Comments on theories of high temperature superconductivity

T M Rice

Theoretische Physik, ETH Zürich, 8093 Zürich, Switzerland

Abstract

The recently discovered MgB_2 superconductors have a record transition temperature for a BCS superconductor due to the high vibration frequencies associated with its light elements. The transition temperatures in the cuprate family of superconductors are much higher but these do not fit the BCS paradigm. The most promising microscopic origin for their many anomalous properties lies in magnetic pairing described by the RVB (Resonant Valence Bond) ansatz. However a comprehensive theoretical description of the key anomalous properties of the cuprates remains to be an open challenge.

Keywords: RVB, Cuprate super conductivity, MgB_2, t-J model

1. Introduction

Superconductivity was discovered by Kamerlingh Onnes in 1911 when he cooled Hg to a temperature below 4 K and observed a sudden drop of the resistivity to an immeasurably small value. Of course at that time before quantum mechanics, the electrical properties of metals etc. were not understood. The discovery of quantum mechanics in the nineteen twenties which led immediately to breakthroughs in our understanding of normal metals, magnetism etc., however did not shed any light on the origin of superconductivity. For three further decades it remained the most glamorous, elusive and mysterious phenomenon in physics. Then in the mid-fifties this changed dramatically with the theoretical breakthrough by Bardeen, Cooper and Schrieffer [1] which overnight transformed superconductivity from the least understood to the best understood phenomenon in metals.

The transition temperature T_c is a key property and considerable effort was spent looking for new superconductors with higher T_c. However this had only modest success for many decades. The BCS breakthrough did not change things. I started research in the early sixties and for the next quarter century the highest T_c remained stuck at values slightly above 20 K. The optimists argued that $T_c \sim 30$ K should be possible but the lack of any movement in T_c made even this value seem to be a pipe dream.

This all changed abruptly with the dramatic discovery by Bednorz and Müller[2] in 1986 of superconductivity above 30 K in $La_{2-x}Ba_xCuO_4$. This breakthrough quickly led to dramatic rises in T_c as more cuprates with slightly hole doped CuO_2-planes were synthesized. The culmination came in May '93 when Schilling and Ott[3] found the record value of T_c = 133 K in $HgBa_2Ca_2Cu_3O_{8+\delta}$ – again in Switzerland. All of the many high T_c cuprates, such as LSCO ($La_{2-x}Sr_xCuO_4$, aka 214), YBCO ($YBa_2Cu_3O_{7-\delta}$, aka 123 and $YBa_2Cu_4O_8$, aka 124), BSCCO ($Bi_2Sr_2CaCu_2O_{8+\delta}$, aka 2212) and many others, are all variants on a single theme namely lightly hole doped CuO_2 planes. [NB There are also a few electron doped cuprates but we will concentrate here the more important hole doped cuprates.] The high-T_c phenomenon is restricted to this narrow class of materials. Note there are many transition metal oxides which are metallic but these are only rarely superconducting and in many cases are even 'vegetables' such as RuO_2, ReO_3 etc. which show nice metallic properties but no phase transition of any kind down to the lowest temperatures. Clearly there must be something very special and unique about this class of cuprates.

Also of special interest is the recent discovery of a superconductor MgB_2 with a T_c = 39 K by Akimitsu and coworkers in 2001[4]. This first surprise of the twenty first century is also of interest and we will discuss it in the next section.

2. Short review of BCS theory and its application to MgB_2

The aim in this short review is just to repeat some key points of the BCS theory which are relevant if one wishes to apply it to the cuprates and also to MgB_2. The superconducting state arises from the pairing instability

of electrons, or more accurately quasiparticles, when one uses Landau Fermi liquid theory to describe the effects of the repulsive Coulomb interactions. Assuming a matrix element V (independent of momentum transfer) for a scattering process of a Cooper pair ($\mathbf{k}\uparrow$, $-\mathbf{k}\downarrow$, energy 2ε) to ($\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow$). An elastic scattering process is modified to second order in V by virtual processes involving all intermediate pair states ($\mathbf{p}\uparrow, -\mathbf{p}\downarrow$: energy 2ξ) leading to a form for the total matrix element

$$\Gamma\left(\varepsilon/D,V\right) = V + \frac{V^2}{\Omega} \sum_{p}^{\xi<D} \frac{2}{2(\varepsilon-\xi)} + ..., \qquad (1)$$

with Ω, volume D: upper energy cutoff. This correction diverges logarithmically ($\sim \ln(D/\varepsilon)$). In higher orders the leading divergences come from independent sums over intermediate states, leading to a geometric series which can be resumed to give

$$\Gamma\left(\varepsilon/D,V\right) = V/\left(1 + N_0 V \ln\left(D/\varepsilon\right)\right), \qquad (2)$$

