From the pseudogap metal to the Fermi liquid using ancilla qubits

Ya-Hui Zhang and Subir Sachdev

Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

We propose a new parton theory of the hole-doped cuprates, describing the evolution from the pseudogap metal with small Fermi surfaces to the conventional Fermi liquid with a large Fermi surface. We introduce two ancilla qubits per square lattice site, and employ them to obtain a variational wave function of a fractionalized Fermi liquid for the pseudogap metal state. We propose a multilayer Hamiltonian for the cuprates, with the electrons residing in the physical layer, and the ancilla qubits in two hidden layers: the hidden layers can be decoupled from the physical layer by a canonical transformation, which leaves the hidden layers in a trivial gapped state. This Hamiltonian yields an emergent gauge theory, which describes not only the fractionalized Fermi liquid, but also the conventional Fermi liquid, and possible exotic intermediate phases and critical points. The fractionalized Fermi liquid has hole pockets with quasiparticle weight which is large only on Fermi arcs, and fermionic spinon excitations, which carry charges of the emergent gauge fields.

I. INTRODUCTION

The structure of the pseudogap metal state in the hole-doped cuprate superconductors has long been the focus of much theoretical and experimental attention [1]. In more recent experiments, it has become clear that the pseudogap state is present only for a hole doping p smaller than a critical value p_c [2–14]. For $p > p_c$, many observables indicate the presence of a conventional Fermi liquid (FL) state, with a large Fermi surface enclosing a volume associated with a hole density $1 + p$. While there have been many theoretical proposals for the pseudogap metal, there is as yet no framework that can capture the essential physics of both the pseudogap metal and the conventional Fermi liquid as different mean-field solutions of the same theory. Such a framework is surely needed as a starting point for a theory of the mysterious strange metal found near $p = p_c$.

We present such a framework here. We show that the introduction of two ancilla qubits per square lattice site leads to a valuable flexibility in describing possible correlated states of mobile electrons on the square lattice. It should be noted that the ancilla qubits are not physical degrees of freedom that can be directly observed; rather, they are theoretical tools, which enable us to capture new varieties of entangled states of the electrons.

We will describe the pseudogap metal as a fractionalized Fermi liquid (FL*) state [15]. This state has electronlike quasiparticles around pocket Fermi surfaces enclosing a volume associated with hole density p. Such small pocket Fermi

surfaces can appear even in the absence of any translational symmetry breaking by charge or spin density wave order. But compatibility with the Luttinger constraint requires that there be additional fractionalized spinon excitations carrying charges of an emergent gauge field [15–17]. Such a FL* state is compatible with many of the experimental observations noted above for $p < p_c$, and several theoretical descriptions have been proposed [18–26]. However, these theories do not describe the termination of the FL* state followed by the appearance of a FL state. A $SU(2)$ gauge theory [27–29] has been proposed to described optimal doping criticality in the cuprates, but this connects naturally to an algebraic charge liquid [30] description of the pseudogap metal, in which the Fermi pockets are initially of spinless fermionic chargons, which have to bind with spinons to obtain the electronlike Fermi surfaces of FL* [31].

Our approach with ancilla qubits is illustrated in Fig. 1. We are interested in the electrons, $c_{i\sigma}$, in the physical layer, where i labels a square lattice site, and $\sigma = \uparrow, \downarrow$ is the electron spin. We use the ancilla qubits in the hidden layers to generate wave functions and field theories for observables in the physical layer. The spin operators $\boldsymbol{S}_{i;1}$, $\boldsymbol{S}_{i;2}$ act on the qubits in the two layers. It is convenient (but not required) to represent the spins by fermions $f_{i;1\sigma}$, $f_{i;2\sigma}$ using

$$\boldsymbol{S}_{i;1} = \tfrac{1}{2} f_{i;1\sigma}^\dagger \boldsymbol{\sigma}_{\sigma\sigma'} f_{i;1\sigma'}, \quad \boldsymbol{S}_{i;2} = \tfrac{1}{2} f_{i;2\sigma}^\dagger \boldsymbol{\sigma}_{\sigma\sigma'} f_{i;2\sigma'} \quad (1)$$

where $\boldsymbol{\sigma}$ are the Pauli matrices, and there must be exactly one fermion on each hidden layer site

$$\sum_\sigma f_{i;1\sigma}^\dagger f_{i;1\sigma} = 1, \quad \sum_\sigma f_{i;2\sigma}^\dagger f_{i;2\sigma} = 1. \quad (2)$$

We can now introduce our trial wave functions for the electrons in the physical layer. We propose a trial Slater determinant state $|\text{Slater}[c, f_1, f_2]\rangle$ for c, f_1 and f_2 fermions. Then we project out components of this wave function in
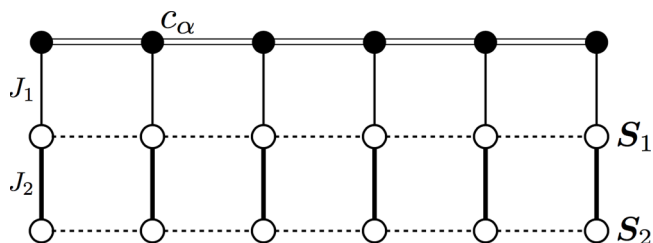
FIG. 1. A Hubbard model of mobile electrons $c_{i\sigma}$ of variable density p in the physical square-lattice layer, coupled to two hidden square-lattice layers of ancilla qubits (spin-1/2 spins) $\boldsymbol{S}_{i;1}$ and $\boldsymbol{S}_{i;2}$. The lattice sites are labeled by i. We develop a theory with the exchange interactions $J_{1,2}$ finite, and take the limit $J_2 \to \infty$ of infinite antiferromagnetic exchange between the hidden layers at the end.

which the ancilla qubits are locked in a spin singlet on each site

$$|s\rangle = \prod_i \frac{1}{\sqrt{2}} (f^\dagger_{i;1\uparrow} f^\dagger_{i;2\downarrow} - f^\dagger_{i;1\downarrow} f^\dagger_{i;2\uparrow})|0\rangle . \quad (3)$$

The wave function on the physical layer is

$$|\Phi[c]\rangle = \sum_a |a\rangle \langle a, s| \operatorname{Slater}[c, f_1, f_2]\rangle, \quad (4)$$

where $|a\rangle$ is a basis of states in the physical layer. Different choices for $|\operatorname{Slater}[c, f_1, f_2]\rangle$ will lead to different physical states $|\Phi[c]\rangle$. A FL* state is obtained by having the c and f_1 fermions occupy linear combinations of states between the physical layer and the first hidden layer, while the f_2 occupy states on the second layer. The total fermion density of the c and f_1 layers is $2 - p$, and so their band structure can exhibit hole pockets of volume p in the conventional Luttinger count (FL* states using an auxiliary band with total fermion density $2 - p$ have also been discussed earlier [20,32]). Concomitantly the f_2 layer realizes the needed fractionalized spinons of the FL* state. In the FL state, we choose $|\operatorname{Slater}[c, f_1, f_2]\rangle$ with decoupled the physical and hidden layers. Then the c layer, with fermion density $1 - p$ will have a large holelike Fermi surface of volume $1 + p$, and the hidden layers will form a trivial gapped insulator.

