Composite fermion theory for pseudogap phenomena and superconductivity in underdoped cuprate superconductors

Youhei Yamaji* and Masatoshi Imada
Department of Applied Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
(Received 12 February 2011; revised manuscript received 5 May 2011; published 30 June 2011)

We study pseudogap phenomena and Fermi-arc formation experimentally observed in typical two-dimensional doped Mott insulators, namely, underdoped cuprate superconductors. To develop a physically unequivocal theory, we start from the slave-boson mean-field theory for the Hubbard model on a square lattice. Our crucial step is to further take into account the charge dynamics and fluctuations. The extra charge fluctuations seriously modify low-energy single-particle spectra of doped Mott insulators near the Fermi level: An electron added around an empty site (or a hole added around a doubly occupied site) constitutes *composite fermion* (cofermion), called *holo-electron* (or *doublo-hole*) at low energy in distinction from the normal quasiparticles. These unexplored composite fermions substantiate the extra charge fluctuation. We show that the quasiparticles hybridize with the holo-electrons and doublo-holes. The resultant hybridization gap is identified as the pseudogap observed in the underdoped region of the high-T_c cuprates. Because the Fermi level crosses the top (bottom) of the low-energy band formed just below (above) the hybridization gap in the hole-doped (electron-doped) case, it causes a Fermi-surface reconstruction, namely, a *topological change in the Fermi surface* forced by the penetration of zeros of the quasiparticle Green's function. This reconstruction signals the emergence of a non-Fermi-liquid phase. The pseudogap and the resultant formation of pocket or arc of the Fermi surface reproduce the experimental results for the cuprate superconductors in the underdoped region. The pairing channel opens not only between two quasiparticles, but also between a quasiparticle and a cofermion. This pairing solves the puzzle of the dichotomy between the d-wave superconductivity and the precursors of the the insulating gap in the antinodal region. We propose and analyze them as the mechanism of the high-temperature superconductivity for the cuprates.

DOI: 10.1103/PhysRevB.83.214522 PACS number(s): 71.10.Hf, 71.10.Fd, 71.30.+h, 74.72.−h

I. INTRODUCTION

The discovery of cuprate superconductors has triggered extensive studies on the nature of low-energy electronic excitations evolving in doped Mott insulators. The extensive interest on the doped Mott insulators exists because it must be directly related with the origin of the high-temperature superconductivity itself.

From the early stage of the studies, experimental observations have indicated that quasiparticle states of normal phases of cuprates change qualitatively with the increase of the doping, together with changes in superconducting transition temperatures. It has been shown[1,2] that the Hall coefficient R_H changes its sign near the so-called optimal doping for the highest superconducting transition temperatures, followed by a steep increase of its amplitude $|R_H|$ with lowering doping concentration, typically indicating a drastic change in low-energy quasiparticles of the normal state between the underdoped and overdoped regions. Such a drastic change is also (but differently) suggested from observations of pseudogap phenomena,[3–8] where the spin and charge excitations are unexpectedly suppressed in the underdoped region. Various types of non-Fermi-liquid properties are accompanied in this region.

Recent improved experimental tools have enabled resolving low-energy single-particle spectra near the Fermi level. In particular, strongly momentum-dependent quasiparticle states in the hole-underdoped cuprates[9,10] observed by angle-resolved photoemission spectroscopy (ARPES) studies have renewed the interest in the low-energy spectrum of the cuprate superconductors. In contrast to the overdoped region, where a large Fermi surface crossing the region around the so-called antinodal points $(\pm\pi,0)$ and $(0,\pm\pi)$ in the two-dimensional (2D) Brillouin zone for the CuO_2 plane is clearly observed, low-energy quasiparticle states around the antinodal points are missing in the underdoped cuprates. It emerges as a truncation of the large Fermi surface observed in the overdoped region. The resultant truncated structure is called the "Fermi arc."