where N_0 is the density of states. This form for Γ has a scaling behavior (dubbed poor man's scaling by Anderson). When the cutoff is reduced from D to D'

$$\Gamma\left(\varepsilon/D',V'\right) = \Gamma\left(\varepsilon/D,V\right), \qquad (3)$$

with the renormalized interaction $V' = V/\left(1 + N_0 V \ln\left(D/D'\right)\right)$. The consequences are that a repulsive interaction ($V>0$) scales away to zero as D' is reduced. But the strength of attractive interactions ($V<0$) grows leading to the Cooper instability. In simple metals there is both a repulsive Coulomb term with a large cutoff $\simeq \varepsilon_F$ (Fermi energy) and an attractive term due to exchange of phonons with a small cutoff $\simeq \theta_D$ (Debye energy). To combine these two interactions one must first rescale the Coulomb term to a cutoff θ_D and then one can simply add them. This procedure leads to the standard result

$$T_c = 1.1\,\theta_D \exp\left(-1/\left(\lambda - \mu^*\right)\right), \qquad (4)$$

with λ as the dimensionless electron-phonon interaction $N_0\left\langle V^{ep}(\mathbf{k} - \mathbf{k}')\right\rangle$ averaged over momentum transfers.

The so called Coulomb pseudopotential $\mu^* = \mu/(1 + \mu \ln(\varepsilon_F/\theta_D))$ has a typical value in the range 0.1 to 0.2, since $\varepsilon_F/\theta_D \sim 10^2$ and $\mu \sim 1$.

In the old days it was widely believed that the way to get a high T_c was to look for metals with ε_F near to a peak in the density of states, N_0, e. g. in the A15 compounds V_3Si, Nb_3Sn. However there is a catch since increasing λ reduces the phonon energies and can lead to a lattice instability caused by a phonon going soft. Note the symmetry of the superconducting state is always simply s-wave since the electron-phonon interaction is attractive for all momentum transfers so a sign change in the pairing amplitude around the Fermi surface is clearly detrimental.

The surprising discovery of a $T_c = 39$ K in MgB_2 a couple of years ago[4] showed a certain myopia and even embarrassment that this readily available metal had been overlooked. MgB_2 consists of graphite-like honeycomb B-layers interspaced with hexagonal Mg-layers. The transfer of 2 electrons/Mg (or 1 electron/B) means that the B-layers are effectively isoelectronic with graphite. However the much stronger interlayer coupling mediated by the Mg-layers leads to a very different Fermi surface. Unlike graphite it has sheets in both σ- and π-bands of the $2p$ electrons in the B-layers. It turns out that these σ-sheets, which do not exist in graphite, are crucial. There is strong electron-phonon coupling on this part of the Fermi surface to a specific B-B optic phonon. The calculations of Kong et al. [5] illustrate this point. The value of $\lambda \sim 0.9$ is not exceptionally large but when combined with the high frequency of the optic phonon due to the light mass of B, a high T_c results. The optic mode frequency is considerably renormalized but because of the small ionic masses it still remains quite high. So far related materials have not been found, apparently due to the difficulty of synthesizing metals in which the σ-bands cross the Fermi energy. The fact that AlB_2 which has a higher Fermi energy lying exclusively in the π-bands, does not superconduct at all shows how crucial this is. One point of interest is the case of $Li_{1+x}BC$. This material would be essentially isoelectronic with MgB_2. Theoretical calculations by Rosner et al. [6] within the framework of BCS theory gave a stronger electron-phonon coupling which in turn led to an estimate of T_c as high as 100 K if the Li content can be varied, much higher than the value for MgB_2. Unfortunately it is not proved possible to synthesize this material. However this estimate shows us that the limits on T_c within BCS theory are a lot higher than previously thought.

The case of MgB_2 illustrates the importance of even moderately strong coupling to high frequency phonons. Similar arguments have been frequently put forward for the cuprates and transition metal oxides based on the presence of high frequency O-optic modes in these metals. Curiously as we remarked earlier, superconductivity occurs only rarely in these oxides. This illustrates the subtlety of superconductivity. However the cause of the absence of superconductivity in many metals and oxide metals in particular, is a much less glamorous question than the origin of high T_c superconductivity in a narrow class of metals. It remains as an open challenge which however is generally ignored. In the specific case of cuprates, I just note that strongly overdoped cuprates are the most metallic and are furthest from the parent Mott insulator and so should have the weakest Coulomb interactions. But they do not superconduct. It follows that we should look elsewhere for the origin of the special superconductivity of the cuprates.