For analytic progress, we need the Hamiltonian of Fig. 1, from which we can derive gauge theories. We consider the Hamiltonian

$$H = H_U + H_a, \quad (5)$$

where H_U is a Hubbard model of electrons $c_{i\sigma}$ on the sites i of a square lattice

$$H_U = -\sum_{i,j} t_{ij} c^\dagger_{i\sigma} c_{j\sigma} - \mu \sum_i c^\dagger_{i\sigma} c_{i\sigma} + U \sum_i c^\dagger_{i\uparrow} c_{i\uparrow} c^\dagger_{i\downarrow} c_{i\downarrow} \quad (6)$$

and H_a describes two hidden layers of ancilla spin $S = 1/2$ qubits $\boldsymbol{S}_{i;1}$, $\boldsymbol{S}_{i;2}$ (see Fig. 1)

$$H_a = \frac{J_1}{2} \sum_i c^\dagger_{i\sigma} \boldsymbol{\sigma}_{\sigma\sigma'} c_{i\sigma'} \cdot \boldsymbol{S}_{i;1} + J_2 \sum_i \boldsymbol{S}_{i;1} \cdot \boldsymbol{S}_{i;2}$$
$$+ H_1(\boldsymbol{S}_{i;1}) + H_2(\boldsymbol{S}_{i;2}), \quad (7)$$

where H_1 represents exchange interactions between the first hidden layer qubits $\boldsymbol{S}_{i;1}$ (which we do not specify), and similarly for H_2. When J_2 is large, the two ancilla layers will form a spin gap state; so we can safely integrate out the ancilla qubits, and induce near-neighbor exchange interactions between the $c_{i\sigma}$ electrons on the physical layer. Alternatively stated, a canonical transformation decouples the physical and hidden layers at the cost of additional exchange interactions in the physical layer. Rather than accounting for these exchange interactions explicitly, more progress is possible by keeping the hidden layers alive, and considering possible states and gauge theories in the expanded Hilbert space. In the end we can take the $J_2 \to \infty$ limit [corresponding to projection onto the singlet hidden layer state $|s\rangle$ in (3)], which reduces the above model to the standard Hubbard model for the cuprates. A similar approach using auxiliary degrees of freedom has been taken by Refs. [33,34] for a description of the bosonic composite Fermi liquid in the lowest Landau level.

The outline of the rest of the paper is follows. We present a gauge theory description of phases of H in Sec. II. This leads to a mean-field description of the FL* and FL phases. The physical properties of the FL* phase, including its photoemission spectrum, are described in Sec. III. The critical region between the FL* and FL phases is discussed in Sec. IV, including a possible intermediate phase, or a direct transition. Some directions for future research are noted in Sec. V, and we summarize in Sec. VI.

II. GAUGE STRUCTURE AND MEAN-FIELD THEORY

A. $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ gauge structure

The spins in the hidden layers can be represented by the standard fermionic partons [1] in (1). Naively we can just form mean-field theories using c, f_1, f_2. However, this kind of analysis does not incorporate the large J_2. For example, let us consider mean-field ansatz for which c decouples from f_1, f_2; then f_1, f_2 can form gapless spin liquids. However, in the large J_2 region, the spin carried by f_1 and f_2 must be gapped. This is similar to a Mott gap in spin channel. In the familiar Hubbard model, we use slave boson theory to describe the Mott transition and incorporate the charge gap at large U. Here we can use a similar slave spin approach to incorporate the spin gap at large J_2.

Therefore, we perform a further fractionalization of f_1 and f_1 in (1):

$$f_{i;a\sigma} = R_{i;\sigma\tilde\sigma} \tilde f_{i;a\tilde\sigma}, \quad (8)$$

where $a = 1, 2$ and the slave spin R is a $SU(2)$ matrix, similar to that introduced in Ref. 27. Basically this means the spin index $\tilde\sigma$ carried by $\tilde f$ can be freely rotated by a $SU(2)$ gauge transformation:

$$\begin{pmatrix} \tilde f_{i;a\uparrow} \\ \tilde f_{i;a\downarrow} \end{pmatrix} \to U_{i;S} \begin{pmatrix} \tilde f_{i;a\uparrow} \\ \tilde f_{i;a\downarrow} \end{pmatrix}, \quad (9)$$

where $U_{i;S} \in SU(2)$. Accordingly the slave spin transforms as $R_i \to R_i U^\dagger_{i;S}$. We label this gauge transformation as $SU(2)_S$. See also Appendix for further discussion on the origin of this expanded gauge structure.

It is well known [1] that the parton representation for each layer in (1) has another $SU(2)$ gauge transformation in the

particle-hole channel:

$$\begin{pmatrix} \tilde{f}_{i;a\uparrow} \\ \tilde{f}^\dagger_{i;a\downarrow} \end{pmatrix} \rightarrow U_{i;a} \begin{pmatrix} \tilde{f}_{i;a\uparrow} \\ \tilde{f}^\dagger_{i;a\downarrow} \end{pmatrix}, \tag{10}$$

where $a = 1, 2$ and $U_{i;a} \in SU(2)$. The $U_{i;1}$ and $U_{i;2}$ are two independent gauge transformations for $\tilde{f}_1$ and $\tilde{f}_2$, respectively, which commute with $U_{i,S}$ [35]. The slave spin R_i remains unchanged under both $U_{i;1}$ and $U_{i;2}$. We label these two gauge transformations as $SU(2)_1$ and $SU(2)_2$, respectively. In total the gauge structure of our parton theory in terms of $\tilde{f}_{i;a\sigma}$ is $(SU(2)_1 \times SU(2)_2 \times SU(2)_S)/Z_2$. Here we need to mod out the transformation: $\tilde{f}_{i;a\sigma} \rightarrow -\tilde{f}_{i;a\sigma}$. The combined transformations in (9) and (10) are similar to the O(4) fractionalization of Ref. [35], but the $SU(2)_S$ transformations for $a = 1, 2$ fermions have been tied to each other by the large J_2.

The fermions $\tilde{f}_1$, $\tilde{f}_2$ are neutral under both physical charge and spin probes because they couple to the $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ gauge fields. If the boson R is gapped, and there is no further Higgs term, in the strong gauge field coupling limit $\tilde{f}_1$, $\tilde{f}_2$ will be confined to form on-site spin singlets. A more interesting possibility can be obtained by a mean-field ansatz in which we Higgs the gauge field by coupling $\tilde{f}_{a;\sigma}$ to the physical electron c: we will explore such an ansatz in the remainder of the paper.