From a more fundamental point of view, the normal state of the cuprates offers a challenge of condensed matter physics as a typical open issue of "Mott physics," namely, nature of strongly correlated metals in the proximity to the Mott insulator.[11–13] The experimentally observed arclike Fermi surface in the underdoped cuprates is a hallmark of the proximity to the Mott insulators. As we illustrate in Fig. 1, global energy spectra of doped Mott insulators are known to consist of essentially three energy "bands": a coherent band around the Fermi level μ and two incoherent bands, namely, the upper Hubbard band (UHB) located above μ and the lower Hubbard band (LHB) located below μ. For the hole-doped (electron-doped) systems, the coherent band is formed around the top of the LHB (the bottom of the UHB). The spectral weight formed just above (below) μ within the coherent band is often called the *low-energy unoccupied spectral weight* (LUSW) in the hole-doped (electron-doped) Mott insulators.

Such a global structure of spectra can be roughly described by a simple picture given by the dynamical mean-field theory.[14] It has unified two scenarios of earlier studies by Hubbard[15] and Brinkman and Rice[16] based on the Gutzwiller approximation.[17] The Hubbard approximation captures the formation of UHB and LHB. In the Brinkman-Rice picture, Mott insulators appear as a consequence of homogeneous vanishing of the quasiparticle weight uniformly on the Fermi surface. Therefore, this approximation draws a picture that

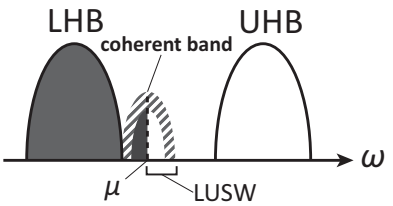

FIG. 1. Schematic global density of states of hole-doped Mott insulators. The coherent band around the Fermi level μ, the upper Hubbard band (UHB), and the lower Hubbard band (LHB) are schematically illustrated. The low-energy unoccupied spectral weight (LUSW) is also indicated.

quasiparticles are renormalized in a momentum-independent fashion and the effective mass diverges independently of the momentum position on the Fermi surface on the verge of the Mott transition.

In infinite-dimensional systems, such a route of the Mott transition has turned out to be correct.[14,18,19] However, in finite dimensions, this picture has turned out to be too simple to understand the low-energy spectra and the nature of the Mott transition: Realistic theory has to capture significant momentum-dependent quasiparticle renormalization, as has been experimentally observed as Fermi-arc formation in the underdoped cuprates, and has also been revealed from various numerical attempts.[20–26] It has now become increasingly clear that a conceptually new idea for describing Mott physics including an emergence of the new non-Fermi-liquid phase in the underdoped region is desired.

In this paper, we extend our previous attempt[27] for describing the momentum-dependent Mott physics, including the mechanism of Mott transitions at the level of LUSW, and for understanding the experimentally supported dramatic change in the electronic states of the cuprates. We give a detailed comparison with experiments and an unconventional superconducting mechanism based on the previous attempt,[27] together with the derivation of our theoretical description in details. We particularly pay attention to the possible reconstruction of the Fermi surface. If the reconstruction emerges, the doped Mott insulator becomes topologically inequivalent to the conventional Fermi liquid and, hence, offers an unexplored possibility for long-standing issue of unconventional metals.