3. Cuprates, brief review of key experiments and phase diagram

High temperature superconductivity is obtained only by lightly hole (or also electron) doping CuO_2 planes with hole concentrations in the range 0.1-0.2 per Cu. The stoichiometric compounds in which all sites have a Cu^{2+} oxidation state are Mott insulators. The gap to charge excitations measured by optical experiments is of order 1.5 to 2 eV[7]. It is an order of magnitude larger than the energy scale of magnetic excitations. These are quite well described by a n. n. AF $s = 1/2$ planar Heisenberg model with a large exchange constant $J \sim 0.15$ eV[8]. Their quasi-2D character lowers the Neel temperature. The persistence of the insulating character for $T > T_N$ demonstrates that it is the strong onsite Coulomb interaction and not AF order, which is the origin of the insulating behavior, i. e. they are Mott insulators. This fact alone shows us that BCS theory does not apply since the Coulomb interaction does not scale away to weak coupling at low energies and cannot be treated perturbatively as in Landau theory. Rather the Coulomb repulsion dominates the low energy behavior and invalidates band theory.

Hole doping rapidly suppresses the AF order and leads to the by now familiar phase diagram. This can be divided into 3 regions according to the hole concentration, δ

a) Overdoped– $\delta \leq \delta_c$: the upper critical concentration, i. e. $T_c \rightarrow 0$, $\delta \rightarrow \delta_c$

In this region the evidence that normal state is a Landau Fermi liquid is compelling. Recently Proust and coworkers[9] studied the thermal conductivity, κ, and electrical conductivity, σ, of an overdoped single layer $Tl_2Ba_2CuO_{6+x}$ sample ($T_c = 14$ K and $B = 0$) in magnetic fields up to $B=15T$ sufficient to suppress all traces of superconductivity. They found perfect agreement with the Wiedemann-Franz law

$$\frac{\kappa}{\sigma T}\bigg|_{T \rightarrow 0} = \frac{\pi^2}{3} \frac{k_B^2}{e^2} \ , \tag{5}$$

in accordance with Landau theory. In this overdoped region weak coupling BCS theory generalized to describe d-wave pairing (see below) should apply.

b) Underdoped region, $\delta < \delta_{opt}$ (defined by the maximum in $T_c(\delta)$)

This region of the phase diagram shows marked deviations from the behavior of standard metals (e. g. see the full review by Timusk and Statt [10]). This deviation was first observed in NMR experiments. A clear difference was found in the Knight shift (which is linearly proportional to the spin susceptibility, χ) vs. T for 2 samples of 123. The sample with O_7 has a hole concentration δ-δ_{opt} while the $O_{6.6}$ sample is in the underdoped region. In the former sample χ is independent of T for $T > T_c$ as expected for the Pauli susceptibility of a metal. However in the underdoped sample $\chi(T)$ starts to fall already at 300 K and shows only a very weak anomaly at $T = T_c$ and $\chi(T_c) \sim \chi$ (300 K)/5. Such a large reduction cannot be ascribed to AF fluctuations (in an AF ordered state the angle averaged value of χ is reduced only by 2/3). It signals a continuous onset of singlet pairing of all Cu-spins starting at a high temperature $T^* \leq T_c$. One speaks of a spin gap setting in at T^*. ARPES experiments in this region show only 4 disconnected arcs centered around the diagonal directions in the Brillouin zone (i. e. near $\mathbf{k} = (\pm\pi/2, \pm\pi/2)$) and energy gaps appearing in a single particle spectrum near the points $\mathbf{k} = (\pm\pi, 0)$ and $(0, \pm\pi)$ [11]. This gap is generally referred to as a pseudogap. This could have its origin in strong AF fluctuations but the behavior of $\chi(T)$ suggests there is more to this behavior than simply strong AF fluctuations. Another intriguing aspect of the underdoped region is the reduction in the density of superconducting carriers, n_s. Early measurements of $1/\lambda^2 (T \rightarrow 0)$ (which varies as n_s) by Uemura and coworkers [12] showed that it scales linearly with T_c and the hole doping, δ. This implies that the materials are behaving as hole doped insulators, not simple metals which of course is quite reasonable on general grounds. However this behavior is a clear sign of strong correlations and the proximity to the Mott insulator. It is a marked departure from a weakly correlated Fermi liquid where n_s should be determined by an integral over the full Fermi surface and so by the electron (not hole) density.

c) Optimum doping and possible Quantum Critical Point

Since there are marked differences between the overdoped and underdoped regions of the phase diagram it is clearly relevant to ask if there is a simple evolution from one limit to the other. T_c for example seems to pass through a smooth maximum at what is called optimal doping. However the properties of the transition to superconductivity seem to change rather abruptly. The specific heat anomaly is a case in point – overdoped samples show a standard jump as in BCS theory while underdoped samples show a weak anomaly at T_c which is line with the interpretation of the spin gap as a consequence of spin pairing in the normal phase which in turn leads to an entropy reduction in the normal phase. Loram and coworkers[13] made extensive experimental studies on the YBCO materials and claim that there is a clear anomaly in thermodynamic quantities which reflects a zero temperature Quantum Critical Point in the superconducting dome at a hole concentration slightly above optimal. The exact form of such a QCP is not a priori clear since superconductivity exists on both sides of the putative QCP, and there is no obvious symmetry difference between the phases on either side.