B. Mean-field theory

Let us define $C = (c_\uparrow, c_\downarrow, c^\dagger_\downarrow, -c^\dagger_\uparrow)^T$, $\Psi_1 = (\tilde{f}_{1\uparrow}, \tilde{f}_{1\downarrow}, \tilde{f}^\dagger_{1\downarrow}, -\tilde{f}^\dagger_{1\uparrow})^T$ and $\Psi_2 = (\tilde{f}_{2\uparrow}, \tilde{f}_{2\downarrow}, \tilde{f}^\dagger_{2\downarrow}, -\tilde{f}^\dagger_{2\uparrow})^T$. We define ρ_a and μ_a as Pauli matrices acting on the spin and particle-hole channels, respectively. Then the $SU(2)_S$ gauge transformation $U_{i,S}$ is generated by ρ_a, and acts on both Ψ_1 and Ψ_2. The $SU(2)_1$ gauge transformation $U_{i;1}$ is generated by μ_a, and acts only on Ψ_1. Similarly, the $SU(2)_2$ gauge transformation $U_{i;2}$ is generated by μ_a and acts only on Ψ_2. In this basis, the gauge transformations $U_{i;1}$, $U_{i;2}$, and $U_{i;S}$ are 4×4 matrices.

After condensation of appropriate Higgs fields, we obtain our proposed mean-field theory:

$$\begin{aligned} H_M = \sum_i (C^\dagger_i \mathbf{B}_{i;1} \Psi_{i;1} + \text{H.c.} + \Psi^\dagger_{i;1} \mathbf{B}_{i;2} \Psi_{i;2} + \text{H.c.}) \\ + H_C + H_1 + H_2 + H_R, \end{aligned} \tag{11}$$

where $\mathbf{B}_{i;1}$ and $\mathbf{B}_{i;2}$ are 4×4 matrices. The Hamiltonian H_C is the bare kinetic term for physical electron c. The Hamiltonians H_1 and H_2 are mean-field ansatzes for Ψ_1 and Ψ_2 which we will specify later. The Hamiltonian H_R is the ansatz for the slave spin R_i; at large J_2 limit, we can assume R is massive and $\langle R_i \rangle = 0$ by invariance under physical global spin rotations.

Under the $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ gauge transformation, $\mathbf{B}_{i;1}$ and $\mathbf{B}_{i;2}$ transform as

$$\begin{aligned} \mathbf{B}_{i;1} &\rightarrow \mathbf{B}_{i;1} U^\dagger_{i;1} U^\dagger_{i;S} \\ \mathbf{B}_{i;2} &\rightarrow U_{i;S} U_{i;1} \mathbf{B}_{i;2} U^\dagger_{i;2} U^\dagger_{i;S}. \end{aligned} \tag{12}$$

Note that the $U_{i;a}$ commute with the $U_{i;S}$ [35], and so their ordering is unimportant.

We also have a global $U(2)_C$ symmetry corresponding to charge conservation and spin rotation. In the basis of C, the

$U(1)$ part is generated by μ_z while the $SU(2)$ spin rotation is generated by $\boldsymbol{\rho}$. Under this global $U(2)_C$ transformation U_C, $C_i \rightarrow U_C C_i$, $\mathbf{B}_{i;1} \rightarrow U_C \mathbf{B}_{i;1}$, while $\mathbf{B}_2$, Ψ_1, Ψ_2 remain unchanged. Hence $\mathbf{B}_{i;1}$ carries both physical charge-spin and gauge charges.

The mean-field characterizations of the phases are

$$\text{FL}^* : \langle \mathbf{B}_1 \rangle = \Phi \rho_0 \otimes \mu_z, \quad \langle \mathbf{B}_2 \rangle = 0$$
$$\text{FL} : \langle \mathbf{B}_1 \rangle = 0, \quad \langle \mathbf{B}_2 \rangle = \Phi' \rho_0 \otimes \mu_z, \tag{13}$$

where Φ and Φ' are real numbers. Here we chose a specific gauge. Equivalent ansatzes can be obtained through gauge transformations. In the FL phase, $\langle \mathbf{B}_2 \rangle \neq 0$ does not need to be put by hand; once $\langle \mathbf{B}_1 \rangle = 0$ and $\langle R_i \rangle = 0$, gauge fluctuation can confine Ψ_1 and Ψ_2 automatically, which is equivalent to the effect of nonzero $\langle \mathbf{B}_2 \rangle$.

The condensate $\langle \mathbf{B}_1 \rangle = \Phi \rho_0 \otimes \mu_z$ Higgses the $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ gauge fields (the hopping terms in H_1 and H_2 also Higgs parts of the gauge symmetry even with $\langle \mathbf{B}_1 \rangle = 0$). In a more precise language, it locks the $SU(2)_C \subset U(2)_C$ background field to the internal gauge fields corresponding to $SU(2)_S$. Thus the spin index $\tilde{\sigma}$ carried by $\tilde{f}_{1\tilde{\sigma}}$ and $\tilde{f}_{2\tilde{\sigma}}$ can now be identified as a physical spin index. The internal $U(1)$ gauge field generated by μ_z in $SU(2)_1$ is locked to the physical electromagnetic field; after the condensation of $\mathbf{B}_1$, $\tilde{f}_1$ can be viewed as electron and $\tilde{f}_2$ can be identified as a spinon. The fermion $\tilde{f}_2$ is still charge neutral because $SU(2)_2$ is not locked to the physical background field.

III. PROPERTIES OF THE PSEUDOGAP METAL

In this section we provide details of the FL* mean-field ansatz for the pseudogap metal in the underdoped region and discuss its properties. When $\langle \mathbf{B}_1 \rangle = \Phi \rho_0 \otimes \mu_z$, $\langle \mathbf{B}_2 \rangle = 0$, we have the following mean-field theory:

$$H_M = H_{c,\tilde{f}_1} + H_{\tilde{f}_2}. \tag{14}$$

The Hamiltonian $H_{c,\tilde{f}_1}$ describes the electron Fermi surface while $H_{\tilde{f}_2}$ describes the phase of the spinon $\tilde{f}_2$. Small Fermi surfaces with Luttinger volume $A_{FS} = p/2$ can be obtained by

$$\begin{aligned} H_{c,\tilde{f}_1} = \sum_{ij} (-t_{c;ij} c^\dagger_{i\sigma} c_{j\sigma} + t_{1;ij} \tilde{f}^\dagger_{i;1\sigma} \tilde{f}_{j;1\sigma} + \text{H.c.}) \\ - \mu_c \sum_i c^\dagger_{i;\sigma} c_{i;\sigma} - \mu_1 \sum_i \tilde{f}^\dagger_{i;1\sigma} \tilde{f}_{i;1\sigma} \\ + \Phi \sum_i (c^\dagger_{i;\sigma} \tilde{f}_{i;1\sigma} + \text{H.c.}), \end{aligned} \tag{15}$$

