Consistent theory of underdoped cuprates: Evolution of the resonating valence bond state from half filling

Sanjoy K. Sarker and Timothy Lovorn
Department of Physics and Astronomy, The University of Alabama, Tuscaloosa, Alabama 35487, USA
(Received 19 October 2009; revised manuscript received 31 May 2010; published 6 July 2010)

Using continuity, we derive a renormalized Hamiltonian from the parent *t*-*J* model to describe the properties of underdoped cuprates. The theory is constrained to agree with the behavior at half filling, which is well described by the Arovas-Auerbach valence bond state in which bosonic spinons are paired into singlets. Spinon states evolve continuously into the doped region preserving their symmetry. We assume that moving holes rapidly destroy magnetic order, which leads to a gap in the spinon spectrum and strongly renormalizes the theory. The spin gap leads to two different types hopping terms for renormalized holes. In one, a fermionic holon hops within the same sublattice accompanied by a singlet backflow, giving rise to a non-Fermi-liquid normal state with novel properties. Spinon singlets condense below a pseudogap temperature T^* ($<$ spin gap temperature T^0), which allows holons to propagate coherently, forming a spinless Fermi liquid, but without an observable holon Fermi surface. Above T^*, holons are localized. This is the so-called strange metal phase, which is actually a new type of insulator since its resistivity would be infinite at $T=0$. In the second term a pair of holons belonging to opposite sublattices hop, accompanied by a singlet backflow. In the presence of the singlet condensate holon pairs condense, leading to *d*-wave superconductivity; the symmetry is primarily determined by the symmetry of the valence bond state at half filling. The metal and the superconductor preserve the two-sublattice character of the valence bond state. A careful examination of the nuclear magnetic resonance, tunneling, and transport data shows that the predictions of the theory is consistent with experimental results. Remarkably, the existence of the spin gap provides a natural explanation for the phenomenon of two dimensionality of the normal state in the presence of interplane hopping. The marked asymmetry between hole-doped and electron-doped cuprates is also easily explained.

DOI: 10.1103/PhysRevB.82.014504 PACS number(s): 74.20.Mn, 74.72.−h, 71.27.+a

I. INTRODUCTION

The origin of high-temperature superconductivity in doped cuprates[1] is believed to be closely linked with the unusual behavior of the normal state, which is not a Fermi liquid. Furthermore, in the underdoped region a pseudogap appears below a temperature T^*, much above the superconducting temperature T_c.[2] Even before the discovery of the pseudogap phase, Anderson argued that the non-Fermi-liquid behavior is due to the two-dimensional (2D) nature of the normal state,[3] and its proximity to the undoped phase, which is a Mott insulator, or equivalently, a quantum antiferromagnet. The 2D behavior is unexpected and its origin is not understood. Anderson also proposed the resonating valence bond (RVB) state in which spins are paired into singlets within the insulator.[3] Upon doping with holes, this state would evolve continuously into the doped region where spinless holons, which carry charge, would propagate coherently and create a metal. The elementary excitations are holons and spin-1/2 spinons. The physical electron is a composite particle. The expectation is that the apparent spin-charge separation would lead to non-Fermi-liquid behavior, and further that the spin singlets would acquire charge via some interaction with holons and become superconducting.

Theoretical studies are usually based on the large-U Hubbard model or, equivalently, the *t*-*J* model on a square lattice, where $J=4t^2/U$ is the exchange coupling between spins, and t describes nearest-neighbor (nn) hopping of electrons such that no site is doubly occupied. For cuprates, $t/J\sim 3-4$. The constraint of no double occupancy corresponds to a $U(1)$ gauge symmetry.[4] Much of the work has been devoted to deriving an effective low-energy theory of propagating spinons and holons (or pairs), coupled to a gauge field. In principle, such a task should not be too difficult since one can invoke continuity. The effective action would depend on the symmetries of the underlying vacuum state, i.e., a renormalized version of the RVB state of paired spinons, which is presumed to describe the insulator at half filling. After more than two decades of intensive work, a successful theory has not been found.

This failure is surprising since, using a mean-field (MF) approximation, Arovas and Auerbach[5] have shown that the behavior at half filling is indeed reasonably well described by an RVB state in which singlets formed by pairing bosonic spinons (Schwinger bosons) condense below a temperature T_{RVB}. (In this paper, we use the term RVB to describe any valence bond state.) One complication is that a fraction of spinons remains unpaired and condenses separately to give rise to long-range antiferromagnetic order. This is not a defect of the theory since it correctly describes the ground state, which is known to be ordered.[6] In other words, the system is in a mixed phase of an RVB state and a Néel state with $T_{\mathrm{RVB}}>T_{\mathrm{AF}}$ (in $d=2$, $T_{\mathrm{AF}}=0$). Essential correctness of the bosonic theory was later confirmed by analytical[6] and numerical work.[7] In cuprates the AF insulating state exists up to a hole density x of about 0.05, beyond which there is a transition to a pseudogap metal (and a *d*-wave superconductor), with no long-range magnetic order of any type. However, a seemingly straightforward extension of the Arovas-Auerbach MF theory (with fermionic holons) to the

doped region leads to a metallic state with spiral magnetic order.[8] This failure has lead to a virtual abandonment of the approach based on the Arovas-Auerbach RVB state and bosonic spinons.[9]

Instead, RVB theories based on the slave-boson representation (spinons are fermions and holons are bosons) have been used widely[10–12] (for a review, see Lee *et al.*[13]). But these do not work at half filling and there are no experimental signatures of a transition to a different RVB state accompanied by statistical transmutation upon doping. There are serious doubts as to their efficacy in describing the physics of underdoped cuprates. Indeed the failure to connect with the physics at half filling is one of the central problems in high-T_c theory.

We have solved this problem by deriving a renormalized theory for small x which is consistent with the physics at half filling but does not have the problems of spiral instability. The key point is that, for the t-J model, the spiral state has been shown to be unstable.[14] We therefore assume that (1) the spin states evolve continuously from half filling, preserving their symmetry and particle statistics, and (2) moving holes rapidly destroy long-range magnetic order at $T=0$ beyond some small critical concentration x_c, strongly renormalizing the theory in the process; but the VB state survives up to a temperature $T_{\mathrm{RVB}}(x)$. These are reasonable assumptions, and formed the basis of early gauge theories based on the Schwinger-boson representation.[15,16] But in those theories it was *assumed* that the initial renormalization leads to a sublattice-preserving t'-J model, where t' is an effective next-nearest-neighbor hopping parameter.

The actual renormalized Hamiltonian is found to be quite different. A brief account of the theory, focusing primarily on the pseudogap phase, and some of the results have been reported earlier.[17] Here we provide a detailed description, and show that it not only accounts for the main features of the phase diagram, but predicts novel properties of the underlying states that require a major revision of our current understanding of the cuprate phenomenology. Additionally, the theory provides a natural explanation for the two dimensionality of the normal state, as well as for the marked asymmetry between hole-doped and electron-doped cuprates.

The continuity requirement ensures that there are exactly the same three spinon states in the doped region as at half filling, and holon motion do not introduce any new order parameters *for the spinons*. Destruction of magnetic order automatically leads to a gap for bosonic spinons in the other two states: the RVB state and the nonordered state.[5,18] The spin gap, not to be confused with the pseudogap, is the distinctive feature in the theory. As described in Sec. IV, the gap allows us to renormalize away nearest-neighbor hopping (t), and use perturbation theory to obtain a minimal Hamiltonian involving sublattice-preserving hopping of renormalized holes, accompanied by singlet backflows. Two types of new hopping terms are generated: in one, a single holon hops within the same sublattice, and in the other, a pair of holons hop together.

We analyze the Hamiltonian in Sec. V, showing that the one-hole term gives rise to exactly two "normal" states, as found in cuprates, whose symmetries are determined by the underlying spinon states (at half filling), no new order parameter is introduced, either in the spin or the charge sector. In the absence of singlet condensation, holons are localized. We identify this state with the so-called strange metal phase since it is stable above $T_{\mathrm{RVB}}(x)$. However, it is not a metal, but a new kind of insulator since its resistivity would become infinite when continued to $T=0$. This is an important prediction. Coupling to the RVB condensate allows holons to hop coherently within the same sublattice, and create a spinless Fermi liquid of concentration x below $T_{\mathrm{RVB}}(x)$, which we identify with the pseudogap temperature T^*. However, gauge invariance ensures that there is no observable (small) Fermi surface, which is another prediction.

As shown in Sec. VI, in the presence of a RVB condensate the pair-hopping term allows holon pairs to Bose condense, giving rise to d-wave superconductivity. The symmetry is largely determined by that of the underlying RVB state known from half filling. As discussed in Sec. II, the valence bond state has a two-sublattice property of its own: on average, singlets connect spins on opposite sublattices. This property is predicted to be preserved in the pseudogap metal and superconducting phases.

The absence of any new order (other than superconductivity) imposes constraints on the number and type of phases; the resulting phase diagram is in qualitative agreement with the experimental one. The constraints allow us to extract many universal features from the structure of the Hamiltonians in various phases without detailed calculations or fine tuning. For example, the spin gap by itself rules out a conventional Fermi-liquid state for small x. In Sec. VII we carefully examine the existing experimental results and find strong support for the predictions of our theory. In particular, results from nuclear magnetic resonance (NMR), transport, optical conductivity, and tunneling measurements, taken together, provide strong evidence for a spin gap, a nonmetallic phase above T^*, and a spinless Fermi liquid without a Fermi surface below. We will point out that these results cannot be reconciled with either a Fermi liquid or earlier RVB theories.

In Sec. VIII we discuss the fundamental issue of the two dimensionality of the normal state, which is unexpected since the cuprates are three dimensional (3D), though layered materials, and localization within a plane would cost too much kinetic energy. Indeed, all the other nearby states: the antiferromagnet in the undoped phase, the superconducting state, and the Fermi-liquid state that appear at large doping exhibit 3D behavior. Hence, theories that are based on the 2D model are suspect if they fail to exhibit 2D confinement when hopping in the perpendicular direction is turned on. Indeed, such a test will probably invalidate most of the existing theories. We show that, in our case, the spin gap provides a protective mechanism against delocalization of holons in the z direction, which is not obvious since holons are spinless. However, the holon pairs can hop coherently, which is essential for the observed 3D superconductivity. We solve the 3D problem and show that condensation energy is enhanced due to 3D coupling.

Another issue that is not completely understood is the significant difference between the hole-doped and electron-doped cuprates. In Sec. IX we will argue that the difference is due to direct intrasublattice hopping t', which breaks electron-hole (e-h) symmetry in the Hubbard model. Nor-

mally the difference is not significant, we will present results to show that in our case it is substantially enhanced because the nearest-neighbor hopping (t) is renormalized away. The issue is subtle and depends on the symmetry of the underlying spinon states.

Our renormalized Hamiltonian belongs to the class of short-range RVB models considered by Kivelson *et al.*[19,20] However, because of the consistency requirements the Hamiltonian differs significantly from those proposed earlier. The present theory is based on a paper published much earlier, in which a self-consistent perturbation expansion in powers of hopping was used to study the destruction of long-range magnetic order and the occurrence of non-Fermi-liquid behavior.[21] The intraplane pair hopping mechanism for superconductivity in the present theory was also discovered there. An effective Hamiltonian for renormalized holons was derived in Ref. 22 but with an (assumed) wrong (spiral) normal state. In Sec. III we review these and earlier works by other authors that are relevant to the present derivation. We draw attention to the early single-hole results which show clear evidence for strong renormalization and effective sublattice-preserving hole hopping. Our conclusions are summarized in Sec. X.

II. MODEL AND SYMMETRIES

The physics of no double occupancy is taken into account by representing the electron as a composite object created by $c_{i\sigma}^{\dagger}=b_{i\sigma}^{\dagger}h_i$, where $b_{i\sigma}^{\dagger}$ creates a spinon of spin σ and h_i destroys a holon at lattice site i, subject to the constraints that number of holons plus spinon at each site is one. Spin operators are represented as $S_i^{+}=b_{i\uparrow}^{\dagger}b_{i\downarrow}$, and $S_i^{z}=\frac{1}{2}(b_{i\uparrow}^{\dagger}b_{i\uparrow}-b_{i\downarrow}^{\dagger}b_{i\downarrow})$. Then the t-J Hamiltonian is given by

$$H=-t\sum_{ij,\sigma}c_{i\sigma}^{\dagger}c_{j\sigma}-2J\sum_{ij}\mathrm{A}_{ij}^{\dagger}\mathrm{A}_{ij}, \tag{1}$$

where the sum is over nearest neighbors, $\mathrm{A}_{ij}=\frac{1}{2}[b_{i\uparrow}b_{j\downarrow}-b_{i\downarrow}b_{j\uparrow}]$ destroys a singlet connecting spinons on nearest-neighbor sites. The second term describes exchange interaction written in terms of A_{ij}'s, which makes the role of the singlets explicit. For the most part we consider the square lattice. However, later we will need to include next-nearest-neighbor hopping within the plane and as well as hopping between planes.