A key question which was hotly debated for many years, is the symmetry of the superconducting state itself. This was settled a decade ago by a series of so-

called phase sensitive experiments. Almost all quantities one can measure in a superconductor depend only on the magnitude of the order parameter and do not give information on the phase, or more specifically, information on possibly angular variations around the Fermi surface of the condensate amplitude. The Knight shift measurements early on established singlet pairing for which the general form of the order parameter is

$$\Delta_{\sigma\sigma'}(\mathbf{k}) = \left\langle c^{+}_{\mathbf{k},\sigma}\, c^{+}_{-\mathbf{k},\sigma'} \right\rangle = i\, f(\mathbf{k})\, e^{i\varphi}\, \sigma_{y\,\sigma\sigma'}\,. \qquad (6)$$

(for details see the review by Sigrist and Ueda [14]). The BCS theory of pairing caused by the attraction caused by phonon exchange predicts singlet pairing with s-wave symmetry (or $f(\mathbf{k})$ approximately constant around the Fermi surface) because all scattering processes for Cooper pairs around the Fermi surface are attractive, irrespective of the momentum transfer. However as noted above the proximity to the Mott insulator immediately calls a simple s-state into question. The key idea to test for a nontrivial angular dependence of $f(\mathbf{k})$ is to form a suitable arrangement of Josephson junctions. In particular as first noted by Geshkenbein, Larkin and Barone [15] in the context of heavy fermion superconductors, the fact that the tunneling probability is greatest for electrons moving perpendicularly to a junction gives us the possibility of probing for sign changes in $f(\mathbf{k})$ around the Fermi surface. In particular if we consider the case that $f(\mathbf{k})=\Delta_0(\cos(k_x)-\cos(k_y))$ – an example of d-wave symmetry in a tetragonal crystal, then the overlap of pairs in a Josephson junction between samples with the x- and y-axis perpendicular to the junction on the right and left sides, acquire a minus sign (see Sigrist and Rice[16]). In this case we can write the Josephson free energy as

$$F_J = -F_0 \cos\left(\varphi_R - \varphi_L + \pi\right), \qquad (7)$$

where $\varphi_R (\varphi_L)$ are Josephson phase on the right (left). Of course in a single isolated junction this addition of π plays no role since it can be absorbed into the Josephson phases. However, if one can arrange the right and left superconductors to join in a ring, then the additional π cannot be gauged away. Further if the Josephson coupling is strong enough it will impose a phase change of π at the junction which leads to a phase change of π and therefore a current flow around the ring. Since the phase change around this ring containing a π-junction is only half that for a standard ring, where the phase change must be a multiple of 2π, the current leads to a magnetic flux exactly enclosed by the ring one half that of the standard superconducting flux quantum of $hc/2e$. We won't review these well known experiments here but refer the reader to the excellent review by Kirtley and Tsuei [17]. One point worth mentioning though is the very recent and elegant work of Hilgekamp and coworkers [18] who fabricated large arrays of these π-junction rings and measured the spontaneous flux patterns that result when the arrays are cooled to low temperatures.

4. Cuprates: microscopic origin of superconductivity within the resonant valence bond framework

As discussed above there are many reasons to abandon BCS theory for the cuprate superconductors. Two of the most relevant are first the $d_{x^2-y^2}$-pairing symmetry rather than the s-wave pairing that follows from BCS theory. Secondly the fact that these superconductors behave as doped Mott insulators in the underdoped region of the phase diagram rather than normal metals with a large Fermi surface shows that the onsite Coulomb interaction between electrons is highly relevant on the low energy scale and has not scaled to weak coupling as is the case in BCS theory. Thus one needs a new microscopic theory to describe the origin of superconductivity. In addition there are many other anomalous properties of these fascinating materials which require explanation. There have been many attempts and literally thousands of papers devoted to this topic in the twenty years since the discovery of high temperature superconductivity in the cuprates. We cannot hope to cover this work in this brief review and instead will restrict this account to the most promising microscopic theory, namely the Resonant Valence Bond (RVB) theory due to Anderson[19].