In this paper, we propose a simple and physically transparent theory that accounts for the strongly momentum-dependent renormalization experimentally suggested in the cuprate superconductors. We focus on charge dynamics involved in the low-energy spectra of the doped Mott insulators. A key idea is that, near the Mott insulator, an electron (a hole) added to an empty (a doubly occupied) site costs much smaller energy than an electron (a hole) added to singly occupied sites and behaves as a component of a band separated from the main quasiparticle. Then, near the Mott insulator, this separated excitation contributes to the LUSW in the hole-doped (electron-doped) Mott insulators. This is in contrast with the weakly correlated regime, in which an added electron (hole) constitutes a uniform excitation irrespective of the added site because the added electron becomes uniformly and spatially extended with the momentum as a good quantum number. In our theory, such an electron (a hole) added to an empty site (a doubly occupied site) forms a composite fermionic excitation (or cofermion), which we call *holo-electron* (*doublo-hole*). The conventional quasiparticles may be scattered by the doubly occupied site (empty site) and are transformed to the cofermions, which introduce finite lifetime into the quasiparticles. This scattering and transformation may alternatively be formulated as the hybridization between the conventional quasiparticles and the holo-electrons and/or doublo-holes. This hybridization naturally generates a hybridization gap and the resultant band splitting causes the Fermi-surface reconstructions, namely, topological changes in the Fermi surface. The hybridization gap and the topological change account for the experimentally suggested pseudogap and Fermi-arc formation, respectively, in the cuprates. Our unique prediction is that the pseudogap is independent of the precursor of the Mott gap itself, while the pseudogap has the structure of the s-wave-type symmetry rather than the d-wave symmetry. In addition, the pseudogap identified here is in clear distinction from the superconducting gap as well. We propose several experimental tests to prove or disprove our theoretical consequences.

A way to understand topological changes in the Fermi surface has recently been proposed to originate from the emergence of zeros of single-particle Green's functions, which are the points in the (k,ω) space satisfying Re $G(k,\omega) = 0$ for the momentum k and the frequency ω. In other words, the single-particle self-energy $\Sigma(k,\omega)$ diverges at the zeros of Green's function. The idea of the emergence of the zeros has a root in the work by Dzyaloshinskii, who examined a possible extension of the Luttinger theorem into Mott insulating states.[28] As is reviewed in Ref. 29, this idea is applied to the doped Mott insulators to explain the "Fermi arcs" observed by ARPES. Recent results of numerical calculations also suggest that the zeros emerging in doped Mott insulators reconstruct the Fermi surface and changes its topology.[23,25,26] Such a topological change is indeed claimed in a recent ARPES measurement.[30]

The Fermi-surface reconstruction itself is not an unconventional phenomenon if it accompanies a spontaneous symmetry breaking such as an antiferromagnetic order. In translational-symmetry broken phases with an ordering wave number Q, electrons with momentum k hybridize, at least, with electrons at momentum $k + Q$. Then, Fermi-surface reconstructions are naturally understood as a consequence of the hybridization gap as we will discuss later in detail. It is indeed able to account for the formation of Fermi pockets or arclike Fermi surfaces if a hybridization with other fermionic excitations exist.

However, in the hole-doped cuprates, in spite of recent reports on time-reversal symmetry[31–33] and rotational symmetry[34] breakings in some of the cuprate superconductors, translational symmetry breakings have not been universally observed.[35] On the other hand, the "Fermi arc" or the truncated Fermi surface have universally been experimentally observed in the underdoped cuprates. To have a unified picture of the Mott physics, it is important to understand whether the Fermi-surface reconstruction occurs as a consequence of the hybridization gap generated by the "hybridization" with some hidden fermionic excitations without assuming a symmetry broken phase. In the present theory, the hybridization between

the conventional quasiparticles and the emergent fermionic excitations called the holo-electron and the doublo-hole naturally explain such a Fermi-surface reconstruction.