The Hamiltonian is invariant under a local $U(1)$ gauge transformation,[4]

$$b_{i\sigma}\rightarrow b_{i\sigma}e^{i\theta_i},\quad h_{i\sigma}\rightarrow h_{i\sigma}e^{i\theta_i}. \tag{2}$$

The number of spinons plus holons,

$$\mathcal{N}_i=b_{i\uparrow}^{\dagger}b_{i\uparrow}+b_{i\downarrow}^{\dagger}b_{i\downarrow}+h_i^{\dagger}h_i$$

is then conserved at each site. The physical model lives in the projected subspace defined by $\mathcal{N}_i=1$, for each i. We can choose any statistics for spinons and holons as long as the statistics of the gauge-invariant quantities are correctly given. However, in an effective theory, in which spinons (and holons) are the low-energy excitations, we expect the system to choose a particular statistics.

Half-filled case. At half filling the hopping term can be dropped. The Arovas-Auerbach RVB state is described by a saddle point obtained by minimizing the effective action, which is equivalent to solving a mean-field problem.[18] In the MF approximation constraints are treated on average and singlets condense so that the valence bond order parameter $A_{ij}=\langle \mathrm{A}_{ij}\rangle$ becomes nonzero. For dimensionality $d\geq 2$, the MF theory works well at low T if spinons are bosons. For the symmetry of the order parameter we can choose,

$$A_{ij}=Ae^{i(1/2)\mathbf{Q}\cdot(\mathbf{r_i}-\mathbf{r_j})}, \tag{3}$$

where $\mathbf{Q}$ is the zone corner wave vector. In $d=2$, $\mathbf{Q}=(\pi,\pi)$. However, for $d\geq 2$ it has been shown that the ground state is not a pure RVB state since a fraction of spinons remains unpaired.[6,8] Since spinons are bosons they condense independently so that $\langle b_{i\sigma}\rangle\neq 0$. This leads to a nonzero value for $\langle S_i^{+}\rangle=\langle b_{i\uparrow}^{\dagger}b_{i\downarrow}\rangle$, giving rise to a two-sublattice antiferromagnetic order (in the x direction).

The ground state is therefore a mixture two competing phases, characterized by distinct order parameters, in which magnetization is reduced from its classical value by the presence of the singlet condensate. The MF theory describes the ground-state properties in $d=2$ rather accurately, and in $d=3$, reproduces the results of Holstein-Primakoff quantum spin-wave theory.[6] This success is remarkable, but not accidental, since the MF equations are identical to the equations describing the spin-wave theory.[5] In fact, the latter can be derived from the Schwinger-boson theory by integrating out the constraints and expanding the action in powers of $1/S$. These results are confirmed by accurate numerical work.[7] At finite T, AF order disappears above T_{AF} (in $d=2$, $T_{\mathrm{AF}}=0$), but the RVB state survives up to $T_{\mathrm{RVB}}>T_{\mathrm{AF}}$.

Symmetry of the RVB state. In the MF approximation, condensation of singlets (and spinons) breaks gauge symmetry. Similarly, coherent motion of holons lead to $\langle h_i^{\dagger}h_j\rangle\neq 0$ for $i\neq j$, which also breaks gauge symmetry. However, such local symmetries cannot be spontaneously broken[23] since there is no rigidity with respect to phases of the order parameter A_{ij} generated by the transformation, Eq. (2). Gauge invariance is restored by averaging over these gauge-equivalent choices, which causes A_{ij} (and other gauge-variant quantities) to vanish. In this sense there is no condensation.

However, usually there is still a transition, since not all phase configurations of the link variables, such as A_{ij}, are gauge equivalent, simply because there are more link variables than there are lattice sites. These issues have been discussed in the literature.[12,24] Briefly, phase configurations are grouped into gauge-equivalence classes: configurations within a given class are related by gauge transformations, Eq. (2), and thus correspond to the same physical state. For example, our choice for the order parameter [Eq. (3)] is a gauge related copy of that used in Ref. 5. Configurations belonging to different classes represent different physical states. These can be characterized by a residual phase, a gauge field, by fixing a gauge, which picks one configuration from each gauge-equivalence class.

The order parameter itself belongs to one particular class, in other words, the system condenses into an entire gauge-

equivalence class, corresponding to a single physical state. There is a rigidity relative to other classes since fluctuations of the gauge field (in a fixed gauge) lead to an excited state. Such gauge fluctuations are studied via a gauge theory, as has been done by many authors. The condensation usually leads to certain type of gauge-invariant lattice order. For example, the RVB state has an important two-sublattice property (see below). Other choices of symmetry lead to flux phases. Similarly, condensation of spinon pairs and holon pairs lead to condensation of gauge-invariant composite entities such as a electron pair, whose symmetry depends on the symmetry of the constituents.

There is thus a transition across T_{RVB}. In the uncondensed state (i.e., above T_{RVB}) both phase and amplitude fluctuations are strong, and spinons and holons remain localized. In the condensed phase, fluctuations are weak. In this paper we will use the term condensation in this sense and the term symmetry of a MF state to reflect the symmetry of the whole gauge-equivalence class.

Although gauge fluctuations are supposed to be weak in the ordered phase at low T, for a detailed quantitative theory (particularly for transport properties) one has to include them.[24] Similarly, the MF theory will not work well near T_{RVB} where it will be necessary to consider both amplitude and phase fluctuations. For the 2D system, these may be strong enough to convert the apparent second-order MF transition to a Kosterlitz-Thouless-type transition (see Ref. 12). We will not study these issues here, but focus on the symmetry related properties, which can be determined by continuing each phase down to $T=0$. MF approximations are adequate for this purpose.

At half filling, the MF theory has been formally extended by Read and Sachdev[18] to situations where AF order is destroyed at $T=0$ by quantum fluctuations. In the ordered phase spinons are gapless, leading to gapless spin-wave excitations. However, once magnetic order is destroyed, a gap Δ_s appears in the spinon spectrum. Both cases can be described by the same MF theory with different values of λ, which acts as the spinon chemical potential. The spinon spectrum has the form

$$\omega(\mathbf{k}) = [\lambda^2 - \phi(\mathbf{k})^2]^{1/2}, \tag{4}$$

where

$$\phi(\mathbf{k}) = 4JA(\sin k_x + \sin k_y) \tag{5}$$

is the spinon "gap" function. The minima of $\omega(\mathbf{k})$ are at $\pm(\pi/2,\pi/2)$. In the ordered state, λ is chosen so that $\omega(\mathbf{k})$ vanishes at these points leading to two-sublattice AF order.

In the absence of AF order, spinons are gapped since $\omega_{min}>0$. Then, as shown by Read and Sachdev,[18] gauge fluctuations lead to a proliferation of instantons, which results in the confinement of spinons into pairs. In this case, the Arovas-Auerbach-type state is unstable relative to a valence bond solid. In our case, we do not need to worry about this, since there is no gap due to the presence of AF order. Away from half filling a spinon gap does appear but, as we shall see, holon motion causes singlets themselves to hop around, which would destabilize the VB solid.

A. Correlation functions

Like the Néel state, the bosonic RVB state has a two-sublattice character of its own. From Eq. (5) we see that although $\phi(\mathbf{k})$ is not gauge invariant, it satisfies the gauge-invariant condition,

$$\phi(\mathbf{k}) = \phi(\mathbf{Q}-\mathbf{k}). \tag{6}$$

Now consider the spinon correlation function $B_{jm} = \frac{1}{2}\Sigma_\sigma \langle b^\dagger_{m\sigma} b_{j\sigma}\rangle$ for any two sites j and m. In the MF approximation this is given by

$$B_{jm} = \frac{1}{N}\sum_{\mathbf{k}} \cos[\mathbf{k}\,.\,(\mathbf{r}_m - \mathbf{r}_j)]\frac{\lambda}{\omega(\mathbf{k})}\{1/2 + n[\omega(\mathbf{k})]\}, \tag{7}$$

where $n(\omega)$ is the Bose function. Using Eq. (5), we find

$$B_{jm} = B_{jm}\cos[\mathbf{Q}\,.\,(\mathbf{r_m} - \mathbf{r_j})]. \tag{8}$$

Therefore, B_{jm} vanishes if j,m are on opposite sublattices. For j,m on the same sublattice B_{jm} is nonzero if the singlet condensate exists ($A\neq 0$). We will call this behavior *even*. Similarly, consider the anomalous correlation function, $A_{jm} = \frac{1}{2}\langle (b_{j\uparrow} b_{m\downarrow} - b_{j\downarrow} b_{m\uparrow})\rangle$, again defined for any two sites j,m. (For nearest neighbors this is the RVB order parameter.) It is given by

$$A_{jm} = \frac{i}{N}\sum_{\mathbf{k}} \sin[\mathbf{k}\,.\,(\mathbf{r}_m - \mathbf{r}_j)]\frac{\phi(\mathbf{k})}{\omega(\mathbf{k})}\{1/2 + n[\omega(\mathbf{k})]\}. \tag{9}$$

From which we find

$$A_{jm} = -A_{jm}\cos[\mathbf{Q}\,.\,(\mathbf{r_m} - \mathbf{r_j})]. \tag{10}$$

Thus A_{jm} is zero if j,m are on the *same* sublattice. For j,m on opposite sublattices, A_{jm} is nonzero as long as the singlet condensate exists. We will call this behavior *odd*. These symmetry properties are intrinsic to the RVB state since they hold even after long-range AF order is destroyed and for any $T<T_{\mathrm{RVB}}$. They are also gauge invariant, even though the correlation functions themselves are not, and therefore have observable consequences. Thus the spin-spin correlation function is given by $\mathcal{S}_{sp,ij}=\langle S^+_i S^-_j\rangle = -|A_{ij}|^2 + |B_{ij}|^2$, which, as expected, alternates in sign.

These properties imply that, on average, RVB singlets connect spinons on opposite sublattices. Therefore, the bosonic RVB state is similar to the short-range RVB state considered by Kivelson *et al.*[19] What we have shown here is that this RVB state has the same red-blue property even in the presence of long-range magnetic order, i.e., even when magnetic correlations are not short ranged.

B. Magnetically disordered phase

Away from half-filling quantum fluctuations due to hole hopping is expected to destroy long-range AF order at $T=0$, leaving behind the RVB phase. Then bosonic spinons acquire a gap Δ_s, which is related to the magnetic correlation length.[5,18] Near its minima the spinon energy has the form: $\omega(\mathbf{k}) \approx [\Delta_s^2 + c_s k^2]$, where $c_s = 2JA \sim T_{\mathrm{RVB}}$ and k is measured relative to the minima. Then the correlation functions B_{ij} and A_{ij} behave as

$$B_{ij} \sim \frac{1}{r_{ij}} e^{-r_{ij}/2\xi} \cos\left[\frac{1}{2}\mathbf{Q} \cdot (\mathbf{r_j} - \mathbf{r_i})\right],$$

$$A_{ij} \sim \frac{1}{r_{ij}} e^{-r_{ij}/2\xi} \sin\left[\frac{1}{2}\mathbf{Q} \cdot (\mathbf{r_j} - \mathbf{r_i})\right],$$

where $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|$ and $\xi=2c_s/\Delta_s$ is the spin-spin correlation length.

The spinon gap and the consequent exponential decay of AF correlations are the distinctive features of the bosonic RVB state. In contrast, spinons are gapless in a fermionic (i.e., slave boson) RVB state, which would lead to power-law correlations. Similarly, a Fermi liquid also has gapless spin excitations. Also, in the latter two cases, the magnetic correlations are peaked at a concentration-dependent wave vector $\mathbf{q}$, which equals $\mathbf{Q}$ only at half filling.

III. AWAY FROM HALF FILLING: EARLY WORK

A. One-hole physics and the t'-J model

Several authors have studied the behavior of a single hole moving in a magnetic background by expanding in powers of hopping and summing a selected a class of diagrams.[25,26] It has been found that because of spin mismatch the hole cannot hop coherently on to the opposite sublattice but it can move coherently within the same sublattice. For large t/J, the bandwidth corresponding to the coherent peak was found to scale with J, reflecting strong renormalization of holes by spin fluctuations, which makes the renormalized hole large and heavy. The short-range incoherent hops due to t leads to an incoherent spectral background of order t. These results are supported by exact numerical work on finite lattices containing one hole.[27] They can be used to estimate renormalized parameters for a many-hole theory. However it is erroneous to construct a low-energy theory based on the one-hole spectrum, since a single hole does not destroy magnetic order, the presence of which leads to a pole in the electron Green's function.

Motivated by the single-hole results, several authors have developed gauge theories for the many-hole system[15,16] based on the Schwinger boson RVB states (thus hoping to establish a connection with the Mott phase), with the assumption that the short-range incoherent hops caused by intersublattice hopping (t) destroys AF order and strongly renormalizes the theory, leading to a sublattice-preserving t'-J model, where $t' \sim J$ is a renormalized next-nearest-neighbor hopping parameter for renormalized holes.[16] While the assumptions are correct, the actual renormalized Hamiltonian, as we show later, is quite different.