All the cuprate superconductors have as key components layers composed of CuO_2-planes with a square lattice of Cu-ions. An O-ion lies between each n. n. Cu-Cu pair. The average concentration of electrons in a hole doped layer corresponds to a formal valence in the layer of $Cu^{(2+\delta)+} O_2^{4-}$, where δ is the hole concentration. The simplest representation of this electronic configuration is as $(1-\delta)$ Cu^{2+}-sites and a concentration δ of Cu^{3+}-sites. The former have a single hole in the 3d-shell and so have a net spin of $S=1/2$. The Cu^{3+}-sites have 2 holes in the 3d-shell but as shown by Zhang and Rice [20] both occupy the highest antibonding level of the CuO_4-square and so have $S=0$. The strong onsite correlation required by the doped Mott insulating character implies that the only charge degrees of freedom come from the motion of these Cu^{3+} configurations in a background of Cu^{2+}-sites. The latter are coupled by a n. n. Heisenberg interaction with an exchange constant, J. The former move rapidly with n. n. and further hoppings allowed through the transfer of an electron with matrix element, t. The resulting Hamiltonian leads to a model on a square lattice generally called a t-J model

$$H_{t-J} = -t \sum_{\langle i,j \rangle,\sigma} \left(1-n_{i,-\sigma}\right) c^{+}_{i\sigma} c_{j\sigma} \left(1-n_{j,-\sigma}\right) +$$
$$H.C. + J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j\,. \qquad (8)$$

This is a difficult model because of the inherent competition between the rapid motion of the holes ($t/J \sim$ 3 in the cuprates) which rearranges the spin configurations and the spin exchange term which favors an antiferromagnetic ordering of the spins. Indeed

twenty years after Anderson [19] proposed it as the essential low energy model that describes the hole doped CuO_2 planes, we still lack a comprehensive analysis of this model.

In his famous 1987 paper Anderson [19] proposed the RVB ansatz as the key. This idea goes back to his earlier work that started from the exceptional stability of a singlet pair of spins which has an energy of $-3J/4$ compared to an antiferromagnetically arranged pair of spins which has an energy of only $-J/4$. This effect implies that the quantum fluctuations about an ordered Heisenberg antiferromagnet for $S = 1/2$ spins are is very strong, so strong that a spin liquid state could be envisaged in which the long range antiferromagnetic order is destroyed, leaving only a short range correlation, i. e. $\langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle \to 0$ as $|i - j| \to \infty$. Such a spin liquid state can be represented as a superposition of many spatial configurations of singlet spin pairs. The effect of doping is to make these electron pairs mobile and so to give rise to a superconducting state. This elegant idea explains the special nature of the superconductivity in these CuO_2 planes. There is however an important difficulty, namely the stoichiometric undoped cuprates have long range antiferromagnetic order and careful analysis showed that even for a single plane with a square lattice of Heisenberg $S = 1/2$ spins, the quantum fluctuations although substantial, are not strong enough to destroy the order [21]. However this is not a fatal difficulty since it by no means rules out an RVB superconductor in the presence of hole doping since the rapid motion of the holes rapidly destroys the magnetic order thus preparing the way for the doped RVB spin liquid beyond a modest hole concentration.

The relationship between the RVB ansatz and superconductivity was established already by Anderson [19] who wrote the RVB wavefunction as a BCS pair superconducting state projected down to the strong correlation Hilbert subspace containing no doubly occupied sites. The relevant projection operator is known as a Gutzwiller projection operator. This operator complicates the analysis but there are two useful methods to handle it. The simplest method known as the Gutzwiller approximation is to replace the operator by a numerical factor which is calculated from the statistical weight of the relevant configurations in the Hilbert subspace, assuming complete statistical independence on each site [22]. Alternatively the projected wavefunction can be treated quantitatively using a variational Monte Carlo algorithm. The results of the two methods agree quite well. More importantly the key features that follow from the RVB ansatz agree well with key experimental properties of the groundstate. This good agreement has been stressed by Anderson and collaborators [23] in a recent review. They emphasize that many of these features were obtained before the relevant experiments were carried out and in that sense represent predictions of the RVB ansatz. These will be summarized below.

First among these is the symmetry of the superconducting states. It was soon shown by Gros[24] and others (see Anderson et al. [23]) that the energy was minimized by d-wave rather s-wave pairing i.e. $f(\mathbf{k}) = \Delta_0 \left(\cos k_x - \cos k_y \right)$ in eq. 6. As discussed above this symmetry was confirmed later by phase sensitive experiments [17].

A second result is that the gap magnitude in the so-called antinodal directions $\left(k_x \approx \pm\pi, k_y \approx 0 \right)$ and $\left(k_x \approx 0, k_y \approx \pm\pi \right)$, $2 \Delta_0(\delta)$, is a strong function of the density, decreasing from a maximum at $x = 0$ to zero at a critical hole density $\delta_c (\approx 0.25)$. However in the underdoped region with small values of δ the true superconducting order parameter vanishes linear in δ since $|\Delta_0|^2$ involves moving a pair of electrons over arbitrarily large distances. The Drude weight which controls the strength of the δ-function in the conductivity at zero frequency, also vanishes linearly in δ since only the holes carry a current. As a result the superconducting transition temperature, T_c must also vanish as $\delta \to 0$. Wen and Lee[25] have given an explicit derivation of this result with $T_c(\delta)$ vanishing linearly with δ. The contrasting behavior of the gap magnitude $\Delta_0(\delta)$ and the transition temperature causes the normal state at $T > T_c$ to be highly anomalous in the underdoped region. The Fermi surface will be mostly truncated and there will be strong spin pairing in singlets. As a result one gets a much reduced Pauli spin susceptibility and Knight shift. This feature of the RVB theory agrees well with the anomalous properties of the pseudogap or spin gap phase [10].