Now we go into some details regarding characteristic low-energy spectra of doped Mott insulators. As is proposed by Meinders, Eskes, and Sawatzky,[36] the doping dependence of the LUSW, formed just above μ, reflects a unique feature of nearly Mott insulating, correlated electron systems. As we have mentioned above, an electron added to a singly occupied site inevitably costs the on-site Coulomb interaction U and gives an excitation in the UHB, whereas electrons added to empty sites exclusively contribute to the low-energy states just above μ in the atomic limit, where the kinetic energy of the electrons t is zero. Here, an empty site can accept an electron irrespective of its spin state and creates two unoccupied states, namely, up- and down-spin states. In the atomic limit, the $N_{\rm h}$ holes create $N_{\rm h}$ empty sites and, then, create $2N_{\rm h}$ unoccupied states as a consequence. When the kinetic energy of the electrons t becomes nonzero, the number of unoccupied states increases because of the pair creation of doubly occupied and empty sites, which is induced by the hopping of electrons and resultant charge fluctuations. In the Mott insulating phases, adding an electron to these empty sites, enabled by the pair creation, costs the *binding energy* $\sim U$ of a doubly occupied and an empty site. This excitation can not contribute to LUSW. However, in the doped Mott insulators, doped mobile carriers screen the on-site Coulomb interaction and weaken the binding energy. As a result, a part of the empty sites originating from these binding states can create LUSW. Therefore, for hole-doping concentration $x = N_{\rm h}/N_{\rm s}$ with $N_{\rm s}$ being the total number of lattice sites, the LUSW in the doped Mott insulator is always larger than $2x$. From the early stage of the studies on the cuprates, the LUSW developing faster than $2x$ has been observed by optical conductivity measurements.[37] This quick increase of the spectral weight larger than $2x$ requires a picture of doped holes very different from the doped semiconductors, where the LUSW is trivially equal to x. Our scheme presented in this paper naturally explains this unconventional feature.

The mechanism of the superconductivity itself is a central open issue of the physics of the cuprate superconductors. When the LUSW in the normal state belongs to an unconventional phase with emergent excitations, the mechanism has to be understood based on this framework because the energy scale of the superconductivity is even smaller than, and governed by, the energy scale of LUSW. Our theory offers an unconventional channel of the pairing and resultant superconductivity emerging from the contribution of composite fermions never considered in the literature. We examine an unexplored type of quasiparticle pairing arising from the pairing potential generated by pairing fields of cofermions, holo-electrons and doublo-holes, and quasiparticles. Although the pseudogap formation by the hybridization gap of quasiparticles and holo-electrons and doublo-holes is destructive to the superconductivity, this unconventional pairing potential serves to create superconductors. This dual character and two sides of the same coin naturally accounts for the recent puzzle under debate on the dichotomy and nature of the gap in the anitinodal region of the underdoped cuprates.[38,39]

The organization of this paper is the following: In Sec. II, we start from the Kotliar-Ruckenstein mean-field theory and review previous extensions for correlated metals for the self-contained description. In Sec. III, we take into account the charge dynamics, which plays an important role in formation of the LUSW, and explain how emergent excitations, namely, the holo-electrons and doublo-holes emerge. The hybridization between quasiparticles and these composite fermionic excitations naturally causes the Fermi-surface reconstruction, namely, the topological change in the Fermi surface. The pseudogap phenomena observed in cuprate superconductors also emerge because of this hybridization. In Sec. III D, we propose an unexplored pairing mechanism evolved from the cofermions as the mechanism of the high-temperature superconductivity. Our results for single-particle spectra, amplitudes of the pseudogap, Fermi-surface topology, the specific-heat coefficient, and the density of states are presented and compared with experimental results in Sec. IV. We also estimate the superconducting gap amplitude and quantitative aspects of the superconducting mechanism. Section V is devoted to discussions and summary.

II. PREVIOUS THEORIES

A. Hubbard model

The Hubbard model,[15,17,40] defined by the Hamiltonian

$$\hat{H} = \sum_{i,j} t_{ij} \hat{c}^{\dagger}_{i\sigma} \hat{c}_{j\sigma} + U \sum_{i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \tag{1}$$

is a canonical model, which describes the competition between the kinetic energy t_{ij} and the on-site Coulomb interaction U, where $\hat{c}^{\dagger}_{i\sigma}$ ($\hat{c}_{i\sigma}$) is a creation (annihilation) operator of σ-spin electron on the ith site and $\hat{n}_{i\sigma} = \hat{c}^{\dagger}_{i\sigma}\hat{c}_{i\sigma}$ is a number operator. Hereafter, we focus on the Hubbard model on the square lattice as a model for the cuprates. In this chapter, the nearest-neighbor hopping and the next-nearest-neighbor hopping are set as $t_{ij} = -t$ and $t_{ij} = +t'$, and further neighbor hoppings are ignored.