B. Spiral and other nearest-neighbor states

One reason why Schwinger boson methods have not been taken seriously is the fear of a spiral instability. For $x>0$, moving holes are expected to rapidly destroy long-range AF order. As a physical hole hops, it carries its spin, creating a string of wrong (ferromagnetic) bonds, which cost energy and cannot be easily repaired. In cuprates the AF insulating phase disappears beyond a small x_c, giving way to a metal with no long-range magnetic order. Now, we can obtain a metallic state by extending the mean-field approximation to the hopping term,[8] which can be written as

$$-t\sum_{ij}[\mathrm{B}_{ij}^{\dagger}\mathrm{D}_{ij} + \mathrm{H.c.}],$$

where $\mathrm{D}_{ij}=h_j^{\dagger}h_i$ and $\mathrm{B}_{ij}=\frac{1}{2}\Sigma_{\sigma}b_{j\sigma}^{\dagger}b_{i\sigma}$. In the MF approximation, in addition to singlets and spinons, the composite bosons created by B_{ij} and D_{ij} also condense, so that the hopping term becomes

$$\sum_{ij} B_{ij}^{*}h_j^{\dagger}h_i + \sum_{ij,\sigma} D_{ij}b_{i\sigma}^{\dagger}b_{j\sigma},$$

where $B_{ij}^{*}=\langle \mathrm{B}_{ij}^{\dagger}\rangle$ and $D_{ij}=\langle \mathrm{D}_{ij}\rangle$ are the average spinon and holon backflows, respectively. Appropriate choices of symmetry for these order parameters allow holons and spinons to hop coherently on to nearest-neighbor (nn) sites, which mixes up the two sublattices. Different choices lead to different metallic states. The state with $B_{ij}=B$, $D_{ij}=D$, has spiral magnetic order with an incommensurate wave vector. The spiral state is energetically favored over other relevant nn states, e.g., the flux state, the ferrimagnetic state, etc.[28] Actually no sign of a spiral state (or any other nn state) has been found in cuprates, or for that matter, in numerical treatments of the model.

Incidentally, the slave-boson RVB theories are also based on nn states, with a similar set of order parameters, although interpretations are different.[11,12] In both cases, the nn states are very complex, characterized by several order parameters (A, B, D, and $\langle b\rangle$). This leads to a complicated phase diagram,[11,12] containing several metallic phases, more complicated than that seen in cuprates. Furthermore, since slave (or Schwinger) bosons condense, the physical electron Green's function has a pole, with a residue $\propto|\langle b\rangle|^2$. Hence, the normal state is in effect a Fermi liquid at low T with gapless spinon and holon excitations.

Also, since slave-boson states do not work at half filling we should see a transition accompanied by a statistical transmutation upon doping. No experimental signature for such a transition has been found. Moreover, Haldane and Levine[29] and, independently, Read and Chakraborty[30] have used Berry phase arguments to show that, in the presence of short-range RVB pairing, spinons are bosons and holons are spinless fermions, as in our case. Finally, energetics also favor the Schwinger-boson representation, for the simple reason that, given the same spectrum, fermions will have much higher energy than bosons because of the Pauli exclusion principle. Therefore, bosonic spinons are better for the exchange term. Interestingly, they are also found to be better for the hopping term for $x<1/3$ even if $J=0$.[31]

Instability of spiral states. Moving holes induce ferromagnetic correlations, which compete with antiferromagnetism. In the spiral state the system compromises by tilting neighboring spins so that the correlations are partially ferromagnetic. While this allows the holons to delocalize and reduce kinetic energy, there is a substantial cost in exchange energy. Therefore such a state may not be stable. Within a Hartree-Fock approximation, several authors have found an instability in the 2D Hubbard model toward localization of holes in

domain walls[32] for small x. An instability toward phase separation was earlier found by Visscher.[33]

By comparing the Hartree-Fock energies, Hu *et al.*[14] showed that spiral states are indeed unstable relative to both phase separation and insulating domain-wall states in the entire region of the parameter space of interest. The actual region of instability is likely to be even larger since contributions from RVB singlets to the exchange energy are not included in the Hartree-Fock treatment. Similar results were found for the t-J model, where both antiferromagnetism and singlets are included.[28] Since the spiral state has the lowest energy among the nn states, it follows that none of these states are stable.

Longer-range Coulomb interaction would destabilize the domain walls and spin-charge separated states, and at very low doping, could stabilize a Wigner crystal of holes. However, since $t > J$ the localized states cost too much kinetic energy, a metal must emerge at higher doping with or without such long-range repulsion. It follows that such a metal would have a lower energy than the spiral state, and can only appear via higher-order hopping processes by renormalized holes, as indicated by the one-hole calculations. To derive a renormalized Hamiltonian and for the results obtained in this paper it is not necessary to know the details of the renormalization procedure. But for the sake of consistency, and for a more detailed theory, it is useful to consider the possible physical processes involved.

C. Destruction of magnetic order and renormalization

Our work is based on an earlier paper[21] in which a self-consistent perturbation expansion in powers of hopping to one-loop order was used to study the destruction of magnetic order upon doping, which is expected to occur, but hard to show theoretically. The point is that localization of electrons into spins (moments) costs considerable amount of kinetic energy. As discussed above, trying to recover the energy via spin-charge separated nn states does not work. Another process which always exist is the binding of spin and charge into physical electrons (or holes). These try to propagate coherently and restore the Fermi liquid, which has low kinetic energy.[31]

In Ref. 21, this process has been treated by RPA, in which the electron is treated as a collective excitation. The zeroth-order Green's function G_{c0} is simply the convolution of the spinon and holon Green's function. Summing the bubbles leads to the full electron Green's function,[21]

$$G_c(\mathbf{k}\omega) = \frac{G_{c0}(\mathbf{k}\omega)}{1 - \epsilon(\mathbf{k})G_{c0}(\mathbf{k},\omega)}, \tag{11}$$

where $\epsilon(\mathbf{k})$ is the energy of the noninteracting electron. This can be viewed as a generalization of the one-hole method where a similar expansion is carried out for one hole. In the many-hole case, the effect of the electron on the spinon and holon self energies is also calculated, using G_c. Unlike moving holons which only change the direction of the local moments, a moving electron has a much more violent effect, as it also tends to delocalize, and hence destroy the moments. It was found that $x > x_c$, where x_c is rather small, long-range magnetic order is destroyed at $T=0$, i.e., there is no Bose condensation of spinons.[21]

It was also found that G_c does not have a pole at low frequencies, that is, the system is not a Fermi liquid for small x. The reason is that in order to move an electron must avoid other electrons of opposite spin.[31] For small x and at low dimensionality, such pathways are rare. Consequently, the electrons (or physical holes) do not become coherent. However, spinons become gapped since long-range magnetic order is destroyed. The gap shows up in G_{c0} and hence in G_c.[34] In contrast, in the one-hole case, the Green's function has a pole, which is not surprising since a single hole cannot destroy long-range AF order. In either case, the holes are strongly renormalized by the short-range incoherent hops.

Interestingly, it was also found that the presence of a RVB condensate leads to a second-order hopping process in which a pair of holons hop together, accompanied by a singlet backflow, which automatically leads to superconductivity. This is the pairing mechanism considered below. The theory was further developed in Ref. 22 where an effective Hamiltonian for the renormalized holons including the pairing term was derived. However, it was not entirely correct since the normal state was assumed to be a spiral metal, with nonzero B and D, but without long-range magnetic order. Such a state is not stable. Below we give a derivation which corrects this flaw.

IV. RENORMALIZED HAMILTONIAN FOR SMALL DOPING

The spinon gap allows us to obtain an effective Hamiltonian involving renormalized holes and singlets by integrating out the spinons and setting $\omega=0$ as was done in Refs. 21 and 22. Here we give a simpler derivation using the usual continuity arguments.[16,19] Since the nn MF state is not stable we have $B=0$ and $D=0$. A moving hole affects antiferromagnetic configurations and singlets differently. In the former case, a string of ferromagnetic bonds are created whose energy increases with the length of the path, and which cannot be easily repaired quantum mechanically. The hole is essentially forced to return, creating a renormalized hole localized within a bag of spin excitations. This process has been considered by many authors in the context of the one-hole problem.[25–27] It is also the reason for localization of holes in domain walls in the Hartree-Fock treatment of the Hubbard model.[14,32] The hole moves rapidly within the bag, gaining kinetic energy of order t. This would contribute an incoherent (essentially $\mathbf{k}$-independent) background to the electron spectral function as well as to the optical conductivity. We assume that disordering of the spins prevents long-range AF order and leads to short-ranged AF correlations. The renormalized hole carries the spin excitations, and is thus larger and more massive.

After this initial renormalization, the Hamiltonian would be similar to the original one by continuity, except for the absence of magnetic order, as long as no additional symmetries are broken. It will be characterized by a hopping amplitude $t_{eff} < t$, and an exchange coupling J_{eff}. For small x, we expect $J_{eff} \sim J$. Also, one-hole calculations[25,27] suggest that

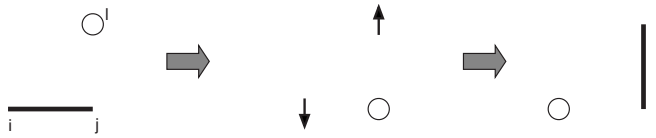

FIG. 1. One hole process. A hole hops from l to j, breaking the singlet (ij) denoted by the solid line. Here i and l are on sublattice a and j is on sublattice b. The hole then hops to i and the singlet is reconstituted at (jl)

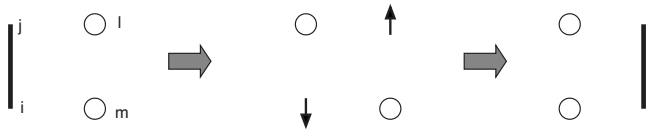

FIG. 2. Two hole process. A hole hops from l to j, breaking the singlet (ij), a second hole hops from m to i and the singlet is reconstructed at (ml).

t_{eff} scales with J for $t>J$. Such estimates can be used since the renormalization process involves only nearby hops, and therefore at low-hole densities it is essentially a one-hole problem. An important aspect of the one-hole calculations is that the local constraints are imposed exactly. The constraints obeyed by the renormalized particles are thus expected to be weaker, although the renormalized Hamiltonian retains gauge invariance.

The singlets are affected less drastically by hole motion since the offending configurations can be repaired more easily, as discussed below. We therefore assume that the problem is already renormalized by the fast processes described above. When a renormalized hole hops from sublattice a to b it breaks a singlet and creates two spinons, costing, say, an energy $\Omega>2\Delta_s$. The system relaxes by a second hop, after which the singlet is reconstructed. In the low-frequency limit $(\omega \ll \Omega)$ these processes can be described by

$$H_{int} = -\frac{t_{eff}^2}{\Omega}\mathcal{P}\sum_{jl\sigma,mi\sigma'} c_{m\sigma}^\dagger c_{i\sigma} c_{l\sigma'}^\dagger c_{j\sigma'}\mathcal{P}, \tag{12}$$

where $c_{i\sigma}=b_{i\sigma}^\dagger h_i$; but now all the operators correspond to renormalized particles. Here $\mathcal{P}\cdots\mathcal{P}$ means that we keep only those terms that involve renormalized singlets and holes. Although nominally similar to superexchange, the intermediate state does not involve double occupancy which has already been projected out. As shown below, the new interaction terms can be rewritten in terms of singlet operators A_{ij}. Now, A_{ij} and B_{ij} decay exponentially as $e^{-r_{ij}/2\xi}$,[18] where ξ is the magnetic correlation length. Therefore, to obtain a minimal Hamiltonian, it is sufficient to retain only the short-range hopping terms since longer range terms are exponentially suppressed. The longer-range hopping terms also preserve the underlying symmetries and neglecting them should not change the physics qualitatively.

Once the singlet is broken, there are three ways to remove the excess energy. First, the hole can hop back and the singlet reconstructed. This is confining and its effect is to renormalize the chemical potentials. The remaining two processes can lead to hole propagation.

(A) *One-hole process.* The hole hops to another site on sublattice a and the singlet is reconstructed on a different link (Fig. 1). Notice that this process involves successive hopping by two electrons of opposite spin but only one holon. Projecting onto the spinon singlet subspace $[\mathcal{P}b_{i\sigma}^\dagger b_{j,-\sigma}\mathcal{P}=\sigma\mathrm{A}_{ij}$, etc.], we obtain

$$-\frac{t_s}{2}\sum_{ijl}\mathrm{A}_{jl}^\dagger \mathrm{A}_{ij} h_i^\dagger h_l(1-h_j^\dagger h_j), \tag{13}$$

where, $\mathrm{A}^\dagger$ creates a singlet and $h^\dagger$ creates a *renormalized* holon, and $t_s=4\frac{t_{eff}^2}{\Omega}$ is the effective intrasublattice hopping parameter. Now, the term $(1-h_j^\dagger h_j)$ arises from the intermediate excited state. It is gauge invariant, and for small x, can be approximated by its average value $1-x$, leading to

$$-\frac{t_s}{2}(1-x)\sum_{ijl}\mathrm{A}_{jl}^\dagger \mathrm{A}_{ij} h_i^\dagger h_l. \tag{14}$$

This term describes hopping by a holon within the same sublattice, accompanied by a backflow of spinon singlets.