where μ_1 is added to fix $n_{\tilde{f}_1} = \sum_\sigma \langle \tilde{f}^\dagger_{i;1\sigma} \tilde{f}_{i;1\sigma} \rangle = 1$, and μ_c fixes $n_c = 1 - p$. For the hopping parameters $t_{c;ij}$, we use $t = 1, t' = -0.22, t'' = 0.19$, which can reproduce the shape of the Fermi surface in the overdoped regime. The hopping parameters for $\tilde{f}_1$ should be determined by minimizing the energy for the wave function in (4). Alternatively, we can also view them as phenomenological parameters and fit them from experimental data. Here we choose $t_1 = 1, t'_1 = -0.1, t''_1 = 0.1$ for the purpose of illustration. In practice, the hopping parameters $t_{1;ij}$ can also have dependences on doping level p. Later we will provide an intuitive explanation why the hoppings of $\tilde{f}_1$ have the opposite sign to that of c.
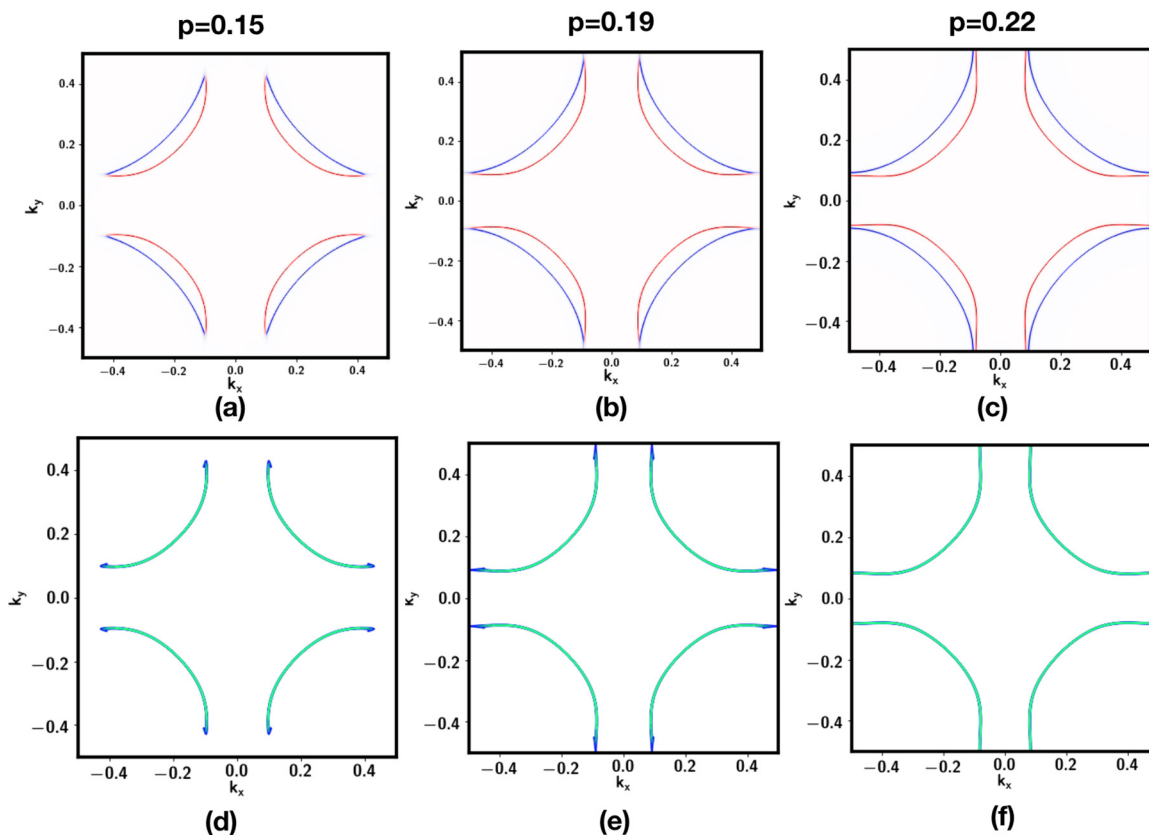
FIG. 2. Spectral function $A(\omega = 0, \mathbf{k})$ from mean-field theory at various doping levels. k_x and k_y are in unit of $\frac{2\pi}{a}$. In (a)–(c), red line is dominated by $A_c(\omega = 0, \mathbf{k})$ while blue line is dominated by $A_{\tilde{f}_1}(\omega = 0, \mathbf{k})$. In (d)–(f) we only show $A_c(\omega = 0, \mathbf{k})$, which clearly shows Fermi arc at $p < p_l \approx 0.19$. At $p_l \approx 0.19$, there is a Lifshitz transition.

For the spinon $\tilde{f}_2$, we use the familiar d wave pairing ansatz [1]:

$$H_{\tilde{f}_2} = -t_2 \sum_{\langle ij \rangle} \tilde{f}^\dagger_{i;2} \tilde{f}_{j;2} + \sum_{i,\hat{\mu}=x,y} \Delta_{\hat{\mu}} (\epsilon_{\sigma\sigma'} \tilde{f}^\dagger_{i;2\tilde{\sigma}} \tilde{f}^\dagger_{i+\hat{\mu};2\tilde{\sigma}'} + \text{H.c.}),$$

(16)

where, $\Delta_{\hat{x}} = -\Delta_{\hat{y}} = \Delta$. This ansatz is equivalent to the staggered flux ansatz, and the spinon $\tilde{f}_2$ is in a $U(1)$ Dirac spin liquid phase.

The gap at the antinode is opened by $\Phi \neq 0$. We choose to use $\Phi(p) = 0.25\sqrt{0.23 - p}$. Then we can calculate spectral densities $A_c(\omega, \mathbf{k}) = (1/\pi)\text{Im}\langle c^\dagger(\omega, \mathbf{k}) c(\omega, \mathbf{k}) \rangle$ and $A_{\tilde{f}_1}(\omega, \mathbf{k}) = (1/\pi)\text{Im}\langle \tilde{f}^\dagger_1(\omega, \mathbf{k}) \tilde{f}_1(\omega, \mathbf{k}) \rangle$. We show plots of the calculated spectral densities in Fig. 2. Between $p_l \approx 0.19$ and $p_c = 0.23$, the antinode is not gapped even if $\Phi \neq 0$. Instead, there is one-hole Fermi surface dominated by c and one electron pocket dominated by $\tilde{f}_1$. The total Hall number is close to p at high temperature[1] and this region still belongs to the pseudogap phase. When we decrease p below p_l, there is a Lifshitz transition and the Fermi surfaces are reconstructed to four small hole pockets centering at node $\mathbf{K_N} = (\pm\frac{\pi}{2}, \pm\frac{\pi}{2})$. For each pocket, one side is dominated by $\tilde{f}_1$ and thus has almost vanishing spectral weight in terms of c. As a result, only Fermi arcs are visible in an ARPES measurement. Note

also the similarity of Fig. 2 to the STM observations in Refs. [4,5].

In Fig. 3 we show $A_c[\omega, \mathbf{k} = (\pi, k_y)]$. We define antinode to be at $\mathbf{K}_{AN} = (\pi, \pm\delta)$, which separates the $n_c(\mathbf{k}) = 1$ and $n_c(\mathbf{k}) = 0$ regions along the cut of $k_x = \pi$. Then at $\mathbf{k} = \mathbf{K}_{AN}$, $A_c(\omega, \mathbf{K}_{AN})$ has a peak at $\omega = -\Delta$. Here the sharp quasiparticle peak at $\omega = -\Delta$ is an artifact of the mean-field calculation, which ignores gauge fluctuation. Even inside the pseudogap phase, the mass of the Higgs gauged fields is at order of $\Phi \sim \Delta$. For high-energy region $|\omega| > \Delta$, gauge fluctuations can not be ignored and may completely destroy the quasiparticle peak at antinode. In contrast, around the node, the Fermi arc is at zero energy and the gauge fluctuations do not have strong influences on spectral function here.