The solution of the Hubbard model on two-dimensional lattices remains an open problem. To get an insight into the nature of the correlated metallic phase, there exist a variety of numerical methods, which give accurate results for finite-size clusters such as exact diagonalization[41] and quantum Monte Carlo.[42] Cluster extensions of the dynamical mean-field theory,[43,44] improved variational Monte Carlo methods,[45,46] Gaussian Monte Carlo method,[47,48] and the path integral renormalization group[49–51] are also available in the literature. However, they all have some limitations. For example, limitations on cluster size are severe in the exact diagonalization and the quantum Monte Carlo, while resolutions in the momentum space are severely limited in the cluster extension of the dynamical mean-field theory.

Without relying on accurate numerical methods, there exist other ways to extract essential physics from analytical or conceptually correct limits. The Landau Fermi-liquid picture[52] offers such an example. Although we have no exact solutions in the thermodynamic limit, low-energy single-particle spectra of the Hubbard model are believed to behave following the Landau Fermi-liquid picture for less correlated systems except for 1D systems. The original proposal for the Mott insulating states itself is not originally based on the numerical

results, but proposed by Peierls and Mott from a gedanken experiment on metallic crystalline hydrogenlike atoms.[53–55] Such phenomenological theories based on physical intuitions offer insights into difficult issues in a wide range of condensed matter physics. In this paper, we try to make a step toward constructing such a physically transparent theory that accounts for the unconventional properties of doped Mott insulators.

B. Kotliar-Ruckenstein formalism

One of the simplest pictures to describe correlated metals and metal-insulator transitions at half-filling (Mott transitions) along the line of the original idea by Mott is the Brinkman-Rice scenario[16] based on the Gutzwiller approximation.[17] As our starting point, we employ the Kotliar-Ruckenstein (KR) slave-boson formalism, which gives the same results as the Gutzwiller approximation. In this section, we briefly outline the KR slave-boson formalism for the Hubbard model[56] to make this paper self-contained. This slave-boson formalism gives a starting point for the mean-field description of strongly correlated electron systems by replacing original electrons with four kinds of bosons and one kind of spinful fermion.

We start with the local Hilbert space of the Hubbard model, which is expanded by a set of the Fock states: the empty state (holon) $|0\rangle$, the singly occupied states $|\uparrow\rangle, |\downarrow\rangle$, and the doubly occupied state (doublon) $|\uparrow\downarrow\rangle$. Corresponding to each Fock state, one slave boson is introduced as $\hat{e}$ for $|0\rangle$, $\hat{p}_\sigma$ for $|\sigma\rangle$, and $\hat{d}$ for $|\uparrow\downarrow\rangle$. In addition to these bosons $\hat{b}$ ($b = e, p_\sigma$, or d), a fermion operator $\hat{f}_\sigma$ is introduced to stand for the σ-spin state and the fermionic nature of the original fermion operator $\hat{c}_{i\sigma}$. The correspondence relation between local basis and the lattice wave functions is given as

$$|0\rangle_i \doteq |\mathrm{vac}\rangle_i^{\mathrm{F}} \otimes \hat{e}_i^\dagger |\mathrm{vac}\rangle_i^{\mathrm{B}}, \tag{2}$$

$$\hat{c}_{i\uparrow}^\dagger |0\rangle_i \doteq \hat{f}_{i\uparrow}^\dagger |\mathrm{vac}\rangle_i^{\mathrm{F}} \otimes \hat{p}_{i\uparrow}^\dagger |\mathrm{vac}\rangle_i^{\mathrm{B}}, \tag{3}$$