The one-hole term above differs from the usual next-nearest-neighbor hopping (t'). In the early RVB-gauge theories based on similar renormalization considerations it was *assumed* that the result of second-order hopping is to generate an effective t' term of the form

$$t'\sum_{il} c_{i\sigma}^\dagger c_{l\sigma} + \mathrm{H.c.},$$

where i and l are next-nearest neighbors. To generate such a term from Eq. (12) would require consecutive hops by two electrons of the *same* spin. This is not easily done when the initial and final states belong to the singlet subspace, as in our case. For after the first hop the remaining electron will have opposite spin. Then to generate a t' one has to flip both spins via additional higher-order intermediate processes, which are not very likely. (It can be done via J' interaction. But for this model J' is usually much smaller than J.) Of course the actual model that applies to real cuprates will have a t' term to start with which, as will see later, will play a role since it also preserves the two-sublattice character, but it would not replace the dominant one-hole term derived above.

(B) *Two-hole process.* The system also relaxes if a second hole hops from sublattice b to a, and the singlet is reconstituted on a different link (Fig. 2). This process yields a term

$$-t_s\sum_{ij;lm}\mathrm{A}_{ml}^\dagger \mathrm{A}_{ij} h_i^\dagger h_j^\dagger h_l h_m, \tag{15}$$

which describes hopping by a singlet, accompanied by the backflow of a holon pair. This is the small x form of the interaction derived earlier[22] but here the normal state is different because of the one-hole term above.

The full Hamiltonian also includes the renormalized exchange term, characterized by J_{eff}. For the purposes of this paper it is sufficient to treat J_{eff}, t_{eff}, and Ω as phenomenological parameters. However, as discussed earlier, one-hole calculations[27] suggest that they all scale with J. Hence the renormalized Hamiltonian describes a strong-coupling problem. For cuprates, $J\sim 0.1$ eV, and is the largest energy scale

in the renormalized problem. If, $t_{eff} > \Omega$ numerically, then a frequency cutoff can be used. By continuity, qualitative physics will not be changed. As it turns out (see below), the effective holon hopping amplitude $t_{eff}A$ is $\sim JA \sim T^*$. Since A is only a small fraction of unity and decreases with x, a cutoff may not be necessary, at least for some values of x.

Because of the presence of the spin gap our renormalized Hamiltonian describes a short-range RVB model for small x.[19] However, the Hamiltonians used by earlier authors[20] were different (for example, the second term which is responsible superconductivity was absent) since they were not required to agree with Arovas-Auerbach state at half filling. Also, holons were thought to be bosons. It was later shown that for short-range RVB models they are actually fermions, which agrees with our choice.[29,30] Our renormalized Hamiltonian cannot possess a conventional Fermi-liquid state since the latter has gapless spin (and charge) excitations.

V. ANALYSIS OF THE RENORMALIZED HAMILTONIAN

The renormalized Hamiltonian has the symmetries of the original model since the singlet variables A_{ij} are not condensed yet. The Hamiltonian still describes a formidable strong-coupling problem. However, we can obtain a number of very important results by continuing the spinon states from half filling, thereby fixing the spin structure. In particular, we can determine the phases in the underdoped region that are consistent with our two basic assumptions, as well as the nature of these phases.

A. Nonordered phase: Strange metal or insulator?

There is always a state in which no symmetry is broken (even at the MF level). Then singlets are not condensed, i.e., $A_{ij}=0$. Hence,

$$\langle h_i^\dagger h_j \rangle = 0$$

for $i \neq j$ so that there is no coherent propagation of holons. Holons are localized due to strong gauge fluctuations. This state has no long-range order and would always occur at sufficiently high T (above T^* in our case). We identify it with the so-called strange metal phase of the cuprates. Since electrons are also incoherent, there are no coherent charge carriers at all in this state. Thus when continued to $T=0$, it would have infinite resistivity since the electrons are also gapped. Therefore, the strange metal is not a metal; it is a new type of quantum insulator. In effect, this state corresponds to a Gutzwiller projected Fermi sea; but one with strong singlet fluctuations, which lead to the gapping of the spinons (and, hence, the electrons), and non-Fermi-liquid behavior. In contrast, in other RVB theories the state above T^* is metallic, with gapless spin and charge excitation.[11,12] In these theories the nonordered state appears at a much higher temperature and is ignored.

B. Pseudogap phase

To proceed further, we first do a MF decomposition of H_{int} to obtain two separate Hamiltonians for spinons and holons, which breaks gauge symmetry. Then, from the one-hole term, we get

$$-\frac{t_s}{2}(1-x)\sum_{ijl}[\langle \mathrm{A}_{jl}^\dagger \mathrm{A}_{ij}\rangle h_i^\dagger h_l + \langle h_i^\dagger h_l\rangle \mathrm{A}_{jl}^\dagger \mathrm{A}_{ij}]. \quad (16)$$

In principle, we can find a state for which $\langle \mathrm{A}_{ij}\rangle = 0$ (no singlet condensation) but $\langle \mathrm{A}_{jl}^\dagger \mathrm{A}_{ij}\rangle \neq 0$. Such a state will not be favored energetically since the energy gained $O(t_s x)$ per bond is low compared with $J_{eff}A^2$ gained from singlet condensation.

This leaves the RVB state, which becomes stable below T^* with $A_{ij} \neq 0$. From Eq. (2) we obtain

$$\langle \mathrm{A}_{jl}^\dagger \mathrm{A}_{ij}\rangle = A_{jl}^* A_{ij} = -A^2 e^{i(1/2)\mathbf{Q}.(\mathbf{r_i}+\mathbf{r_l})}.$$

Using this in Eq. (15) we obtain an effective holon hopping Hamiltonian,

$$H_{h0} = -\sum_{ij} t_{h,ij} h_i^\dagger h_j, \quad (17)$$

which describes coherent holon propagation within the same sublattice. Here $t_{h,ij} = \pm t_h$ for i,j next-nearest neighbors along the $(1,\pm 1)$ diagonal direction, $t_{h,ij} = t_h/2$ for next-next-nearest neighbors (along the x and y directions), and $t_h = t_s A^2(1-x)$. The holon energy is then given by

$$\epsilon_{h1}(\mathbf{k}) = -2t_h + 2t_h(\sin k_x + \sin k_y)^2. \quad (18)$$

The holon band (hence, metallic conduction) appears as soon as $A \neq 0$ without additional symmetry breaking. As expected, the spectrum has the two-sublattice character since $\epsilon_{h1}(\mathbf{k}) = \epsilon_{h1}(\mathbf{Q}-\mathbf{k})$.

We can do a similar MF decomposition of the two-hole term [Eq. (14)], and replace the A_{ij}'s by their expectation values, which yields a holon-holon interaction term of the form

$$H_{h,int} = -t_s A^2 \sum_{ij;lm} [\mathrm{F}_{ij}^\dagger \mathrm{F}_{ml} + \mathrm{H.c.}], \quad (19)$$

where $\mathrm{F}_{ij}^\dagger = h_i^\dagger h_j^\dagger$ creates a holon pair on the link ij and the sum is over plaquettes. The order of the indices follows from the symmetry of A_{ij} and is very important.

Once the holons are propagating, the pair-hopping term will also contribute to the normal-state holon spectrum. Doing a MF decomposition of $H_{h,int}$ we obtain a term

$$H_2 = t_s A^2 \sum_{ij,ml} [D_{li} h_j^\dagger h_m + D_{mj} h_i^\dagger h_l - D_{li} D_{mj} + \mathrm{H.c.}],$$

where

$$D_{li} = \langle h_i^\dagger h_l \rangle = \frac{1}{N}\sum_{\mathbf{k}} e^{i\mathbf{k}.(\mathbf{r}_l - \mathbf{r}_i)} f[\epsilon_h(\mathbf{k}) - \mu] \quad (20)$$

is the average holon hopping amplitude and f as the Fermi-Dirac function. Using the symmetry of the holon spectrum, we find that $D_{li}=0$, when l and i are nearest neighbors. And, $D_{li} = \pm D_1$ for l and i along the $(1,\pm 1)$ diagonal directions, respectively, with

$$D_1 = -\frac{1}{N}\sum_{\mathbf{k}} \sin k_x \sin k_y f[\epsilon_h(\mathbf{k}) - \mu]. \tag{21}$$

Therefore the extra diagonal hopping terms have the same symmetry as the original ones. Furthermore, since the minima of the holon spectrum is at $\pm\frac{1}{2}(\pi,-\pi)$, D_1 is positive. Using the fact that hopping along (1,1) is accompanied by a backflow along $(1,-1)$ the extra diagonal terms are found to have the same signs as before. Including these, the full holon spectrum becomes

$$\epsilon_h = -2t_h + 2t_h(\sin k_x + \sin k_y)^2 + 4D_1 t_s \sin k_x \sin k_y. \tag{22}$$

The parameter D_1 has to be calculated self-consistently. However, for small x, the main contribution comes from near the minima of the spectrum $[\mathbf{k} = \pm(\pi/2,-\pi/2)]$, where we can set $\sin k_x \sin k_y = -1$, which yields $D_1 \approx x$.

So far we have considered only the holon part. The motion of holons (or holon pairs) is accompanied by a backflow of singlets. The mean-field singlet Hamiltonian is then given by $H_b = H_J + H_{b1} + H_{b2}$, where H_J is the usual exchange term [see Eq. (1)] with an exchange constant J_{eff},

$$H_{b1} = -\frac{t_s}{2}(1-x)\sum_{ijl} \mathrm{A}^{\dagger}_{jl}\mathrm{A}_{ij}\langle h_i^{\dagger} h_l\rangle \tag{23}$$

is the singlet backflow term associated with the one-hole process and

$$H_{b2} = -t_s \sum_{ij;lm} \mathrm{A}^{\dagger}_{ml}\mathrm{A}_{ij}\langle h_i^{\dagger} h_j^{\dagger} h_l h_m\rangle \tag{24}$$

is that associated with the two-hole process. The last two terms only appear below T^*. They will contribute to the singlet condensation energy in the pseudogap phase. They are suggestive of the emergence of coherent singlet excitations, which are spin-0, chargeless vector bosons. These will contribute to low-T specific heat, and thermal conductivity, but not to electronic transport. We will study them in a future paper.

Symmetry properties of the pseudogap phase. The pseudogap state preserves the two-sublattice (red-blue) character of the RVB state. To see this consider the correlation functions for the renormalized particles. Using $\epsilon_h(\mathbf{k}) = \epsilon_h(\mathbf{Q} - \mathbf{k})$, we find $D_{ij} = \langle h_j^{\dagger} h_i\rangle = D_{ij}\cos[\mathbf{Q}.(\mathbf{r_j} - \mathbf{r_i})]$, which is thus nonzero only on the same sublattice (even). The spinon correlation functions, and hence the gauge-invariant magnetic correlation functions have the same symmetry as in the Mott phase. The electron hopping amplitude $P_{ij,\sigma} = \langle c_{i\sigma}^{\dagger} c_{j\sigma}\rangle = -B_{ij}D_{ij}$. Since B_{ij} and D_{ij} are both even, P_{ij} is nonzero only on the same sublattice. The Fourier transform of P_{ij} is the momentum distribution function, which thus satisfies $n_c(\mathbf{k}) = n_c(\mathbf{Q} - \mathbf{k})$. In a Fermi liquid, the electron hopping amplitude shows a power-law decay. In our case, the holon hopping amplitude D_{ij} shows a metallic, i.e., power-law behavior since holons are gapless. However, since spinons are gapped, B_{ij}, hence the full electron hopping amplitude P_{ij} decay exponentially, reflecting non-Fermi-liquid behavior.

The metallic character of the holons can also be seen in the gauge-invariant charge structure factor. Let $\rho_i = h_i^{\dagger} h_i - \langle h_i^{\dagger} h_i\rangle$ measure the excess hole density. Then the charge structure factor is given by: $\mathcal{S}_{ch,ij} = \langle \rho_i \rho_j\rangle = -|D_{ij}|^2$, for $i \neq j$, and $\mathcal{S}_{ch,ii} = x(1-x)$. This is nonzero on the same sublattice and exhibits the long-range oscillatory structure of a metal. In the momentum space it satisfies $\mathcal{S}_{ch}(\mathbf{k}) = \mathcal{S}_{ch}(\mathbf{Q} - \mathbf{k})$ (Fig. 3). In contrast, $\mathcal{S}_{ch}(\mathbf{k})$ of an ordinary metal increases from zero at $\mathbf{k} = 0$ and becomes a constant for $q > 2k_F$. In our case, an image of the behavior near $\mathbf{k} = 0$ appears near $\mathbf{k} = \mathbf{Q}$. Experiments probe the bare correlation functions. These are dominated by short-range incoherent processes that do not preserve the two-sublattice property, which is therefore not easy to see. The best candidate is $\mathcal{S}_{ch}(\mathbf{k})$ since holon motion is coherent. The experimental $\mathcal{S}_{ch}(\mathbf{k})$ would no longer vanish at $\mathbf{Q}$ but there will still be a dip.

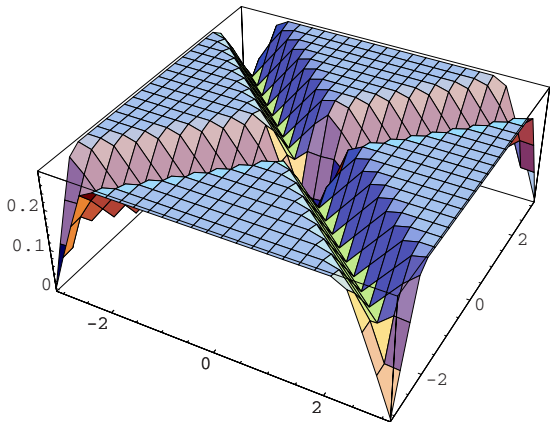

FIG. 3. (Color online) Charge structure factor showing the two-sublattice symmetry. The calculation is approximate. Note the symmetry between $\mathbf{k} = 0$ and $\mathbf{k} = (\pi,\pi)$, which is distributed among the four corners of the Brillouin zone. The valley region is elevated relative to $\mathbf{k} = 0$. In a normal metal, the structure near (π,π) is absent.