We define Δ as the gap of the antinode. The dependence of Δ and Φ on the doping level is shown in Fig. 4. The doping p_c is the true quantum critical point QCP at the end of the pseudogap phase. For our parameters, the antinode gap Δ onsets at a smaller doping p_l through a Lifshitz transition. In slightly overdoped region $p_c < p < p_d$, there may be ghost Fermi surfaces decoupled from the Fermi liquid, which get confined and disappear at a larger doping p_d.

A. Physical meaning of auxiliary fermions

At the $J_2 \to \infty$ limit, the wave function of (4) can be written as

$$|\Phi\rangle = (\langle s|\text{Slater}[c, \tilde{f}_1]\text{Slater}[\tilde{f}_2]\rangle) |s\rangle,$$ (17)

---

[1]At low temperature it depends on the scattering rate of c and $\tilde{f}_1$.
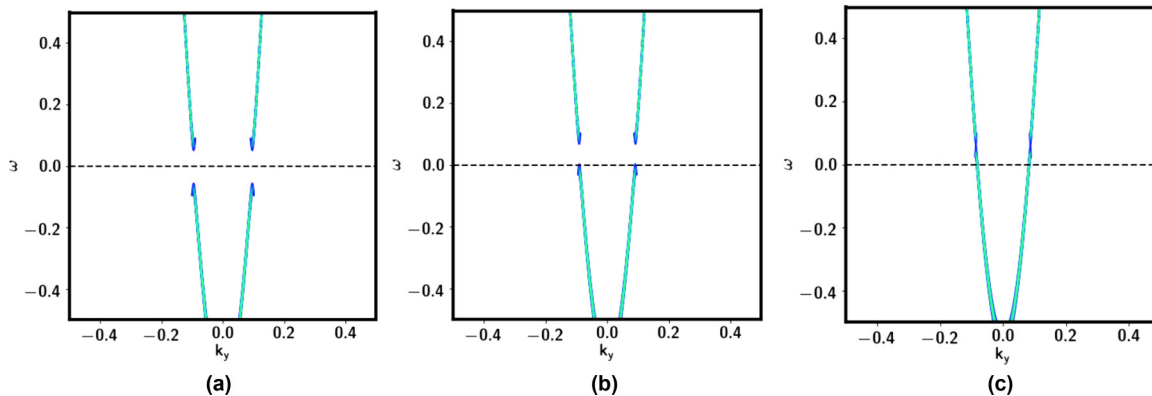
FIG. 3. Spectral function $A_c[\omega, \mathbf{k} = (\pi, k_y)]$ calculated from mean-field theory at various doping levels. The doping is (a) $p = 0.15$; (b) $p = 0.19$; and (c) $p = 0.21$. In ARPES only the $\omega < 0$ region can be measured at low temperature. Because of the gauge fluctuations at high energy, the true spectral function may be quite different and no sharp quasiparticle peak exists.

where $|s\rangle = \prod_i (\tilde{f}^\dagger_{i;1\uparrow} \tilde{f}^\dagger_{i;2\downarrow} - \tilde{f}^\dagger_{i;1\downarrow} \tilde{f}^\dagger_{i;2\uparrow})/\sqrt{2}|0\rangle$. Slater$[c, \tilde{f}_1]$ is the ground state of $H_{c,\tilde{f}_1}$ in Eq. (15). Slater$[\tilde{f}_2]$ is the ground state of $H_{\tilde{f}_2}$ in (16).

The above wave function is a state purely in the Hilbert space of the physical layer. Thus $\tilde{f}_1$ and $\tilde{f}_2$ should correspond to physical degrees of freedom. A natural question is: what is the physical meaning of these auxiliary fermions inside the physical Hilbert space? To gain intuition, let us look at the zero doping case first. In this case, Φ can gap out both c and $\tilde{f}_1$ and we have a Mott insulator. Therefore we should interpret Φ as Mott gap. The Φ condensate binds electron $c^\dagger$ to hole $\tilde{f}_1$, and so we should view $\tilde{f}_1$ as creation operator for a correlation hole. This correlation hole may be quite nonlocal and should not be confused with on-site hole operator c_i. In a certain sense, the physics is similar to that of fractional quantum Hall effect (FQHE). In the FQHE system, because of Coulomb interaction, electron also binds with its correlation hole (also called a vortex) and only the bound state (a composite fermion) can move coherently. In our case,

the Hubbard U also favors the binding between electron and correlation hole, which causes Mott localization. Unlike the FQHE, here in (15) we use a BCS description of the exciton binding instead of viewing the exciton as the fundamental particle. The exciton binding means that the exciton $c^\dagger \tilde{f}_1$ moves coherently, and thus the hopping of $\tilde{f}_1$ should be similar to that of $c^\dagger$. Thus it is natural that the hopping of $\tilde{f}_1$ has the opposite sign to that of c.

At zero doping, c, $\tilde{f}_1$ are gapped and generate the upper Hubbard band, and the lower Hubbard band for the Mott insulator. The fermion $\tilde{f}_2$ can be identified as spinon at low energy. At small doping p, the Mott gap does not close immediately and the doped hole just enters the lower Hubbard band, and forms small hole pockets. In this sense, the pseudogap is inherited from the Mott gap of the undoped parent compound. This picture is illustrated in Fig. 5. If the spinon part is in a spin liquid phase as we assume here, the phase is a fractionalized Fermi liquid (FL*). FL* phase has been proposed before [15,21,23], but our construction is, as far as we know, the first theory to describe the FL* phase fully in mean-field level and naturally explain the observed Fermi arcs. Another advantage of our theory is that the spinon part explicitly decouples from the hole pockets. Here we assume the spinon band $\tilde{f}_2$ forms the $U(1)$ Dirac spin liquid. However, Neel ordered phase at small doping is also possible and can be viewed as descendant of the Dirac spin liquid. In principle,
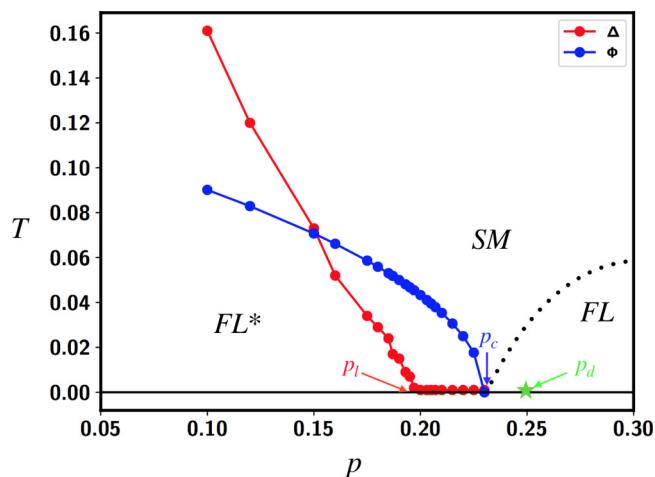