$$\hat{c}_{i\downarrow}^\dagger |0\rangle_i \doteq \hat{f}_{i\downarrow}^\dagger |\mathrm{vac}\rangle_i^{\mathrm{F}} \otimes \hat{p}_{i\downarrow}^\dagger |\mathrm{vac}\rangle_i^{\mathrm{B}}, \tag{4}$$

$$\hat{c}_{i\uparrow}^\dagger \hat{c}_{i\downarrow}^\dagger |0\rangle_i \doteq \hat{f}_{i\uparrow}^\dagger \hat{f}_{i\downarrow}^\dagger |\mathrm{vac}\rangle_i^{\mathrm{F}} \otimes \hat{d}_i^\dagger |\mathrm{vac}\rangle_i^{\mathrm{B}}, \tag{5}$$

where $|\mathrm{vac}\rangle_i^{\mathrm{F(B)}}$ is a vacuum of the ith site for fermionic (bosonic) degrees of freedom. Equations (2)–(5) represent the mapping between wave functions written by the original electrons and the fermions $\hat{f}$ combined with bosons $\hat{b}$. This mapping is derived when the electron operators are replaced with composite ones as

$$\hat{c}_{i\sigma}^\dagger \doteq (\hat{p}_{i\sigma}^\dagger \hat{e}_i + \hat{d}_i^\dagger \hat{p}_{i\bar{\sigma}}) \hat{f}_{i\sigma}^\dagger. \tag{6}$$

We should note that the replacement given by Eq. (6) is not a unique one. It is known that operators equivalent to the right-hand side of Eq. (6) can be given as

$$\hat{c}_{i\sigma}^\dagger \doteq \hat{z}_{i\sigma} \hat{f}_{i\sigma}^\dagger, \tag{7}$$

where $\hat{z}_{i\sigma}$ is defined[56,57] as

$$\hat{z}_{i\sigma} = \hat{g}_{i\sigma}^{(1)} (\hat{p}_{i\sigma}^\dagger \hat{e}_i + \hat{d}_i^\dagger \hat{p}_{i\bar{\sigma}}) \hat{g}_{i\sigma}^{(2)}, \tag{8}$$

$$\hat{g}_{i\sigma}^{(1)} = (1 - \hat{p}_{i\bar{\sigma}}^\dagger \hat{p}_{i\bar{\sigma}} - \hat{e}_i^\dagger \hat{e}_i)^{p_1}, \tag{9}$$

$$\hat{g}_{i\sigma}^{(2)} = (1 - \hat{p}_{i\sigma}^\dagger \hat{p}_{i\sigma} - \hat{d}_i^\dagger \hat{d}_i)^{p_2}. \tag{10}$$

The operators $\hat{g}_{i\sigma}^{(1)}$ and $\hat{g}_{i\sigma}^{(2)}$ act as identities when these are operated to $(\hat{p}_{i\sigma}^\dagger \hat{e}_i + \hat{d}_i^\dagger \hat{p}_{i\bar{\sigma}})$ for any powers p_1 and p_2. This ambiguity of the correspondence has been utilized before.[56]

In the expanded Hilbert space, the Hubbard Hamiltonian [Eq. (1)] is rewritten as

$$\hat{H} = \sum_{i,j} t_{ij} \hat{z}_{i\sigma} \hat{f}_{i\sigma}^\dagger \hat{f}_{j\sigma} \hat{z}_{j\sigma}^\dagger + U \sum_i \hat{d}_i^\dagger \hat{d}_i. \tag{11}$$

The on-site Coulomb interaction is replaced by a "chemical potential" for doublons $\hat{d}_i^\dagger$. Then, the correlation among electrons is now contained as a hopping process of fermions $\hat{f}_{i\sigma}^\dagger$ disturbed by the associated motion of slave bosons. The fermion hopping is accompanied with four kinds of bosonic motions generated by $\hat{z}_{i\sigma} \hat{z}_{j\sigma}^\dagger$, namely, physical processes of hopping of holons $\hat{e}_i^\dagger$ and doublons $\hat{d}_i^\dagger$, and pair creations and annihilations of holons and doublons.