VI. SUPERCONDUCTING STATE

It is clear from the structure of the two-hole term [Eq. (19)] that the system can lower its kinetic energy if two holons hop as a pair. Pairs will condense, leading to $F_{ij} = \langle \mathrm{F}_{ij}\rangle \neq 0$. Let us define the pairing order parameter for physical electrons as a singlet,

$$\mathcal{C}_{ij} = \langle (c_{j\downarrow}c_{i\uparrow} - c_{j\uparrow}c_{i\downarrow})/2\rangle. \tag{25}$$

Then, $\mathcal{C}_{ij} = -A_{ij}F_{ij}^* \neq 0$ since the spinons are already condensed, giving rise to superconductivity below $T_c \leq T^*$. The order parameters A_{ij} and F_{ij} are not gauge invariant but $\mathcal{C}_{ij}$ is. The symmetries of A_{ij} and the holon spectrum are of course already known.

We solve the superconductivity problem by a mean-field approximation. Since holons are fermions the holon pairing order parameter satisfies $F_{jm} = -F_{mj}$. We denote the vector F_{jm} by two components: $F_{jm} = iF_x$ if m is to the right of j and $F_{jm} = iF_y$ if m is above j in the y direction. The prefactor i is chosen so that order parameter for the electron (Cooper) pair is real. For a uniform system, we can take $|F_{jm}| = F_0$, but the phases along x and y need not be the same. Without loss of

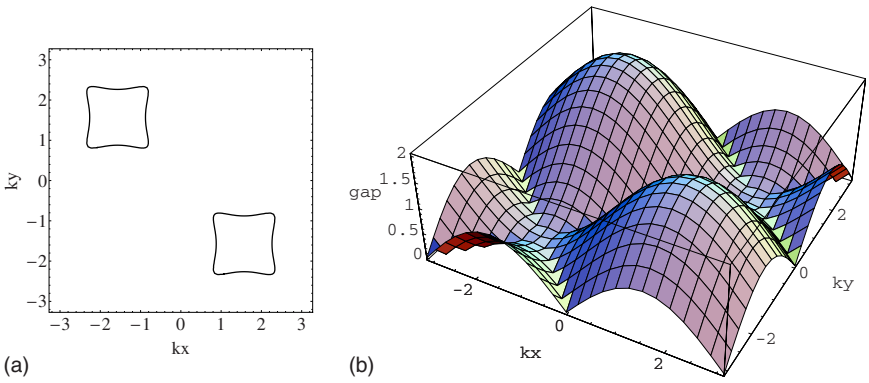

FIG. 4. (Color online) Origin of d-wave symmetry. Upper panel: holon Fermi surface. Holons live within the pockets centered at $\frac{1}{2}(\pi,-\pi)$ and $\frac{1}{2}(-\pi,\pi)$. The shape of the FS is somewhat different when compared with Ref. 17 because the holon spectrum has corrections from the two-hole term. The FS is not directly observable since under a gauge transformation it is moved and deformed. Lower panel: the symmetry factor $|\sin k_x+\alpha \sin k_y|$. It has broad maxima in the hole-rich region for $\alpha=-1$, resulting in maximum condensation energy. In contrast, the symmetry factor vanishes in this region for $\alpha=1$ (s wave). Under a gauge transformation both the Fermi surface and the symmetry factor move together to preserve these gauge-invariant results.

generality we can choose, $F_x=F_0$ and $F_y=\alpha F_0$ with $\alpha=e^{i\theta}$. The choice of $\alpha=\pm 1$ leads to $\mathcal{C}_x=\pm\mathcal{C}_y$, corresponding to s-wave (d-wave) symmetry for the electron pair wave function. But other values of α can, in principle, lead to $s+id$ or $s+d$ symmetries. To find the correct symmetry we compare the free energies for different choices. The MF Hamiltonian is given by

$$H_{\mathrm{MF}}=\sum_{\mathbf{k}} \xi(\mathbf{k})h_{\mathbf{k}}^{\dagger}h_{\mathbf{k}}+\frac{1}{2}\Delta_h(\mathbf{k})(h_{\mathbf{k}}^{\dagger}h_{-\mathbf{k}}^{\dagger}+\mathrm{H.c.}), \tag{26}$$

where $\xi(\mathbf{k})=\epsilon_h(\mathbf{k})-\mu$, μ is the holon chemical potential, and $\Delta_h(\mathbf{k})=2t_sF(\mathbf{k})$ is the holon gap function with

$$F(\mathbf{k})=2F_0(\sin k_x+\alpha \sin k_y). \tag{27}$$

The holon spectrum $\epsilon_h(\mathbf{k})$ is assumed to contain the Hartree contribution from the interaction term. The Hamiltonian is diagonalized by the Bogulyubov transformation. The self-consistent equations for the order parameters are then given by

$$F_\eta=\frac{4t_s}{N}\sum_{\mathbf{k}} \sin k_\eta(\sin k_x+\alpha \sin k_y)\frac{\tanh[\beta E(\mathbf{k})/2]}{E(\mathbf{k})}, \tag{28}$$

where $\eta=(x,y)$, and

$$E(\mathbf{k})=[\xi^2(\mathbf{k})+|\Delta_h(\mathbf{k})|^2]^{1/2}.$$

Since F_x is real we find that solutions exist only for $\sin\theta=0$, i.e., for $\alpha=\pm 1$. Combining the equations for F_x and F_y, we obtain

$$\frac{1}{t_s}=\frac{1}{N}\sum_{\mathbf{k}} W(\mathbf{k})(\sin k_x+\alpha \sin k_y)^2\frac{\tanh[\beta E(\mathbf{k})/2]}{E(\mathbf{k})}, \tag{29}$$

$W(\mathbf{k})$ is a suitably chosen cutoff function.

To find the symmetry let us consider the $T=0$ case, for which $\tanh[\beta E(\mathbf{k})/2]=1$. We have solved this equation numerically with and without a cutoff function. Quite generally, we find that $\alpha=-1$, (i.e., d wave) is favored, as it leads to the largest F_0, and hence the largest condensation energy. The origin of this result can be understood from the following simple considerations. The dominant contribution to the sum comes from the region where $|\xi(\mathbf{k})|=|\epsilon_h(\mathbf{k})-\mu_h|$ is small, and the symmetry factor $|\sin k_x+\alpha \sin k_y|$ is large. As shown in Fig. 4, the holon Fermi surface is in the second and fourth quadrant, exactly where $\sin k_x+\alpha \sin k_y$ has maxima for $\alpha=-1$ (d wave) and vanishes for $\alpha=1$ (s wave). Hence, d wave always wins. Thus the symmetry of the superconducting order parameter is determined by the symmetry of the underlying RVB state.

Note that since $F(\mathbf{k})=F(\mathbf{Q}-\mathbf{k})$, the two-sublattice property is also preserved in the superconducting state. The holon pairing function is odd since for any two i,j, it satisfies $F_{ij}=-F_{ij}\cos \mathbf{Q}.(\mathbf{r_i}-\mathbf{r_j})$. Since A_{ij} is also odd, the electron pairing function $\mathcal{C}_{ij}=-A_{ij}F_{ij}^*$ is also odd. Similarly, the symmetries of $n_c(\mathbf{k})$ and spin-spin correlation function remain unchanged in the superconducting state. The charge structure factor, however, picks up an additional contribution: $\mathcal{S}_{ch,ij}=|F_{ij}|^2-|D_{ij}|^2$, and is no longer restricted to the same sublattice; but like the spin-spin correlation function, it oscillates in sign.

Note that the renormalized Hamiltonian describes a strong-coupling problem since the one-hole and two-hole terms are essentially governed by the same energy scale t_s. The holon wave functions depend on x but not on t_s. Hence, the BCS-type MF approximation is expected to work only for $T\ll T_c$. In particular, treatment of the transition region and the calculation of $T_c(x)$ will require strong-coupling techniques. Similarly, MF approximations would not work near T^*, where fluctuations will be strong. The phase diagram shown in Fig. 5 is thus only schematic. The main point is, unlike earlier RVB theories, the number and type of phases for small x are similar to those in cuprates. Moreover, as discussed below, the qualitative behavior of various phases can be easily understood. Given the numerous unusual properties of the cuprates, and the highly constrained nature of the renormalized Hamiltonian, the predictions can be put to severe experimental tests. So far we have found no contradictions.

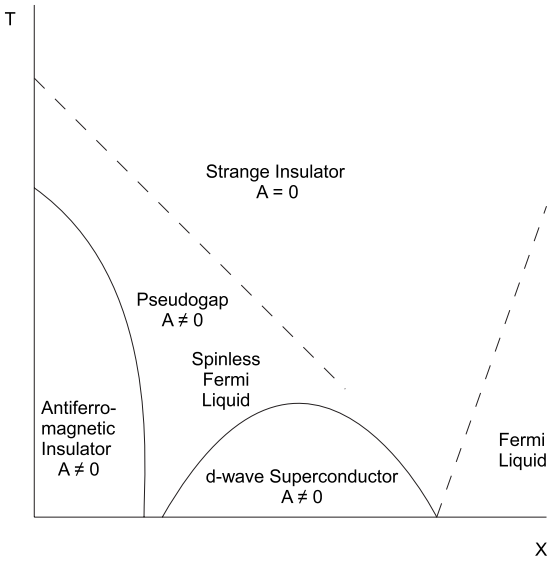

FIG. 5. Schematic phase diagram. The state above T^* is a novel insulator and is unordered. The pseudogap metallic state is a spinless Fermi liquid. Except for the strange insulator, the remaining phases are characterized by RVB order parameter $A \neq 0$

VII. COMPARISON WITH EXPERIMENTS

Clearly, the present theory is consistent with the important features of the experimental phase diagram. By construction, it is consistent with the physics at half filling. The normal state has spin-charge separation and is not a Fermi liquid. There is a transition from a strange phase to a pseudogap metal below T^*. As discussed later in Sec. VIII, these phases are two dimensional, whereas the superconducting state exhibits 3D behavior. However, the phases themselves are predicted to have novel characteristics that defy conventional wisdom. The key ones are: (1) there is a true spin gap Ω which exists in all three phases and which is distinct from the pseudogap that appears below T^*, (2) holons are confined above T^* so that the strange phase is an insulator rather than a metal; (3) below T^* holons form a spinless Fermi liquid of concentration x without an observable (small) Fermi surface. The pseudogap metal and the superconductor are characterized by the RVB order parameter and the two-sublattice property, whereas the strange phase does not have any order. These results are robust as they follow from symmetry and particle statistics. A large number of important conclusions can be drawn from the very structure of the effective Hamiltonian in each phase, which do not require detailed calculations, and which can be tested against experiments. We now examine recent experimental data to show that there is very strong evidence in support of these predictions. We separate the experimental findings into three groups: the spin sector, the charge sector, and the electron sector.

A. Spin sector

The presence of the spin gap implies the existence of an additional temperature scale $T^0 \sim \Omega/k$. In the pseudogap region the AF correlation length ξ_{spin} is expected to be short, no more than a few lattice spacings, due to strong renormalization processes implicit in our theory. Then, $\Omega \sim J_{eff}(x)$. On the other hand, $T^* \sim J_{eff} A < T^0$ since A is only a fraction of unity (from the MF theory). At half filling, RVB ordering arises from long-range singlet-singlet interactions mediated by gapless spinons. In the doped region, this interaction is weaker and short ranged because of the gap. However, now the renormalized hopping terms favor singlet condensation since holons can propagate coherently only if singlets are correlated. The qualitative behavior is easy to understand.

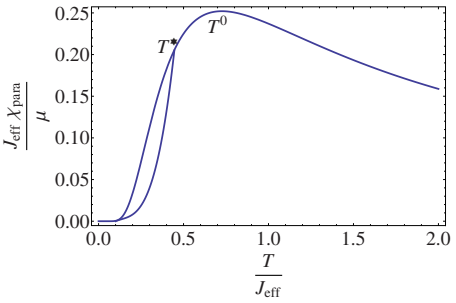

FIG. 6. (Color online) Paramagnetic susceptibility for a two-site problem. Note the maximum at T^0, the spin-gap scale. The line on the right below T^* is schematic, showing how singlet condensation at T^* for a many-site problem would further depress χ_{para}.