FIG. 4. Phase diagram in $T - x$ space. We choose $\Phi = 0.25\sqrt{0.23 - p}$. Δ is calculated from mean-field theory and is in unit of $t = 1$. The dashed line is drawn by hand to show the other crossover line of the quantum critical region.
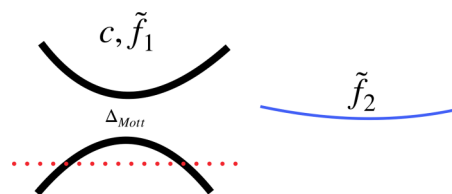

FIG. 5. An illustration of the pseudogap metal in our theory. The red dashed line is the chemical potential. c, $\tilde{f}_1$ form the upper Hubbard band and the lower Hubbard band of the Mott insulator. The chemical potential moves from inside the Mott gap to the lower Hubbard band upon hole doping, resulting in small hole pockets. $\tilde{f}_2$ is the neutral spinon band, which decouples from the hole pockets. For cuprate we assume $\tilde{f}_2$ forms the famous $U(1)$ Dirac spin liquid.

the spinon part can provide additional contributions to thermal transport. Actually, the recent observed giant thermal Hall effect in the pseudogap metal region [36] may be naturally attributed to the Dirac spin liquid part of this FL* phase. We will explore this aspect in the future.

IV. CRITICAL THEORY

In this section we provide the theory at the critical point p_c, where a nonzero Φ onsets. At the critical point, the Higgs condensate $\mathbf{B}_1$ fluctuates along the manifold generated by the gauge transformations. The hopping terms for $\tilde{f}_1$ and $\tilde{f}_2$ in (15) and (16) break the $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ down to $[U(1)_1 \times U(1)_2 \times SU(2)_S]/Z_2$. Basically there is no gauge transformation rotating a particle to a hole. As a consequence, the fluctuation of $\mathbf{B}_1$ is generated by a $U(2) = [U(1)_1 \times SU(2)_S]/Z_2$ transformation. We define $\tilde{C}_i = (c_{i;\uparrow}, c_{i;\downarrow})^T$, $\tilde{\Psi}_{i;1} = (\tilde{f}_{i;1\uparrow}, \tilde{f}_{i;1\downarrow})^T$ and $\tilde{\Psi}_{i;2} = (\tilde{f}_{i;2\uparrow}, \tilde{f}_{i;2\downarrow})^T$. We define the $U(1)$ gauge fields for $U(1)_1$ and $U(1)_2$ as a_1 and a_2. The $SU(2)_S$ gauge field is labeled as $\boldsymbol{\alpha}$.

The critical theory is

$$L = L_C + L_{\Psi_1, a_1, \boldsymbol{\alpha}} + L_{\Psi_2, a_2, \boldsymbol{\alpha}} + L_{B, a_1, \boldsymbol{\alpha}}$$
$$+ [\tilde{C}^\dagger(\tau, \mathbf{x}) B(\tau, \mathbf{x}) \tilde{\Psi}_1(\tau, \mathbf{x}) + \text{H.c.}], \quad (18)$$

where $B(\tau, \mathbf{x})$ is a 2×2 matrix parameterized as $B_1(\tau, \mathbf{x}) = \Phi(\tau, \mathbf{x}) U(\tau, \mathbf{x})$. Here $\Phi(\tau, \mathbf{x})$ is a complex field and $U(\tau, \mathbf{x})$ is a $SU(2)$ matrix field.

The Lagrangian L_C is the action for the fermi liquid theory of physical electron. $L_{\Psi_2, a^2, \boldsymbol{\alpha}}$ is the action for the spinon coupled to $[U(1)_2 \times SU(2)_S]/Z_2$ gauge field. For our Dirac spin liquid ansatz, it has $N = 2$ Dirac fermions in the fundamental representation coupled to the $U(2)$ gauge field.

The theory for the critical boson B is

$$L_{B, a^1, \boldsymbol{\alpha}} = \frac{1}{g_B} \left| \partial_\mu B_{\alpha\beta} - a_\mu^1 B_{\alpha\beta} - \alpha_\mu^a B_{\alpha\beta'} \rho_{\beta'\beta}^a \right|^2 - m|B_{\alpha\beta}|^2, \quad (19)$$

which is a nonlinear σ model coupled to $U(2)$ gauge field. We used Einstein summation for $\alpha, \beta, \beta' = \uparrow, \downarrow$. The fermion Ψ_1 forms a Fermi surface and couples to both a^1 and $\boldsymbol{\alpha}$:

$$L_{\Psi_1, a^1, \boldsymbol{\alpha}} = \Psi_1^\dagger \left(\partial_\tau - a_0^1 - \alpha_0^a \rho^a \right) \Psi_1$$
$$- \frac{\hbar^2}{2m^*} \Psi_1^\dagger \left(\partial_i - a_i^1 - \alpha_i^1 \rho^a \right)^2 \Psi_1. \quad (20)$$

The quantum critical point is tuned by the mass term m for the critical boson. At $m < 0$, $\langle B \rangle = \Phi \rho_0$ and this is a FL* phase. When $m > 0$, $\tilde{\Psi}_1$ and $\tilde{\Psi}_2$ decouple from the physical electron. In this case $\tilde{\Psi}_1$ forms ghost a Fermi surface, which couples to neither charge nor spin probes. $\tilde{\Psi}_1$ couples to a $U(1)$ and a $SU(2)$ gauge field. It is known that $U(1)$ gauge field suppresses pairing, while $SU(2)$ gauge field induces pairing [37]. Thus it is not clear whether this ghost Fermi surface is stable or not. It can certainly be stable and survive until a larger doping $p_d > p_c$. If the ghost Fermi surface is unstable with infinitesimal $m > 0$ but is stable at the $m = 0$ point, there is hope to have a direct transition between the FL* phase and the FL phase. It is also possible that the ghost Fermi surface is not stable to pairing even at $m = 0$. The pairing of $\tilde{f}_1$ will also induce pairing for physical electron

c through the coupling $C^\dagger B \Psi$. In this case the QCP will be hidden inside a superconductor dome and the superconductor T_c may be enhanced at $m = 0$ because of the $SU(2)_S$ gauge field. In any case, our theory implies that there is a ghost Fermi surface coexisting with the physical Fermi surface in the strange metal region and the density of states (DoS) should be significantly larger than that of the overdoped Fermi liquid. A recent experimental measurement of specific heat indeed found that $\gamma = C/T$ close to critical region is almost four times larger that that of the overdoped FL [9]. An independent measurement of the effective mass is needed to subtract the contribution from the Fermi liquid part and test the existence of ghost Fermi surface. The ghost Fermi surface can also give an contribution to thermal conductivity, in addition to that from the Fermi liquid part. However, the scattering of the ghost Fermi surface by gauge fields may make its thermal conductivity much smaller than that from the Fermi liquid part. Thus it is not clear whether the violation of the Wiedemann-Franz law is significant enough to be easily detected in the experiments.