The fermion $\hat{f}_{i\sigma}$ has been interpreted to stand for the Landau's quasiparticle excitation,[58] especially in the mean-field treatment for the slave bosons, which will be discussed in the following Sec. II C. In the KR formalism, the fermions $\hat{f}_{i\sigma}$ do not interact with each other, and their motions are disturbed by the slave bosons as already mentioned above. In addition, the slave bosons just renormalize the hopping and spectral weight of the fermions $\hat{f}_{i\sigma}$ in the mean-field treatment. Thus, the correspondence between $\hat{c}_\sigma$ and $\hat{f}_\sigma$ in the KR formalism gives us a nonperturbative realization of the Landau's correspondence between an interacting many-body fermion system and a noninteracting Fermi gas. Therefore, we call the fermion $\hat{f}_{i\sigma}$ quasiparticle, hereafter.

When we employ the path integral description of the system by making use of the coherent states for bosons and fermions, we need to introduce a set of constraints to eliminate unphysical states in the expanded Hilbert space, which arise when we introduce slave bosons to describe the local Fock states. First, only one boson should occupy each local state. There are only four local physical Fock states, and these four states are exhausted by four different kinds of slave bosons. Therefore, we need the first constraint

$$\hat{e}_i^\dagger \hat{e}_i + \sum_\sigma \hat{p}_{i\sigma}^\dagger \hat{p}_{i\sigma} + \hat{d}_i^\dagger \hat{d}_i = 1. \tag{12}$$

Second, the number operator of the σ-spin quasiparticle is necessarily given as

$$\hat{f}_{i\sigma}^\dagger \hat{f}_{i\sigma} = \hat{p}_{i\sigma}^\dagger \hat{p}_{i\sigma} + \hat{d}_i^\dagger \hat{d}_i. \tag{13}$$

In the path integral form, the partition function for the Hubbard model is given by

$$Z = \int \mathcal{D}[\hat{c}_\sigma^\dagger, \hat{c}_\sigma] e^{-\mathcal{S}[\hat{c}_\sigma^\dagger, \hat{c}_\sigma]}, \tag{14}$$

where the action is

$$\mathcal{S}[\hat{c}_\sigma^\dagger, \hat{c}_\sigma] = \sum_{ij\sigma} \int_0^\beta d\tau\, \hat{c}_{i\sigma}^\dagger(\tau) [(\partial_\tau - \mu)\delta_{ij} + t_{ij}] \hat{c}_{j\sigma}(\tau) + U \sum_i \int_0^\beta d\tau\, \hat{c}_{i\uparrow}^\dagger(\tau) \hat{c}_{i\uparrow}(\tau) \hat{c}_{i\downarrow}^\dagger(\tau) \hat{c}_{i\downarrow}(\tau). \tag{15}$$

We use the same notation for both of the operators and the corresponding Grassmann fields for simplicity.

On the other hand, the partition function in the slave-boson formalism is

$$Z = \int \mathcal{D}[\hat{f}^\dagger_\sigma, \hat{f}_\sigma]\mathcal{D}[\hat{\boldsymbol{B}}^\dagger, \hat{\boldsymbol{B}}]\mathcal{D}[\boldsymbol{\lambda}]\, e^{-\mathcal{S}}. \tag{16}$$