Recently Paramekanti and coworkers [26] extended the variational Monte Carlo calculations using the Gutzwiller wavefunction to obtain the properties of the quasiparticles with wavevectors along the nodal directions, i. e. along the Brillouin zone diagonals, $k_x = \pm k_y$. Along these directions the d-wave gap function vanishes so that there are simple quasiparticles even in the superconducting state. They found two interesting results that agree well with ARPES (Angle Resolved Photo Emission Spectroscopy) measurements [11]. First the coherent quasiparticle weight is a strong function of the hole density, δ and also vanishes linearly with, δ. Secondly, the Fermi velocity of the quasiparticle is essentially independent of the hole density δ, which also agrees with ARPES measurements. It is interesting to note that the combination of a Fermi velocity which remains constant and a vanishing Drude weight as $\delta \to 0$ are not easy to reconcile within a Landau Fermi liquid description. The simplest resolution of the conflict is to partially truncate the Fermi surface away from the diagonal directions. We will return to this point later.

The conclusion of this analysis is that the RVB ansatz gets the key properties at least of the ground state correctly, which is itself a considerable achievement. However these approaches are not easily generalized to finite temperature.

5. Gauge theory formulation of RVB

A great effort has been made to develop an analytic formulation of the RVB concept. The approach tries to formulate an appropriate mean field approximation which could be improved by adding fluctuations, hopefully in a controlled way, so that the key properties are not drastically altered. Early on it was suggested a suitable approach is to introduce a factorization of the electron operators into charge and spin parts (Kotliar and Liu [27], Suzumura et al. [28]). We know that in one dimension a Luttinger liquid forms. If an electron is removed at the Fermi level, it does not form a Landau quasiparticle as in three dimensions but decays into a superposition of a holon, which carries the charge, and a spinon, which carries the spin. These interact only weakly and one calls this phenomenon spin-charge separation. Here however, one is introducing a radical ansatz which needs to be justified. The factorization in its simplest form is written as

$$c_{i\sigma}^+ = f_{i,\sigma}^+ \, b_i , \qquad (9)$$

where the fermion, $f_{i\sigma}^+$ is called a spinon and the charged boson b_i is a holon. The t-J model is now rewritten as

$$H_{tJ} = -t \sum_{\langle ij\rangle,\sigma} f_{i\sigma}^+ \, b_i \, b_j \, f_{j\sigma} + \text{h.c.} +$$

$$J \sum_{\langle i,j\rangle} \left(b_i \, b_i^+ \, b_j \, b_j^+ \right) \mathbf{S}_i \cdot \mathbf{S}_j - \mu \sum_{i,\sigma} f_{i\sigma}^+ \, b_i \, b_i^+ \, f_{i\sigma} +$$

$$\sum_i \lambda_i \left(\sum_\sigma f_{i\sigma}^+ f_{i\sigma} + b_i^+ b_i - 1 \right) , \qquad (10)$$

where the spin operator $\mathbf{S}_i = f_{i\sigma}^+ \, \sigma_{\sigma\sigma'} \, f_{i\sigma}$, and λ_i is an independent Lagrange multiplier to enforce the strong coupling constraint, i. e. that on each site there must be either a spinon or a holon. This factorization is invariant under a simultaneous local gauge transformation of the holon and spinon i. e. $f_{i\sigma} \to f_{i\sigma} \, e^{i\varphi_i}$ and $b_i \to b_i \, e^{i\varphi_i}$, i. e. a U(1) gauge transformation, so we have a gauge theory. The simplest approach is to make a suitable mean field approximation with the hope that the fluctuation corrections are small. Mean field factorizations based on a series of expectation values, $\chi_{ij} \left(= \sum_\sigma \left\langle f_{i\sigma}^+ \, f_{j\sigma} \right\rangle \right)$