Above T^0 we expect free spins to dominate. Then the uniform paramagnetic susceptibility χ_{para} should be Curie type, i.e., decrease with increasing T. Below T^0 singlets will form and χ_{para} will start decreasing with *decreasing* T. Note that there is no transition at T^0; it is just a broad crossover scale. Hence, we can associate it with the maximum of χ_{para}. With decreasing T, there will be more singlets and they will become increasingly more correlated, and eventually condense at T^*. Below T^*, χ_{para} will decrease much more rapidly, and vanish as $T \to 0$ (because of the gap), even in the absence of superconductivity. The magnetic behavior above T^* is determined by fluctuating singlets correlated over a distance $\xi_{singlets}(T)$, together with weakly correlated free spins. The behavior near T^0 and above, where singlets are weakly correlated, can be crudely understood in terms of a one singlet (two-site) problem, for which χ_{para} is given by

$$\chi_{para2} = \frac{\mu^2 \beta}{3 + e^{\beta J_{eff}}}.$$

This has gap $\Omega = J_{eff}$, and it vanishes exponentially as $e^{-\beta J_{eff}}$ as $T \to 0$, shows a Curie-type $1/T$ behavior at large T with a maximum at $T^0 \approx 0.62 J_{eff}/k$. Figure 6 shows χ_{para2} as a function of T.

The pseudogap has been observed in cuprates by using nuclear magnetic resonance (NMR) techniques.[35–39] There is a rapid decline in χ_{para}, as measured by the Knight shift, below T^*, which is far above T_c. A similar downturn is seen in the spin-lattice relaxation rate, $1/T_1$. More interestingly, the higher crossover temperature T^0 was also found in many of these materials (for a review, see Ref. 2). That T^0 has not been seen in some materials is more a matter of not going to high enough temperatures. For example, T^0 was not seen in initial measurements on $YBa_2Cu_4O_8$ up to 400 K. Curro *et al.*[40] extended the measurement to 715 K and found a

broad maximum at $T^0 \sim 500$ K with χ_{para} decreasing slowly on both sides of T^0. In this material T^* is about 240 K and $T_c \sim 81$ K. It is interesting to note that T_0 and T^* vary roughly the same way, decreasing with increasing x, consistent with the theoretical expectation that they are proportional to the same energy scale $J_{eff}(x)$.

Our theory also predicts that χ_{para} would vanish in the normal state as $T \to 0$. This is harder to confirm since superconductivity intervenes. However, as discussed in Ref. 2, in highly underdoped systems such as $Y_1Ba_2Cu_3O_{6.5-6.7}$ (Ref. 35 and 36) and $YBa_2Cu_4O_8$,[41] there is almost no sign of a superconducting transition in the NMR data at T_c, indicating that there is no additional spin pairing at T_c, although χ_{para} continues to decrease. Similar results have also been seen in moderately underdoped $Bi_2Sr_2CaCu_2O_8$ (Bi-2212) (Ref. 36) by Walstedt *et al.* who argue that the lack of any effect at T_c indicates that the gap is unrelated to superconductivity, and represents spin-charge separation.

B. Electron sector: Tunneling DOS

Further support for the theory comes from tunneling experiments, which probe the electron spectra (charge plus spin). Unlike magnetic properties, tunneling density of states (DOS) (Refs. 42 and 43) shows a sizable effect near T_c in underdoped Bi2212. For $T \ll T_c$, there is a gap in the DOS, with well-developed conductance peaks at $\pm\Delta_{sc}$, except that $2\Delta_{sc}/kT_c$ is much larger than the BCS value. As T approaches T_c from below, conductance peaks decrease in height and disappear at T_c, as does the zero-bias peak originating from the Josephson current. This is what is expected. However, while the gap decreases as T_c is approached, it does not close; instead a gaplike structure with a depressed DOS continues to exist above T_c. Although, sometimes it is referred to as the pseudogap in the tunneling literature, the gap actually exists far above T^* (as measured in the NMR experiments), that is, it also exists in the strange phase up to high temperatures, i.e., it is the spin gap. Such a behavior cannot be understood in terms of nn (e.g., slave-boson) states or a Fermi liquid.

The tunneling data are clearly consistent with our theory in which the spin gap exists in all three phases. The gap and the singlet condensation qualitatively account for the NMR data in the normal state. Superconductivity arises from a pairing of spinless holons. Since there is spin-charge separation χ_{para} and $1/T_1$ would largely be unaffected below T_c, as seen in the experiments. However, tunneling experiments measure the DOS of the electron, containing both spin and charge. At the MF level this is a convolution of spinon and holon DOS. Above T_c holons are gapless, so the gap in the electron Green's function reflects the spin gap. This is why tunneling gap exists far above T^*. Below T_c, a gap also opens up in the charge spectrum, as holons pair; this explains the increase in the *total* gap size with decreasing T, and the appearance of the conductance peaks. Also, experimentally the $T=0$ gap increases with decreasing x, mirroring the behavior of T^0 associated with the much larger spin gap, as expected from the theory.

Experimental observation of a spin gap far above T^* rules out a mechanism based on pre-existing electron pairs, carrying both spin and charge. To account for the gap, such pairs should be formed near T^0, and presumably condense below T_c. It is then hard to explain T^*. There are other difficulties. The condensation at T_c must increase the binding energy of the Cooper pairs, enough to cause the large depletion observed in the DOS below T_c, and the observed increase in the size of the gap. However, since electrons forming a Cooper pair also carry spin, such a large increase will be accompanied by a large change in χ_{para} and $1/T_1$ below T_c, a behavior not seen in experiments.[36]

C. Charge sector

Strange insulator. A key prediction of our theory is that holons are confined (nonpropagating) above T^* since $A=0$. Furthermore, since electrons are incoherent and gapped, the "strange metal" is not a metal but a new type of insulator. A strong evidence for this comes from the measured in-plane dc resistivity, ρ. In a metal, ρ is supposed to saturate at the Mott value. Actually, while the linear T dependence of the $\rho(T)$ has received most of the attention, it has been pointed out by many authors that ρ shows no sign of saturation, and in the underdoped regime, far exceeds the Mott value.[44–46] The fact that $\rho(T)$ increases with T does not make the system a metal since ρ also increases in a disorder-induced (Anderson) localized insulator as long as inelastic mean-free path $\ell_{inel}(T)$ is less than the localization length, ξ_{loc}. A quantitative theory is beyond the scope of this paper; but the qualitative behavior can be understood, as follows. In the case of Anderson localization, ξ_{loc} is independent of T; then $\rho(T)$ would show an upturn at low T, as $\ell_{inel(T)}$ exceeds ξ_{loc}. In our case, the appropriate localization length is the distance over which holons can move coherently, which is the singlet-singlet correlation length $\xi_{singlet}(T)$ (not the shorter AF correlation ξ_{spin}). Since $\xi_{singlet}(T)$ increases with decreasing T and becomes infinite at T^* (at the MF level) where ℓ_{inel} is finite. Hence there cannot be a low-T upturn, which is consistent with experiments. On the other hand, since there is no coherent holon motion in the perpendicular direction the c-axis resistivity should show an upturn, as seen.

A clear indication is also obtained from the frequency-dependent conductivity $\sigma(\omega)$ which, for a metal, should exhibit a Drude peak. But in our case there would be no such peak above T^*. Recent experiments in $La_{2-x}Sr_xCuO_4$ (LSCO) (Ref. 46) and Bi-2212 (Ref. 47) show that such a peak does not exist for small x. Instead, one has nearly a flat spectrum over a range of frequencies of order 1 eV. The lack of a Drude peak indicates the absence of coherent charge carriers. The flat spectrum would come from incoherent hopping. At higher doping, a broad peak develops, which is not surprising since in this case the physical electron comes into play as one approaches the Fermi liquid.

Pseudogap state: emergence of metallic conduction. The theory predicts that below T^* holons become coherent and form a spinless Fermi liquid of concentration x. Therefore, $\rho(T)$ should drop rapidly with decreasing T and become metallic. Indeed, experimentally $\rho(T)$ is found to drop faster than linearly[44,48,49] below a temperature which approximately agrees with T^* obtained from NMR experiments.[2,49]

At lower T, $\rho(T)$ becomes metallic and has a residual impurity contribution,[45,46,50] as expected of a metal. At low-T, $\rho(T) \propto T^2$ due to fermion-fermion scattering,[44,51] also as expected. For LSCO, one also finds that ρ does show an upturn at very low T, which is believed to be due to disorder-induced localization.[44]

Recent experiments[51,52] in underdoped LSCO and $Y_1Ba_2Cu_3O_x$ show that the Hall coefficient R_H is independent of T and proportional to $1/x$ in the pseudogap regime, as expected of a metal. This should be compared with the strong T dependence in the "strange" insulator phase. With increasing temperature, $\rho(T)$ is seen to deviate from the T^2 to eventually a linear-T behavior above T^*.

Strong evidence also comes from optical conductivity $\sigma(\omega)$, which shows a Drude peak at low T, with an integrated area (spectral weight) $\propto x$.[52,53] It is characterized by a small plasma frequency, consistent with a small bandwidth, or heavy holons, as in our theory. With increasing T, $\sigma(\omega)$ broadens, but at a rate too large to be attributed to thermal effects in a Fermi liquid, and merges into the incoherent background above T^*.[46,47,52]

These transport properties, taken together, strongly support the view that the charge carriers are holes, rather than electrons, and they form a Fermi liquid below T^*, and when combined with the NMR data, they suggest that these holes are spinless, supporting our view. They also suggest that at small x there are no coherent charge carriers above T^*.

Lack of a Fermi surface

If the pseudogap state is a Fermi liquid of concentration x as the transport experiments suggest, where is the corresponding, presumably, small Fermi surface? This is a real puzzle. Attempts to understand this[51] in terms of the "Fermi arc" found in the photoemission experiments[54,55] does not make sense for many reasons. First, there are no quasiparticles associated with the Fermi arc. The electron Green's function is completely incoherent in the normal state; there are no sharp peaks in the photoemission spectrum. Therefore, the Fermi liquid is not formed by electrons. Furthermore, as discussed earlier, neither the NMR data nor the gap in the tunneling spectra is consistent with an electron Fermi liquid.

In contrast, our theory predicts that the holon Fermi surface is unobservable, even in principle. This is because the holon Green's function, and thus the holon spectrum, are not gauge invariant. Hence, the holon Fermi surface can be deformed and moved around by a gauge transformation. The holon Fermi surface does not exist when averaged over all the gauge-equivalent copies. This is why experiments do not see it.

Another prediction is that since χ_{para} decreases rapidly below T^*, and eventually vanishes as $T \to 0$, the *total* uniform magnetic susceptibility of pseudogap metal would become more diamagnetic with decreasing T due to the orbital motion of holons. The effect should be small near T^* but should become more prominent in the metallic region (at lower T). Unfortunately, superconductivity intervenes. Note that the diamagnetic response discussed here should not be confused with strongly T-dependent response found by Wang *et al.*[56] considerably above T_c. The latter is presumably due to pairing above T_c. Such strong-coupling effects are not ruled out by our theory but its treatment would require more sophisticated techniques. We point out that in the analysis of Wang *et al.* the normal-state diamagnetic contribution appear to have been subtracted out along with other weakly temperature-dependent terms. Therefore more experimental work will be needed to see the effect in the normal state.

In our theory superconductivity appears via a pairing of holons. Since both terms in the effective Hamiltonian arises from hopping, condensation energy is $\propto -t_s^2/t_h$, which causes a reduction in the kinetic energy, as observed.[57,58] Our mean-field treatment implies that spin-charge separation continues to exist in the superconductor. For small x, there is some evidence for this since χ_{para} is essentially unaffected at T_c but more work is needed to clarify this issue.

VIII. ORIGIN OF TWO DIMENSIONALITY

In this section we consider the origin of two dimensionality of the normal state, which is central to the occurrence of high-T_c superconductivity in cuprates. These are highly anisotropic 3D materials. Ordinarily the metallic state would show 3D behavior below some $T_\perp$ since confinement within a plane would cost too much kinetic energy. For cuprates, this would imply $T_\perp < T_c$, which is small at low doping, and actually vanishes at some critical x. In other words, the normal state appears to remain two dimensional as $T \to 0$ (in the absence of superconductivity). On the other hand, the other phases in cuprates: the AF state at half filling, the superconducting state, and the Fermi liquid state at large x are all three dimensional, as one would expect.

In theoretical studies, the 2D nature of the normal state is usually *assumed*, not established. One simply analyzes the 2D Hamiltonian. The question arises: does the corresponding normal state remain two dimensional when hopping is turned on in the perpendicular direction? If it does not, the theory should be discarded. This requirement, seldom tested, provides a powerful constraint on any theory.

Out-of-plane hopping can be modeled by adding

$$-t_\perp \sum_{i,z} [c^\dagger_{i\sigma}(z) c_{i\sigma}(z+1) + \text{H.c.}]$$

to the original model, where $t_\perp \ll t$; and for simplicity we consider a tetragonal lattice. Suppose we put $U=0$ and treat the anisotropic hopping model as an effective model for the Fermi liquid state. Obviously, this would show 3D behavior. The same is true if the particles are bosons, as in a superconductor or a quantum magnet. Similarly, the nn (e.g., the spiral or the slave-boson) states should be three dimensional since they are modeled by similar effective MF Hamiltonians, in which fermions and bosons hop "freely" in all three directions. The issue of confinement has been studied for coupled one-dimensional Luttinger chains, which have gapless excitations.[59]

In our theory, the gapping of the spin excitations provides a protective mechanism for 2D confinement. This is far from obvious since, as we have seen, holons can delocalize in the plane even though spinons are gapped. So why not in the z direction? First, note that a nonzero $t_\perp$ leads to an exchange

interaction $J_\perp = 4t_\perp^2/U = J(t_\perp/t)^2 \ll J$. Since a spin can belong to only one singlet at a given time it follows that the spin singlets are formed only on the x-y plane in the ground state. In order to form them along the z direction one has to break singlets in the plane, which will cost an energy of order J-$J_\perp$ per singlet, *as long as spin excitations are gapped*.