V. FUTURE DIRECTIONS

There are several directions to generalize our theory. In this paper we restricted our analysis to the symmetric FL* phase, but it is easy to incorporate symmetry breaking orders. For example, we can let the ansatz of $\tilde{f}_1$ have nematic order or loop current order. In this case, the nematic order or loop current order will onset when $\Phi \neq 0$, and coincides with the onset of the pseudogap phase. However, the symmetry breaking order is just a byproduct of the pseudogap phase and does not play any essential role. Alternatively, $\tilde{f}_2$ can be put in an ansatz with antiferromagnetic order. Thus the theory describes evolution from an antiferromagnetic metal with small Fermi surfaces towards the Fermi liquid phase with large Fermi surface. In contrast to the standard Hertz-Millis theory of antiferromagnetic critical point, this theory allows a jump of carrier density across the critical point and may be relevant for the quantum critical point in heavy fermion systems.

Our framework can also be easily generalized to $SU(N)$ Hubbard model with any N. At integer filling n_c, there is a Mott insulator with n_c electrons per site. We also introduce an auxiliary fermion $\tilde{f}_1$ at density $n_{f_1} = N - n_c$ and an auxiliary fermion $\tilde{f}_2$ with density $n_{\tilde{f}_2} = n_c$. The fermions $\tilde{f}_1$ and $\tilde{f}_2$ can again form trivial $SU(N)$ singlet per site, and there is a $U(1)_1 \times U(1)_2 \times SU(N)_S$ gauge structure. Then upon doping at filling $n = n_c - x$, we can have a coupling like $-\Phi c_\alpha^\dagger \tilde{f}_{1\alpha}$, which can lead to small Hall number $\eta_H = -x$. The fermion $\tilde{f}_2$ can be viewed as a spinon, and be put in spin liquid phase or ordered phase. This implies that pseudogap metal is a quite universal phenomena upon doping a generic Mott insulator. Recently an approximate $SU(4)$ Hubbard model is shown to be realized in graphene moiré superlattice [38,39] and thus this generalized theory may be relevant there.

VI. SUMMARY

We have proposed a new framework for describing the pseudogap phase obtained from doping a Mott insulator,

which can also be extended towards understanding its evolution towards the conventional Fermi liquid at larger doping. We showed that the use of ancilla qubits allows us to address the complete doping evolution in a single mean-field framework, which has not been possible in previous work. We applied our theory to the hole doped cuprates. At small doping, we provide a simple parton mean-field theory of the fractionalized Fermi liquid (FL*) with small Fermi surfaces and reproduce the Fermi arc in ARPES experiments. We also provide a critical theory at the end of the pseudogap phase across which the carrier density jumps from p to $1 + p$. Our theory finds a ghost Fermi surface close to the optimal doping which should significantly enhance the density of states.

Finally, we note the relationship of the present gauge theory of optimal doping criticality to a recent $SU(2)$ gauge theory [26,40] of the same regime. The common features are a $SU(2)_S$ gauge field and a large Fermi surface of gauge-neutral electrons c_σ. The differences are that the other theory [26,40] has (i) multiple Higgs fields that transform in the adjoint of $SU(2)_S$, (ii) the Higgs fields transform nontrivially under the space group of the square lattice, (iii) bosonic spinon excitations, which remain gapped across the transition. In contrast, our present theory has (i) a single Higgs field that transforms as a $SU(2)_S$ fundamental, and also under a separate emergent $U(1)$ gauge field, (ii) the Higgs fields is also a fundamental of the global spin rotation $SU(2)$ and the electromagnetic $U(1)$, (iii) there are gapless ghost fermionic excitations; which carry neither spin nor charge, but which carry the charges of the $SU(2)_S$ gauge field, the emergent $U(1)$ gauge field, and another $SU(2)_2$ gauge field. For our present theory, it is possible to take a linear combination of the emergent and electromagnetic $U(1)$'s and transfer the electromagnetic charge from the Higgs field to a ghost fermion [41].

ACKNOWLEDGMENT

APPENDIX: CONSTRAINTS AND GAUGE STRUCTURE

We make some additional comments here on the origin of the $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ gauge structure. The complete constraints from the single occupancy of the hidden layers are [generalizing (2)] [1]:

$$\sum_{\tilde\sigma} \tilde{f}^\dagger_{i;1\tilde\sigma} \tilde{f}_{i;1\tilde\sigma} = 1 , \qquad \sum_{\tilde\sigma} \tilde{f}^\dagger_{i;2\tilde\sigma} \tilde{f}_{i;2\tilde\sigma} = 1,$$

$$\sum_{\tilde\sigma\tilde\sigma'} \epsilon_{\tilde\sigma\tilde\sigma'} \tilde{f}_{i;1\tilde\sigma} \tilde{f}_{i;1\tilde\sigma'} = 0 , \qquad \sum_{\tilde\sigma\tilde\sigma'} \epsilon_{\tilde\sigma\tilde\sigma'} \tilde{f}_{i;2\tilde\sigma} \tilde{f}_{i;2\tilde\sigma'} = 0. \quad \text{(A1)}$$

Writing $\psi_{i;a} = (\tilde{f}_{i;a\uparrow}, \tilde{f}^\dagger_{i;a\downarrow})^T$, these constraints as the vanishing of the Nambu pseudospin operator for each i and $a = 1, 2$ (see also Ref. [42])

$$\boldsymbol{T}_{i;a} = \tfrac{1}{2}\psi^\dagger_{i;a}\boldsymbol{\sigma}\psi_{i;a} = 0. \quad \text{(A2)}$$

So (A1), (A2) contain a total of six real constraints on each lattice site i. In the limit $J_2 \to \infty$, the spins in the hidden layers are projected onto singlets on each site, and so we have three additional constraints

$$\boldsymbol{S}_{i;1} + \boldsymbol{S}_{i;2} = 0, \quad \text{(A3)}$$

after replacing the f_σ in (1) by $\tilde{f}_{\tilde\sigma}$. We note here that these nine constraints per site correspond precisely to the nine generators of the $[SU(2)_1 \times SU(2)_2 \times SU(2)_S]/Z_2$ gauge symmetry.

This pseudospin operator in (A2) transforms as an adjoint under $SU(2)_1 \times SU(2)_2$, and so its vanishing is maintained under these transformations. The expressions in (A1) are explicitly spin rotation invariant, and so are also invariant under the $SU(2)_S$ transformation in (9).

Another feature of the $SU(2)$ gauge transformation in Ref. [1] is that it leaves the spin operator invariant. So (A3) is invariant under $SU(2)_1 \times SU(2)_2$. Finally, we note that the $SU(2)_S$ spin rotation in (9) performs an adjoint rotation of (A3) in spin space, and so the vanishing of the total spin per site is also obeyed after the $SU(2)_S$ transformation.

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

[2] C. Proust and L. Taillefer, Annu. Rev. Condens. Matter Phys. **10**, 409 (2019).