$\Delta_{ij} \left(= \left\langle f_{j\uparrow} \, f_{i\downarrow} - f_{j\downarrow} \, f_{i\uparrow} \right\rangle \right)$ and $B_i = \left\langle b_i^+ \right\rangle$, are introduced. In addition the Lagrange multiplier on each site λ_i, is set to a single value, λ. This relaxes the local constraint and replaces it with a global constraint. The resulting mean field phase diagram as a function of temperature, T, and hole doping, δ, looks encouraging. It contains regions with d-wave paired spinons and Bose condensed holons and a true superconducting phase when both are nonzero. The phase with $\Delta_{ij} \neq 0$ but $B = 0$ can be interpreted as a spin gap phase without superconductivity and the phase with $\Delta_{ij} = 0$ but $B_i \neq 0$ is a Landau Fermi liquid. Further the phase with both Δ_{ij}

and B_i finite is a d-wave superconductor. In addition there is a region with $\chi \neq 0$ but Δ and B both zero which is not a standard metallic phase. Thus overall there is encouraging qualitative agreement with the experimental phase diagram. However none of the expectation values Δ, χ and B is invariant under local gauge transformations so they cannot be true physical order parameters. Further the lines where they become finite should not represent physical phase transitions. Ubbens and Lee [29] examined the corrections that follow when gauge fluctuations are included. Their phase diagram has only a superconducting dome as a broken symmetry phase and the unphysical phase transition from the mean field approximation have disappeared.

However it is not clear how good this approximation scheme is at higher temperatures. The basic problem lies in the inherent defect that follows from introducing fermions, bosons and gauge fields. One has described the strong correlation constraints which reduce the physical degrees of freedom, by expanding the degrees of freedom. This conflict was emphasized by Hlubina and coworkers [30] who compared the entropy as a function of temperature calculated in these gauge theories to numerical calculations of Putikka. Reasonable agreement could only be reached if the longitudinal fluctuations of the gauge field were included (usually they are omitted) which in turn lead to strong interactions between spinons and holons. However if the spinons and holons are strongly interacting a description in terms of weakly interacting spinons and holons fails and the whole basis to argue for spin-charge separation as a good starting point is questionable.

Recently a more sophisticated and improved version of the gauge theory has been introduced by Lee, Wen and coworkers [31]. Their SU(2) gauge theory seeks to incorporate fluctuations between the d-wave paired state and other states connected by the SU(2) gauge transformations which apply at half-filling, in particular fluctuations into staggered flux (or d-density wave) states. Since the free energy difference between the phases vanishes in the stoichiometric limit, it should be small at finite but small doping. This is very reasonable on physical grounds. To this end they introduce two independent bosonic holons. In this improved gauge theory only vortices with the standard superconducting flux quantum of $hc/2e$ appear, thereby removing the problems that flux quanta of hc/e can appear in the U(1) gauge theory because of the single holon Bose condensate. A nice feature of the theory is that vortices are cheap in the words of Patrick Lee, i. e. they cost little core energy because a rotation to a nonsuperconducting core with a staggered flux is possible rather than a complete suppression of the RVB phase (see Honerkamp and Lee [32]). Thus the region above T_c will have strong vortex fluctuations which can be the origin of the large Nernst effect reported by Ong and coworkers [33]). The introduction of an extra bosonic particle does nothing to

cure the problem of excess entropy mentioned above. So this description should also be limited to relatively low temperatures, i. e. $T \geq T_c$ but $T \leq T^*$.

A comprehensive review of the gauge theories has recently been published by Lee, Nagaosa and Wen [34].

6. Conclusions

This brief review has concentrated on the RVB ansatz which is the most promising of all the theories that have been advanced to explain the spectacular phenomena exhibited by the high-T_c cuprate superconductors. While the basic outline of this ansatz is elegant and simple to understand, the derivation of a comprehensive theory has proved extraordinarily difficult. There are two good reasons for this. One is the lack of a small parameter which would enable us to make a controlled expansion around a well understood limit. Although the hole density is small, the introduction of fermionic holes into the Mott insulating state of the stoichiometric cuprates is an essential perturbation which immediately changes the character of the low energy states and therefore of the many body problem from a purely spin problem to a fermionic one. The second reason is the absence of symmetry breaking and the presence of only short range correlations in the basic RVB phase. This does not allow us to use a mean field approach based on long range order. Yet these short range correlations control the character of the low lying excitations and lead to essential modifications.

One model system which shows similar behavior is the two leg Hubbard ladder. At half-filling this model has a unique groundstate which has finite energy gaps to excitations in the charge and spin sectors (for a review see Dagotto and Rice [35]). The pairing (with approximate d-wave character) and antiferromagnetic correlation functions are enhanced but are strictly short range. This model has the advantage that it can be comprehensively analyzed and shows a continuous crossover between weak and strong coupling as the onsite Coulomb repulsion is varied. Particularly the weak coupling limit can be comprehensively analyzed using a combination of renormalization group (RG) theory and bosonization to obtain the properties of the strong coupling low energy sector (e. g. see Lin, Balents and Fisher [36]). The RG approach can be extended to a two dimensional square lattice. Interestingly although the RG flow equations are quite different, there are strong parallels in the flow of the correlation functions between the two models (Honerkamp et al. [37]). Indeed a recent numerical analysis by Läuchli and coworkers [38] of the strong coupling phase at low energies in the two dimensional model confirmed the similarity and its further evidence in favor of the RVB ansatz.