Now, suppose a hole (or a projected electron) hops on to the adjacent layer, breaking a singlet, which creates two unpaired spinons, one in each plane. The cost in energy is $\sim\Omega$, which cannot be removed by the hole hopping on to a third layer since that would create two more unpaired spinons costing more energy. A singlet can be reconstructed between the two unpaired spins in the intermediate layer if they are in opposite direction, which happens half the time. But that still leaves two unpaired spins. The other 50% of the time there will be four unpaired spins. Thus, broken singlets proliferate as the hole hops farther and farther in the z direction.

Now, there are only two processes by which the system can relax. The hole can come back to the original layer, which means that the normal state is two dimensional. The second one is the two-hole process similar to the one discussed earlier, that is, a second hole follows the first. As in the plane, the pair hopping is accompanied by a singlet backflow. Therefore superconductivity is three dimensional, which explains the 3D enhancement of T_c.

Although this interlayer *holon pair* hopping is nominally similar to the interlayer pair tunneling (ILT) mechanism,[60] the physics is quite different. In the ILT theory, electron pairs tunnel. It plays the primary role; there is no *intralayer* pair hopping. The spinon spectrum is gapless, the interlayer hopping matrix is diagonal in k_{xy}, i.e., long ranged. In our case, the hopping is localized in real space. The primary mechanism for superconductivity is the intralayer holon pair hopping, which contributes the main part of the condensation energy. The interlayer hopping process makes superconductivity three dimensional and enhances T_c.

Superconductivity in the 3D system. The second process described above contributes the following interplane pair hopping term

$$H_{inter} = -t_{zs}\sum_{ij,z}[\mathrm{F}_{ij}^\dagger(z)\mathrm{F}_{ij}(z+1) + \mathrm{H.c.}], \tag{30}$$

where $t_{zs} > 0$ is an effective pair hopping amplitude which is ($\ll t_s$), the intraplane pair hopping amplitude. In this paper we will treat it as free parameter. Note that the vector fields $F_{ij} = \langle h_j h_i \rangle$ lie on the plane. Furthermore, this term does not modify the normal holon spectrum at the Hartree-Fock level.

The MF problem for the superconducting state is again reduced to a 2D problem with an additional pairing term. The equation for the order parameter (with d-wave symmetry) now becomes

$$\frac{1}{t_s + t_{zs}} = \frac{1}{N}\sum_{\mathbf{k}} (\sin k_x - \sin k_y)^2 \frac{\tanh[\beta E(\mathbf{k})/2]}{E(\mathbf{k})}, \tag{31}$$

where, as before,

$$E(\mathbf{k}) = [\xi^2(\mathbf{k}) + |\Delta_h(\mathbf{k})|^2]^{1/2}$$

but with a modified gap function

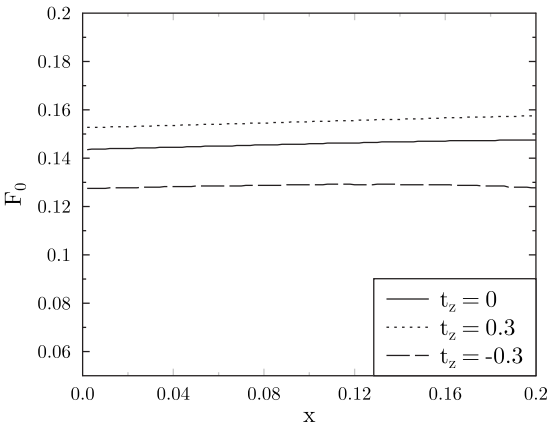

FIG. 7. 3D enhancement of superconducting condensation energy: order parameter vs interplane pair hopping amplitude t_z, which is the same as t_{zs} in the text. It is enhanced in the actual case (dotted line). However, it is reduced if the sign of the interaction is reversed by hand (bottom line).

$$\Delta_h(\mathbf{k}) = 4(t_s + t_{zs})F_0(\sin k_x - \sin k_y). \tag{32}$$

The equation for x is unaffected. As shown in Fig. 7, F_0 increases with t_{zs} for given values of t_h, the effect increasing slowly with increasing x. Note that the sign of t_{zs} is determined by the symmetry of the underlying RVB state, and is of considerable importance since, as seen from the figure, F_0 actually decreases if the sign is reversed by hand (bottom line).

IX. ELECTRON DOPING VS HOLE DOPING

High-T_c superconductivity also occurs in electron-doped systems. This is expected theoretically since the square-lattice Hubbard (or t-J) model has an electron-hole symmetry about half filling. The symmetry is realized by the transformation: $c_{i\sigma} \to c_{i\sigma}^\dagger$ on one sublattice and $c_{i\sigma} \to -c_{i\sigma}^\dagger$ on the other. If we assume that the orbital structure in the CuO_2 plane is roughly the same for the two cases, an electron-doped system at an electron concentration $1+x$ should exhibit roughly the same behavior as a hole-doped system at a concentration $1-x$. Experimentally this is not found to be the case. In the electron-doped system the superconducting state occupies a much smaller region in the parameter space and T_c's are also much smaller. The AF insulator phase occupies a much larger region.

In a real system there will always be an asymmetry since one has to include an intrasublattice hopping term in the original model,

$$H' = t'\sum_{il} c_{i\sigma}^\dagger c_{l\sigma} + \mathrm{H.c.}, \tag{33}$$

where i and l are next-nearest neighbors, and this term changes sign under the e-h transformation. A rough estimate from angle-resolved-photoemission data and electronic-structure calculations is: $t'/t \approx (0.1-0.3)$.[61] The t' term would generate a $J' = 4t'^2/U$, which is very small since $J'/J = (t'/t)^2$, and can be neglected. In general, t' should not change the physics qualitatively; and for a Fermi liquid or a

nn state the quantitative difference should not be too large since physics is dominated by the t term. The observed difference is much larger than what is expected.

In our case the difference can be significant since t is renormalized away, and the hopping parameter t_h that characterizes renormalized intrasublattice hopping [see Eqs. (17) and (18)] is only $\sim J$. It should be noted this term *does not* change under the e-h transformation. Therefore, the effect of t' is to increase the net intrasublattice hopping amplitude in one case and decrease it in the other. Of course, t' will also be renormalized to t'_{eff}, but renormalization effect will be smaller since this term moves an electron within the same sublattice, which has ferromagnetic correlations. Now, in the MF approximation H' becomes

$$H' = -t'_{eff}\sum_{ij}[2B_{ij}h_j^\dagger h_i + D_{ij}\sum_\sigma b_{i\sigma}^\dagger b_{i\sigma} + \text{H.c.}], \quad (34)$$

which generates intrasublattice holon hopping. Here $B_{ij}=\langle b_{i\sigma}^\dagger b_{j\sigma}\rangle$ and $D_{ij}=\langle h_j^\dagger h_i\rangle$ are the average hopping amplitudes for spinons and holons along the diagonal. Above T^* this term does not contribute since $B_{ij}=0=D_{ij}$. We assume that t'_{eff} is small so that it does not by itself break gauge symmetry and generate coherent hopping of holons above T^*.

The situation is quite different below T^* since the diagonal holon hopping amplitude $D_{ij}\neq 0$ in the pseudogap metal; and the diagonal B_{ij} is also nonzero in the presence of the singlet condensate. The symmetry of B_{ij}^* is easily calculated from the MF theory, and we find $B_{ij}^*=\mp B_2$, along the diagonal $(1,\pm 1)$ directions, where

$$B_2 = \frac{1}{N}\sum_{\mathbf{k}} \sin k_x \sin k_y \frac{\lambda}{\omega(\mathbf{k})}\{1/2 + n[\omega(\mathbf{k})]\}. \quad (35)$$

Since the spinon spectrum has minima at $\mathbf{k}=(\pi/2,\pi/2)$, $B_2>0$. Thus the t' term contributes

$$\epsilon'_h(\mathbf{k}) = -4t'_h \sin k_x \sin k_y \quad (36)$$

to the holon spectrum, where $t'_h=t'_{eff}B_2$. Therefore, the extra holon hopping term in the diagonal direction has the same symmetry as the earlier one; except that the sign is opposite for the hole-doped case for which, $t'_h>0$. As shown in Fig. 8 (upper panel), the extra term has maxima at $\pm(\pi/2,-\pi/2)$, where the original spectrum is a minimum, making the band shallower and holons heavier (and increases the holon DOS). Since t' term does not contribute to the pairing interaction, the net effect is to reduce pair breaking, which increases F_0 and T_c. This is seen in Fig. 9 where we plot F_0 vs x at $T=0$.

The problem of an electron-doped system at an electron density of $1+x$ is the same as that of a hole-doped system at a hole density of x, except that the sign of t', (and hence, that of t'_h) is now changed. In this case, the total holon hopping amplitude along the diagonal add. The extra term has minima at the minima of the original spectrum (Fig. 8, lower panel), making the band steeper, which making holons lighter and thus decreases the DOS. As shown in Fig. 9, F_0 is now smaller, as one would expect. We stress that these effects are amplified because the pairing interaction and holon hopping are of the same order of magnitude.

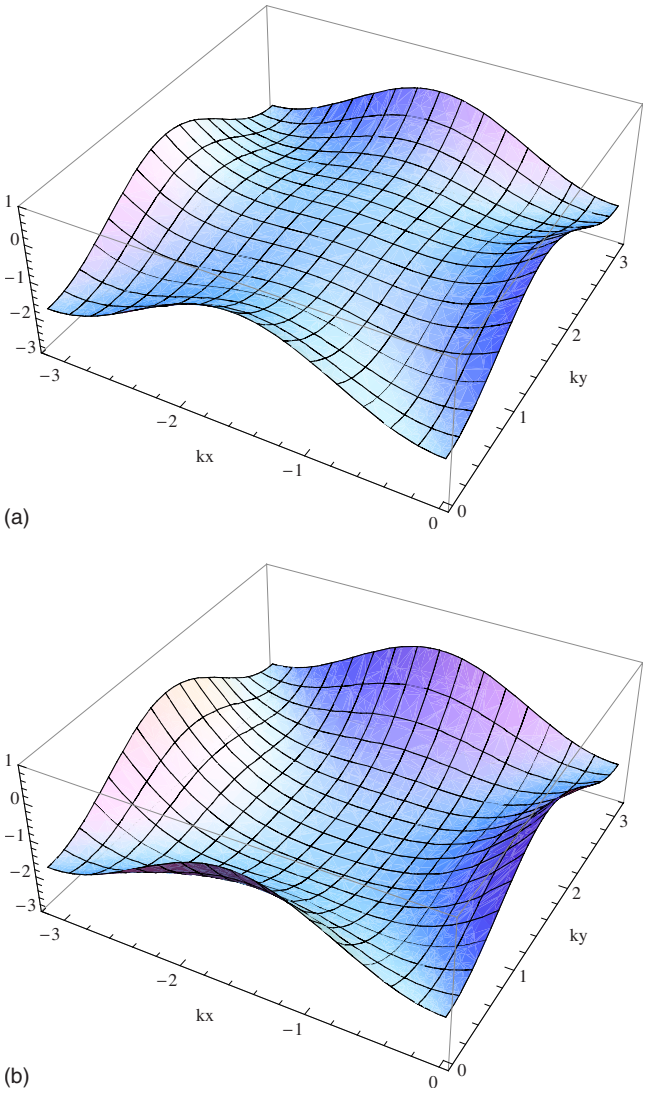

FIG. 8. (Color online) The origin of electron-hole asymmetry: the holon spectrum in the second quadrant in the presence of next-nearest-neighbor hopping $t'_{eff}=\pm 0.2t_{eff}$. For positive t'_{eff} (hole-doped case) the minimum becomes shallower (upper panel), which would to lower Fermi velocity and higher F_0. For the electron-doped case, t'_{eff} is negative, and minimum deepens (lower panel), which would lead to higher Fermi velocity and lower F_0.

Interestingly the theory also provides a straightforward explanation for why the size of the AF region is larger for the electron-doped case simply on the basis of the symmetry of the RVB state. The point is that average spinon hopping amplitude B_{ij} measures *ferromagnetic* correlations between sites i and j. It is nonzero when i and j are on the *same* sublattice, and has the largest value when the system has long-range AF order. This can be verified from Eq. (35) which shows that right-hand side is largest when the spinon energies $\omega(\mathbf{k})$ are smallest, i.e., when the spectrum is gapless (corresponding to long-range AF order). We also see from Eqs. (20) and (21) that $D_{ij}=\pm D_1$ with $D_1>0$ along the $(1,\pm 1)$ directions. Hence, the contribution of the t' term to the energy per bond is $t'_{eff}B_2D_1$ which has the same sign as t'. It follows that AF ordering will be opposed by this term for the hole-doped

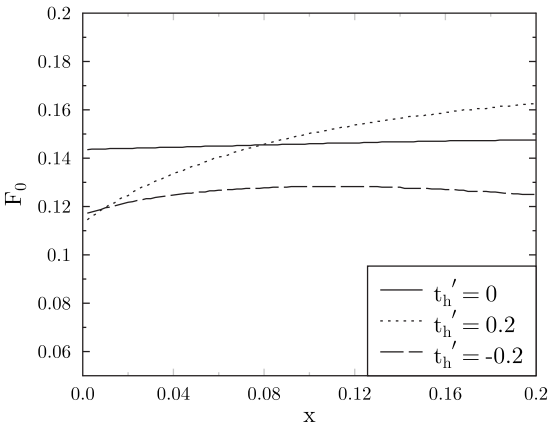

FIG. 9. Electron-hole asymmetry: the condensation energy vs the effective diagonal hopping amplitude t_h' for fixed x. The hole-doped case (dotted line) corresponds to $t_h' > 0$ and the electron-doped case (broken line) corresponds to $t_h' < 0$.

case ($t' > 0$) and will be favored for the electron-doped case ($t' < 0$).

