now proceed to provide an analytical description of the transition, focusing on the most important quantity Δ. Using the relation (2), we rewrite G_0 as, $G_0^{-1}(z) = z - (D^2/4)\int \bar{\rho}(\epsilon)d\epsilon/(z-\epsilon)$, where $\bar{\rho}(\epsilon)$ is the DOS of the fully interacting system shown in Fig. 3. Its three peak feature can be represented by a Lorentzian at the center with width Δ, and two semicircles centered at $\pm U/2$, with width D. The weight of the Lorentzian peak is determined by demanding that $\mathrm{Im}G_0^{-1}(0) = D/2$. A simple and good approximation for G_0, close to the transition can then be obtained: $G_0^{-1}(z,\Delta) = z - D\Delta/(z + i\Delta \mathrm{sgn}z)$. We checked numerically that the parametrized G_0 and the self-energy calculated from it are in good agreement with the actual G_0 and Σ. The impurity self-energy has an explicit Δ dependence when calculated from the parametrized G_0. $\Sigma(\omega,\Delta) = U^2 \int F^3(\lambda) e^{-|\lambda|\omega} d\omega$, where $F(\lambda) = \int_0^\infty \rho(\epsilon) e^{-i\lambda\epsilon} d\epsilon$, and $\pi\rho(\omega) = -\mathrm{Im}G_0 = 2D\Delta^2/[(\omega^2 - D\Delta)^2 + \Delta^2\omega^2]$. As $\Delta \to 0$, $\rho(\omega)$ peaks sharply at $\omega_0^2 = D\Delta - \Delta^2/2$, approaching a δ function with unit weight. Thus, for low frequencies, we find the real part of the self-energy,

$$\Sigma(\omega,\Delta) = -U^2\omega/9(D\Delta - \Delta^2/2). \quad (3)$$

On the other hand, the self-consistency equation for G_0 in Eq. (1),

$$G_0^{-1}(\omega,\Delta) = \tfrac{1}{2}(\omega + \Sigma(\omega,\Delta) + \{[\omega - \Sigma(\omega,\Delta)]^2 - D^2\}^{1/2}),$$

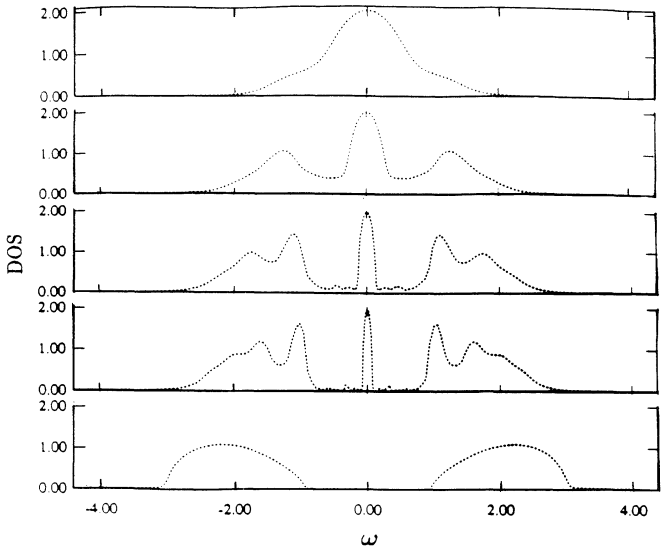

FIG. 3. Density of states $-\mathrm{Im}G$ at zero temperature as a function of U. From top to bottom, $U = 1, 2, 2.7, 3, 4$.