Very recently Yang and coworkers [39] introduced a phenomenological ansatz for the single particle Greens function within the RVB theory. Their approach was based on an analogy with the form that Konik et al. [40] derived for a particular form of interladder hopping between an array of two leg Hubbard ladders. The ansatz of Yang et al. [39] gives a good fit to the ARPES results on underdoped cuprates and resolves a number of open questions about the relation of the data to the Luttinger sum rule.

There are still many open challenges in the field and we are far from having a comprehensive and complete theory for the high T_c cuprate superconductors as we have for the conventional superconductors. These cuprate materials of course show many anomalous properties also in the phase at temperatures above the superconducting transition so that the complete description requires us to understand these properties too and their relationship to their exceptional superconductivity.

References

1. J Bardeen, L N Cooper and J R Schrieffer, *Phys. Rev.* **108** (1957) 1175.
2. J G Bednorz and K A Müller, Z Physik B **64** (1986) 189.
3. A Schilling et al., *Nature*. **363** (1993) 56.
4. J Nagamatsu et al., *Nature*. **410** (2001) 63.
5. Y. Kong et al., *Phys. Rev.* B **69** (2001) 020501.
6. H Rosner et al., *Phys. Rev. Lett.* **68** (2002) 127001.
7. S Uchida et al., *Phys. Rev.* B **43** (1991) 7942.
8. R Coldea et al., *Phys. Rev. Lett.* **86** (2001) 5377.
9. C Proust et al., *Phys. Rev. Lett.* **89** (2002) 147003.
10. T Timusk and B Statt, *Rep. Prog. Phys.* **62** (1999) 61.
11. M R Norman et al., *Nature*. **392** (1998) 157.
12. Y J Uemura et al., *Phys. Rev. Lett.* **62** (1989) 2317.
13. J W Loram et al., *Phys. Rev. Lett.* **71** (1993) 7140 and *Physica* C. **282-7** (1997) 1405.
14. M Sigrist and K Ueda, *Rev. Mod. Phys.* **63** (1991) 239.
15. V B Geshkenbein, A I Larkin and A Barone, *Phys. Rev.* B. **36** (1987) 323.
16. M Sigrist and T M Rice, *J. Phys. Soc. Japan* **61** (1992) 4283.
17. J R Kirtley and C C Tsuei, Rev. Mod. Phys. **72** (2000) 969.
18. H Hilgenkamp et al., *Nature*. **422** (2003) 50.
19. P W Anderson, *Science*. **235** (1987) 1196.
20. F C Zhang and T M Rice, *Phys. Rev.* B **37** (1988) 3759.
21. E Manousakis, *Rev. Mod. Phys.* **83** (1991) 1.
22. F C Zhang et al., *Supercond. Sci. Technol.* **1** (1988) 36.
23. P W Anderson et al., *J. Phys. Condens. Matter* **16** (2004) R755.
24. C Gros, *Phys. Rev.* B **38** (1988) 931; *Ann. Phys.* **189** (1989) 153.
25. X G Wen and P A Lee, *Phys. Rev. Lett.* **76** (1996) 503.
26. A Paramekanti, M Randeria and N Trivedi, *Phys. Rev. Lett.* **87** (2001) 217002; *Phys. Rev.* B **70** (2004) 054504.
27. G Kotliar and J Liu, *Phys. Rev.* B **38** (1988) 5142.
28. Y Suzumura, Y Hasegawa and H Fukuyama, *J. Phys. Soc. Japan* **57** (1988) 2768.

29. M U Ubbens and P A Lee, *Phys. Rev.* B **49** (1994) 6853.

30. R Hlubina et al., *Phys. Rev.* B **46** (1992) 11224 .

31. P A Lee and X G Wen, *Phys. Rev.* B **63** (1997) 224517; see also J. Kishine et al., *Phys. Rev. Lett.* **86** (2001) 5365.

32. C Honerkamp and P A Lee, *Phys. Rev. Lett.* **90** (2003) 246402 and **92** (2004) 177002.

33. Y Wang et al., *Phys. Rev. Lett.* **88** (2002) 257003.

34. P A Lee, N Nagaosa and X G Wen, *Rev. Mod. Phys.* **78** (2006) 17.

35. E Dagotto and T M Rice, *Science* **271** (1996) 618 .

36. H H Lin, L Balents and M P A Fisher, *Phys. Rev.* B **58** (1998) 1794.

37. C Honerkamp, M Salmhofer and T M Rice, *Eur. Phys.* J. B **27** (2002) 127.

38. A Läuchli, C Honerkamp and T M Rice, *Phys. Rev. Lett.* **92** (2004) 037006.

39. K-Y Yang, T M Rice and F C Zhang, *Phys. Rev.* B **73** (2006) 174501.

40. R M Konik, T M Rice and A Tsvelik, *Phys. Rev. Lett.* **96** (2006) 086407.