[3] I. M. Vishik, M. Hashimoto, R.-H. He, W.-S. Lee, F. Schmitt, D. Lu, R. G. Moore, C. Zhang, W. Meevasana, T. Sasagawa, S. Uchida, K. Fujita, S. Ishida, M. Ishikado, Y. Yoshida, H. Eisaki, Z. Hussain, T. P. Devereaux, and Z.-X. Shen, Proc. Natl. Acad. Sci. **109**, 18332 (2012).

[4] Y. He, Y. Yin, M. Zech, A. Soumyanarayanan, M. M. Yee, T. Williams, M. C. Boyer, K. Chatterjee, W. D. Wise, I. Zeljkovic, T. Kondo, T. Takeuchi, H. Ikuta, P. Mistark, R. S. Markiewicz, A. Bansil, S. Sachdev, E. W. Hudson, and J. E. Hoffman, Science **344**, 608 (2014).

[5] K. Fujita, C. K. Kim, I. Lee, J. Lee, M. H. Hamidian, I. A. Firmo, S. Mukhopadhyay, H. Eisaki, S. Uchida, M. J. Lawler, E.-A. Kim, and J. C. Davis, Science **344**, 612 (2014).

[6] S. Badoux, W. Tabis, F. Laliberté, G. Grissonnanche, B. Vignolle, D. Vignolles, J. Béard, D. A. Bonn, W. N. Hardy, R. Liang, N. Doiron-Leyraud, L. Taillefer, and C. Proust, Nature (London) **531**, 210 (2016).

[7] J. Loram, J. Luo, J. Cooper, W. Liang, and J. Tallon, J. Phys. Chem. Sol. **62**, 59 (2001).

[8] J. L. Tallon, J. G. Storey, J. R. Cooper, and J. W. Loram, arXiv:1907.12018.

[9] B. Michon, C. Girod, S. Badoux, J. Kačmarčík, Q. Ma, M. Dragomir, H. A. Dabkowska, B. D. Gaulin, J. S. Zhou, S. Pyon, T. Takayama, H. Takagi, S. Verret, N. Doiron-Leyraud, C. Marcenat, L. Taillefer, and T. Klein, Nature (London) **567**, 218 (2019).

[10] Y. Tang, L. Mangin-Thro, A. Wildes, M. K. Chan, C. J. Dorow, J. Jeong, Y. Sidis, M. Greven, and P. Bourges, Phys. Rev. B **98**, 214418 (2018).

[11] S.-D. Chen, M. Hashimoto, Y. He, D. Song, K.-J. Xu, J.-F. He, T. P. Devereaux, H. Eisaki, D.-H. Lu, J. Zaanen, and Z.-X. Shen, Science **366**, 1099 (2019).

[12] C. Panagopoulos, J. L. Tallon, B. D. Rainford, T. Xiang, J. R. Cooper, and C. A. Scott, Phys. Rev. B **66**, 064501 (2002).

[13] C. Panagopoulos, A. P. Petrovic, A. D. Hillier, J. L. Tallon, C. A. Scott, and B. D. Rainford, Phys. Rev. B **69**, 144510 (2004).

[14] M. Frachet, I. Vinograd, R. Zhou, S. Benhabib, S. Wu, H. Mayaffre, S. Krämer, S. K. Ramakrishna, A. Reyes, J. Debray, T. Kurosawa, N. Momono, M. Oda, S. Komiya, S. Ono, M. Horio, J. Chang, C. Proust, D. LeBoeuf, and M.-H. Julien, arXiv:1909.10258.

[15] T. Senthil, S. Sachdev, and M. Vojta, Phys. Rev. Lett. **90**, 216403 (2003).

[16] T. Senthil, M. Vojta, and S. Sachdev, Phys. Rev. B **69**, 035111 (2004).

[17] A. Paramekanti and A. Vishwanath, Phys. Rev. B **70**, 245118 (2004).

[18] X.-G. Wen and P. A. Lee, Phys. Rev. Lett. **76**, 503 (1996).

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

[20] Y. Qi and S. Sachdev, Phys. Rev. B **81**, 115129 (2010).

[21] J.-W. Mei, S. Kawasaki, G.-Q. Zheng, Z.-Y. Weng, and X.-G. Wen, Phys. Rev. B **85**, 134519 (2012).

[22] M. Punk and S. Sachdev, Phys. Rev. B **85**, 195123 (2012).

[23] M. Punk, A. Allais, and S. Sachdev, Proc. Natl. Acad. Sci. **112**, 9552 (2015).

[24] J. Feldmeier, S. Huber, and M. Punk, Phys. Rev. Lett. **120**, 187001 (2018).

[25] B. Verheijden, Y. Zhao, and M. Punk, SciPost Phys. **7**, 074 (2019).

[26] S. Sachdev, H. D. Scammell, M. S. Scheurer, and G. Tarnopolsky, Phys. Rev. B **99**, 054516 (2019).

[27] S. Sachdev, M. A. Metlitski, Y. Qi, and C. Xu, Phys. Rev. B **80**, 155129 (2009).

[28] D. Chowdhury and S. Sachdev, Phys. Rev. B **91**, 115123 (2015).

[29] M. S. Scheurer, S. Chatterjee, W. Wu, M. Ferrero, A. Georges, and S. Sachdev, Proc. Natl. Acad. Sci. **115**, E3665 (2018).

[30] R. K. Kaul, Y. B. Kim, S. Sachdev, and T. Senthil, Nat. Phys. **4**, 28 (2008).

[31] S. Sachdev, Rep. Prog. Phys. **82**, 014001 (2019).

[32] E. G. Moon and S. Sachdev, Phys. Rev. B **83**, 224508 (2011).

[33] V. Pasquier and F. D. M. Haldane, Nucl. Phys. B **516**, 719 (1998).

[34] N. Read, Phys. Rev. B **58**, 16262 (1998).

[35] C. Xu and S. Sachdev, Phys. Rev. Lett. **105**, 057201 (2010).

[36] G. Grissonnanche, A. Legros, S. Badoux, E. Lefrancois, V. Zatko, M. Lizaire, F. Laliberte, A. Gourgout, J.-S. Zhou, S. Pyon *et al.*, Nature (London) **571**, 376 (2019).

[37] M. A. Metlitski, D. F. Mross, S. Sachdev, and T. Senthil, Phys. Rev. B **91**, 115111 (2015).

[38] G. Chen, L. Jiang, S. Wu, B. Lyu, H. Li, B. L. Chittari, K. Watanabe, T. Taniguchi, Z. Shi, J. Jung, Y. Zhang, and F. Wang, Nat. Phys. **15**, 237 (2019).

[39] Y.-H. Zhang and T. Senthil, Phys. Rev. B **99**, 205150 (2019).

[40] H. D. Scammell, K. Patekar, M. S. Scheurer, and S. Sachdev, arXiv:1912.06108.

[41] L. B. Ioffe and A. I. Larkin, Phys. Rev. B **39**, 8988 (1989).

[42] M. S. Scheurer and S. Sachdev, Phys. Rev. B **98**, 235126 (2018).