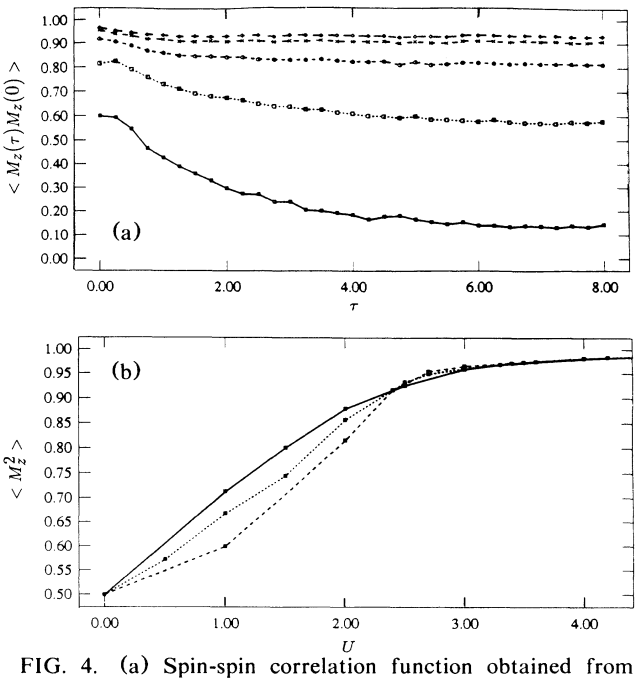

FIG. 4. (a) Spin-spin correlation function obtained from Monte Carlo for $\beta = 16$, and $U = 1, 2, 2.4, 2.7, 3$. The correlation length $\sim\Delta^{-1}$ diverges as we approach the insulating side. (b) Local moment as a function of U for three different temperatures: $\beta = 4$ (solid line), 8 (dotted line), 16 (dashed line).

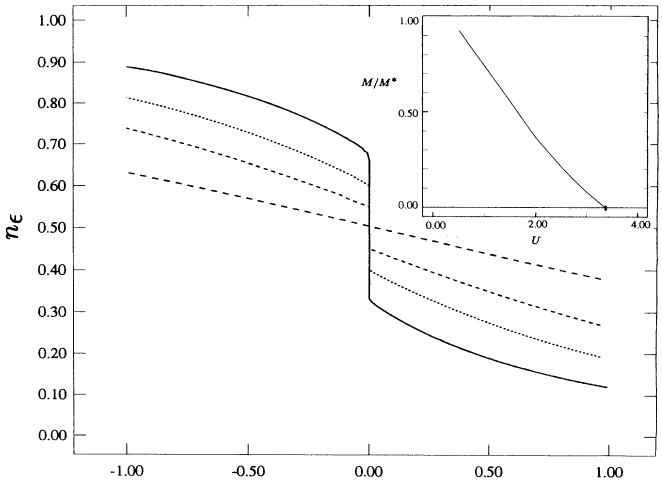

FIG. 5. Particle occupation number vs noninteracting energy ϵ as a function of $U=2,2.5,3,4$ obtained from perturbative calculations. Inset: m/m^* vs U calculated from the slope of the real part of the self-energy at the Fermi level.

requires that to linear order in ω, $\Sigma=(D/\Delta)\omega$. Since the self-consistency is an iterative procedure, we can equate the above two expressions for the self-energy iteratively:

$$\Delta_{n+1}=(9D/U^2)(D\Delta_n-\Delta_n^2/2)\,. \tag{4}$$

There are two fixed points of the iteration, $\Delta^*=2D(1-U^2/U_c^2)$ which is stable for $U<U_c=3D$, and $\Delta^*=0$. The first one is the Fermi liquid fixed point discussed in [6]; the second one is the insulating fixed point. We focus here on the fate of the metallic state. The destruction of the $\Delta=0$ (insulating) solution will be discussed elsewhere. When $U=U_c$ the two fixed points merge into one: the unstable Mott point. We note here that $\Delta=0$ is the exact atomic solution, because from the parametrized G_0 we have $G_0(\tau)=\theta(\tau)-\frac{1}{2}$, and, consequently, $\Sigma(\tau)=(U^2/4)[\theta(\tau)-\frac{1}{2}]$, or $\Sigma=(U^2/4)G_0$ (see Rozenberg, Zhang, and Kotliar [8]).