X. CONCLUSION

In this paper, we have shown that assumptions of continuity from half filling and renormalization give rise to the number and type of phases that are seen in the experimental phase diagram. The theory makes three major predictions: a spin gap which plays a pivotal role in all three phases; a spinless Fermi liquid below T^* but one without a Fermi surface; and a state above T^* which is not a metal but a novel quantum insulator. Except for superconductivity, the motion of holons do not introduce any new order, other than those at half filling. The normal states are predicted to be rather different when compared with other theories. We have presented a careful review of the old and new experimental results in the charge, spin, and electron sectors, and found very strong evidence in support of these predictions.

Superconductivity also occurs naturally, via pair hopping of holons, driven by the existence of previously paired spinons. It is found that the possible symmetries of the superconducting order parameter are severely restricted because the symmetry of the spin part of the electron pair is already determined at half filling. We have shown that this leads to a robust d wave. Remarkably, the theory provides a natural explanation for the two dimensionality of the normal state and also of the pronounced difference between hole-doped and electron-doped cuprates. Most of results follow from the type and symmetry of the renormalized Hamiltonians, detailed calculations are not necessary.

The theory provides a basis for future calculations of finite-temperature properties, which would require careful treatment of fluctuations, including gauge fluctuations, separately for each phase. In particular, because of the low dimensionality, and since the energy scales of all the terms in the renormalized Hamiltonian scale with J and are comparable, the mean-field approximation would not work near the transitions. For this reason, probing the nature of the transition near T^*, and the calculation of $T_c(x)$ or, possibility of a Nernst effect, would require a more sophisticated treatment. We also do not discuss the issue of collective modes, the treatment of which would require going beyond the MF approximations. There are a number of such collective excitations. An important one is the physical electron itself, which can be probed by the photoemission spectrum. Experimentally one sees something like a Fermi surface above T^*, and a Fermi arc in the pseudogap normal state, although there are no quasiparticle excitations associated with these. One important issue is the possible appearance of sharp peaks in the gap region of the superconducting state. Our theory does not rule out the emergence of a more conventional Fermi-liquid state at large x, where the physical electron would appear as a collective excitation via spin-charge recombination.[21] These and other issues would be discussed in a future paper.

[1] J. G. Bednorz and K. A. Muller, Z. Phys. **64**, 189 (1986).
[2] T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).
[3] P. W. Anderson, Science **235**, 1196 (1987).
[4] G. Baskaran and P. W. Anderson, Phys. Rev. B **37**, 580 (1988).
[5] D. P. Arovas and A. Auerbach, Phys. Rev. B **38**, 316 (1988).
[6] S. Sarker, C. Jayaprakash, H. R. Krishnamurthy, and M. Ma, Phys. Rev. B **40**, 5028 (1989).
[7] Y.-C. Chen and K. Xiu, Phys. Lett. A **181**, 373 (1993); Y.-C. Chen, Mod. Phys. Lett. B **8**, 1253 (1994); E. Manousakis, Rev. Mod. Phys. **63**, 1 (1991).
[8] C. Jayaprakash, H. R. Krishnamurthy, and S. K. Sarker, Phys. Rev. B **40**, 2610 (1989); D. Yoshioka, J. Phys. Soc. Jpn. **58**, 1516 (1989).
[9] Z. Y. Weng, D. N. Sheng, and C. S. Ting, Phys. Rev. Lett. **80**, 5401 (1998). These authors connect with half filling by using a representation in which both holons and spinons are bosons. But they have not shown their states are stable relative to spiral states.
[10] G. Baskaran, Z. Zou, and P. W. Anderson, Solid State Commun. **63**, 973 (1987).
[11] G. Kotliar and J. Liu, Phys. Rev. B **38**, 5142 (1988).
[12] I. Ichinose, T. Matsui, and M. Onoda, Phys. Rev. B **64**, 104516 (2001).
[13] P. A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78**, 17 (2006).
[14] F. M. Hu, S. K. Sarker, and C. Jayaprakash, Phys. Rev. B **50**, 17901 (1994).
[15] P. B. Wiegmann, Phys. Rev. Lett. **60**, 821 (1988).
[16] P. A. Lee, Phys. Rev. Lett. **63**, 680 (1989); X.-G. Wen, Phys. Rev. B **39**, 7223 (1989).
[17] S. K. Sarker, Phys. Rev. B **77**, 052505 (2008).
[18] N. Read and S. Sachdev, Phys. Rev. Lett. **62**, 1694 (1989); Phys. Rev. B **42**, 4568 (1990).

[19] S. A. Kivelson, D. S. Rokhsar, and J. P. Sethna, Phys. Rev. B **35**, 8865 (1987).
[20] D. S. Rokhsar and S. A. Kivelson, Phys. Rev. Lett. **61**, 2376 (1988).
[21] S. K. Sarker, Phys. Rev. B **46**, 8617 (1992).
[22] S. K. Sarker, Phys. Rev. B **61**, 8663 (2000).
[23] S. Elitzur, Phys. Rev. D **12**, 3978 (1975).
[24] L. B. Ioffe and A. I. Larkin, Phys. Rev. B **39**, 8988 (1989).
[25] C. L. Kane, P. A. Lee, and N. Read, Phys. Rev. B **39**, 6880 (1989).
[26] S. Schmitt-Rink, C. M. Varma, and A. E. Ruckenstein, Phys. Rev. Lett. **60**, 2793 (1988).
[27] D. Poilblanc, T. Ziman, H. J. Schulz, and E. Dagotto, Phys. Rev. B **47** 14267 (1993); for a review see, E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[28] S. K. Sarker, Phys. Rev. B **47**, 2940 (1993).
[29] F. D. M. Haldane and H. Levine, Phys. Rev. B **40**, 7340 (1989).
[30] N. Read and B. Chakraborty, Phys. Rev. B **40**, 7133 (1989).
[31] S. K. Sarker, Phys. Rev. B **68**, 085105 (2003).
[32] D. Poilblanc and T. M. Rice, Phys. Rev. B **39**, 9749 (1989); H. J. Schulz, Phys. Rev. Lett. **64**, 1445 (1990); J. Phys. (Paris) **50**, 2833 (1989); J. Zaanen and O. Gunnarsson, Phys. Rev. B **40**, 7391 (1989); M. Kato, K. Machida, H. Nakanishi, and M. Fujita, J. Phys. Soc. Jpn. **59**, 1047 (1990).
[33] P. B. Visscher, Phys. Rev. B **10**, 943 (1974).
[34] S. K. Sarker, C. Jayaprakash, and H. R. Krishnamurthy, Physica C **228**, 309 (1994).
[35] W. W. Warren, R. E. Walstedt, G. F. Brennert, R. J. Cava, R. Tycko, R. F. Bell, and G. Dabbagh, Phys. Rev. Lett. **62**, 1193 (1989).
[36] R. E. Walstedt, R. F. Bell, and D. B. Mitzi, Phys. Rev. B **44**, 7760 (1991).
[37] M. Bankay, M. Mali, J. Roos, and D. Brinkmann, Phys. Rev. B **50**, 6416 (1994).
[38] R. Stern, M. Mali, I. Mangelschots, J. Roos, D. Brinkmann, J.-Y. Genoud, T. Graf, and J. Muller, Phys. Rev. B **50**, 426 (1994).
[39] M. H. Julien, P. Carretta, M. Horvatić, C. Berthier, Y. Berthier, P. Ségransan, A. Carrington, and D. Colson, Phys. Rev. Lett. **76**, 4238 (1996); J. Bobroff, H. Alloul, P. Mendels, V. Viallet, J.-F. Marucco, and D. Colson, *ibid.* **78**, 3757 (1997).
[40] N. J. Curro, T. Imai, C. P. Slichter, and B. Dabrowski, Phys. Rev. B **56**, 877 (1997).
[41] H. Zimmermann, M. Mali, I. Mangelschots, J. Roos, L. Pauli, D. Brinkmann, J. Karpinski, S. Rusiecki, and E. Kaldis, J. Less-Common Met. **164-165**, 138 (1990).
[42] Ch. Renner, B. Revaz, J.-Y. Genoud, K. Kadowaki, and O. Fischer, Phys. Rev. Lett. **80**, 149 (1998).
[43] N. Miyakawa, J. F. Zasadzinski, L. Ozyuzer, P. Guptasarma, D. G. Hinks, C. Kendziora, and K. E. Gray, Phys. Rev. Lett. **83**, 1018 (1999); N. Miyakawa, P. Guptasarma, J. F. Zasadzinski, D. G. Hinks, and K. E. Gray, *ibid.* **80**, 157 (1998).
[44] H. Takagi, B. Batlogg, H. L. Kao, J. Kwo, R. J. Cava, J. J. Krajewski, and W. F. Peck, Jr., Phys. Rev. Lett. **69**, 2975 (1992).
[45] K. Segawa and Y. Ando, Phys. Rev. Lett. **86**, 4907 (2001).
[46] K. Takenaka, J. Nohara, R. Shiozaki, and S. Sugai, Phys. Rev. B **68**, 134501 (2003).
[47] A. F. Santander-Syro, R. P. S. M. Lobo, N. Bontemps, Z. Konstantinovic, Z. Z. Li, and H. Raffy, Phys. Rev. Lett. **88**, 097005 (2002).
[48] B. Bucher, P. Steiner, J. Karpinski, E. Kaldis, and P. Wachter, Phys. Rev. Lett. **70**, 2012 (1993).
[49] B. Batlogg, H. Y. Hwang, H. Takagi, R. J. Cava, H. L. Kao, and J. Kwo, Physica C **235-240**, 130 (1994).
[50] T. Ito, K. Takenaka, and S. Uchida, Phys. Rev. Lett. **70**, 3995 (1993).
[51] Y. Ando, Y. Kurita, S. Komiya, S. Ono, and K. Segawa, Phys. Rev. Lett. **92**, 197001 (2004).
[52] W. J. Padilla, Y. S. Lee, M. Dumm, G. Blumberg, S. Ono, K. Segawa, S. Komiya, Y. Ando, and D. N. Basov, Phys. Rev. B **72**, 060511(R) (2005).
[53] J. Orenstein, G. A. Thomas, A. J. Millis, S. L. Cooper, D. H. Rapkine, T. Timusk, L. F. Schneemeyer, and J. V. Waszczak, Phys. Rev. B **42**, 6342 (1990); S. Uchida, T. Ido, H. Takagi, T. Arima, Y. Tokura, and S. Tajima, *ibid.* **43**, 7942 (1991); S. L. Cooper, D. Reznik, A. Kotz, M. A. Karlow, R. Liu, M. V. Klein, W. C. Lee, J. Giapintzakis, D. M. Ginsberg, B. W. Veal, and A. P. Paulikas, *ibid.* **47**, 8233 (1993).
[54] A. Ino, C. Kim, M. Nakamura, T. Yoshida, T. Mizokawa, A. Fujimori, Z.-X. Shen, T. Kakeshita, H. Eisaki, and S. Uchida, Phys. Rev. B **65**, 094504 (2002).
[55] T. Yoshida, X. J. Zhou, T. Sasagawa, W. L. Yang, P. V. Bogdanov, A. Lanzara, Z. Hussain, T. Mizokawa, A. Fujimori, H. Eisaki, Z.-X. Shen, T. Kakeshita, and S. Uchida, Phys. Rev. Lett. **91**, 027001 (2003).
[56] Y. Wang, L. Li, M. J. Naughton, G. D. Gu, S. Uchida, and N. P. Ong, Phys. Rev. Lett. **95**, 247002 (2005).
[57] G. Deutscher, A. F. Santander-Syro, and N. Bontemps, Phys. Rev. B **72**, 092504 (2005).
[58] H. J. A. Molegraaf, C. Presura, D. van der Marel, P. H. Kes, and M. Li, Science **295**, 2239 (2002); A. B. Kuzmenko, H. J. A. Molegraaf, F. Carbone, and D. van der Marel, Phys. Rev. B **72**, 144503 (2005).
[59] D. G. Clarke, S. P. Strong, and P. W. Anderson, Phys. Rev. Lett. **72**, 3218 (1994).
[60] S. Chakravarty, A. Sudbø, P. W. Anderson, and S. Strong, Science **261**, 337 (1993).
[61] E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, and O. K. Andersen, Phys. Rev. Lett. **87**, 047003 (2001).