Since $\Delta/D\approx m/m^*$, the fixed point solution also provides the critical behavior of the effective mass close to the transition, $m^*/m\propto[1-(U/U_c)^2]^{-1}$ which is the same as BR. The numerical results are plotted in Fig. 5. The transition occurs at $U_c=3.37D\approx 8\bar{\epsilon}$, as in BR. However, above the transition in the insulating side, the two band features are not quite symmetric around $\pm U/2$, meaning there are still some residual kinetic energies gained from local hoppings [15]. The fact that virtual double occupancy is present in the insulating side which gives a finite exchange constant is also clear from QMC measurements of the *local* spin-spin correlation function, which shows a reduced moment on the insulating side, $\langle m_z^2\rangle=1-2\langle n_{i\uparrow}n_{i\downarrow}\rangle$, Fig. 4(b). QMC also indicates that the paramagnetic spin susceptibility does not diverge as the effective mass does [16]. These are the crucial differences between the exact solution in $d=\infty$ and that of BR.

Another important quantity is the occupation number n_k. Since Σ is independent of k, it is more convenient to label n with the noninteracting energy ϵ, $n_\epsilon=-\pi^{-1}\times\mathrm{Im}\int_{-\infty}^{0}d\omega/[\omega-\epsilon-\Sigma(\omega)]$. Figure 5 shows, in accordance with the divergence of m^*, the jump at the Fermi level Z continuously shrinking to zero. The monotonic behavior found here should be contrasted with the results using Gutzwiller wave function in 1D [17].

In conclusion, we have found a simple and reliable perturbative scheme to treat the impurity model. It enables us to provide a detailed analysis of the Mott transition in the $\infty-d$ limit. Our solution connects two very different approaches, the BR and the Hubbard III. In particular it shows a continuous increase of the effective mass as $U\rightarrow U_c$ from below, and contrary to one's intuition, followed by the discontinuous opening of a gap in the one-particle spectrum at U_c. On the insulating side, there is an unsaturated local magnetic moment with long-range order in the imaginary time indicating a nonzero double occupancy and consequently a finite magnetic exchange and a nondivergent magnetic susceptibility.

We would like to thank Elihu Abrahams, P. Coleman, J. Ferrer, F. Gebhard, A. George, A. Ruckenstein, Q. Si, P. Wolfle, and V. Yakovenko for useful discussions. This work was supported by the NSF under Grant No. DMR 92-24000.

[1] N. F. Mott, Philos. Mag. **6**, 287 (1961).
[2] J. Hubbard, Proc. R. Soc. London A **281**, 401 (1964). This paper is commonly known as the Hubbard III.
[3] D. M. Edwards and A. C. Hansen, Rev. Mod. Phys. **40**, 810 (1968).
[4] W. F. Brinkman and T. M. Rice, Phys. Rev. B **2**, 4302 (1970).
[5] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
[6] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992).
[7] V. Janis, Z. Phys. B **83**, 227 (1991); Phys. Rev. B **40**, 11 331 (1989); V. Janis and D. Vollhardt, Int. J. Mod. Phys. B **6**, 731 (1992).
[8] M. Jarrel, Phys. Rev. Lett. **69**, 168 (1992); M. J. Rozenberg, X. Y. Zhang, and G. Kotliar, Phys. Rev. Lett. **69**, 1236 (1992); Antoine Georges and Werner Krauth, Phys. Rev. Lett. **69**, 1240 (1992).
[9] P. G. J. van Dongen, Mod. Phys. Lett. B **5**, 861 (1991).
[10] E. Müller-Hartmann, Z. Phys. B **74**, 507–512 (1989); **76**, 211–217 (1989).
[11] T. Pruschke, D. Cox, and M. Jarrel (to be published)
[12] K. Yamada, Prog. Theor. Phys. **53**, 970 (1975); K. Yosida and K. Yamada, Prog. Theor. Phys. **53**, 1286 (1975).
[13] W. Metzner, Phys. Rev. B **43**, 8549 (1991).
[14] V. Zlatic, B. Horvatic, and D. Sokcevic, Z. Phys. B **59**, 151 (1985).
[15] This observation was made by Florian Gebhard.
[16] M. Jarrel and Thomas Pruschke (to be published).
[17] H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **56**, 1490 (1987); C. Gros, R. Joynt, and T. M. Rice, Phys. Rev. B **36**, 381 (1987); W. Metzner and D. Vollhardt, Phys. Rev. Lett. **59**, 121 (1987); Phys. Rev. B **37**, 7382 (1988).