The action is rewritten as

$$\begin{aligned}\mathcal{S} = &\int_0^\beta d\tau\, \hat{f}^\dagger_{i\sigma}(\tau)\big[(\partial_\tau - \mu + \lambda^{(2)}_{i\sigma})\delta_{ij} + \hat{\zeta}_{ij\sigma}(\tau)t_{ij}\big]\hat{f}_{j\sigma}(\tau)\\ &+\int_0^\beta d\tau\, U\hat{d}^\dagger_i(\tau)\hat{d}_i(\tau) + \int_0^\beta d\tau\, \Bigg[\sum_i \Bigg(\hat{e}^\dagger_i(\tau)\partial_\tau\hat{e}_i(\tau)\\ &+\sum_\sigma \hat{p}^\dagger_{i\sigma}(\tau)\partial_\tau\hat{p}_{i\sigma}(\tau) + \hat{d}^\dagger_i(\tau)\partial_\tau\hat{d}_i(\tau)\Bigg)\\ &+\sum_i \lambda^{(1)}_i\Bigg(\hat{e}^\dagger_i(\tau)\hat{e}_i(\tau)+\sum_\sigma \hat{p}^\dagger_{i\sigma}(\tau)\hat{p}_{i\sigma}(\tau)+\hat{d}^\dagger_i(\tau)\hat{d}_i(\tau)-1\Bigg)\\ &-\sum_{i\sigma}\lambda^{(2)}_{i\sigma}(\hat{p}^\dagger_{i\sigma}(\tau)\hat{p}_{i\sigma}(\tau)+\hat{d}^\dagger_i(\tau)\hat{d}_i(\tau))\Bigg],\end{aligned} \tag{17}$$

where $\hat{\boldsymbol{B}}_i = (\hat{e}^\dagger_i, \hat{p}^\dagger_{i\uparrow}, \hat{p}^\dagger_{i\downarrow}, \hat{d}^\dagger_i)$, $\boldsymbol{\lambda}_i = (\lambda^{(1)}_i, \lambda^{(2)}_{i\uparrow}, \lambda^{(2)}_{i\downarrow})$, and $\hat{\zeta}_{ij\sigma}(\tau) = \hat{z}_{i\sigma}(\tau)\hat{z}^\dagger_{j\sigma}(\tau)$. The Lagrange multipliers $\lambda^{(1)}_i$, $\lambda^{(2)}_{i\uparrow}$, and $\lambda^{(2)}_{i\downarrow}$ are introduced to enforce the set of constraints (12) and (13).

C. Mean-field theory for KR formalism

The mean-field approximation for Eq. (16) corresponds to replacing bosonic fields $\hat{\boldsymbol{B}}_i$ and $\boldsymbol{\lambda}_i$ with the homogeneous saddle point values for them as

$$\begin{gathered}\hat{e}^\dagger_i, \hat{e}_i \to \overline{e}_0, \quad \hat{p}^\dagger_{i\sigma}, \hat{p}_{i\sigma} \to \overline{p}_{0\sigma}, \quad \hat{d}^\dagger_i, \hat{d}_i \to \overline{d}_0,\\ \lambda^{(1)}_i \to \lambda^{(1)}, \quad \lambda^{(2)}_{i\sigma} \to \lambda^{(2)}_\sigma.\end{gathered}$$

These saddle-point values are determined self-consistently through minimizing the free energy f given in Eq. (19) below. Then, the action for the fermionic degrees of freedom contains only quadratic terms of fermionic fields $\hat{f}^\dagger_{i\sigma}$ after the slave bosons are replaced with c numbers $\overline{e}_0$, $\overline{p}_{0\sigma}$, and $\overline{d}_0$ as

$$S_0 = \int_0^\beta d\tau \sum_{ij\sigma}\hat{f}^\dagger_{i\sigma}(\tau)[(\partial_\tau - \mu)\,\delta_{ij} + \zeta_{0\sigma}t_{ij}]\hat{f}_{j\sigma}(\tau). \tag{18}$$

We can easily integrate out the remaining fermionic degrees of freedom and obtain the mean-field free energy for homogeneous phases as

$$\begin{aligned}f = &-\frac{T}{N_s}\sum_{k,\sigma}\ln[1 + e^{-\beta(\zeta_{0\sigma}\epsilon_k - \mu + \lambda^{(2)}_\sigma)}] + U\overline{d}_0^2\\ &+\lambda^{(1)}\left(\overline{e}_0^2 + \sum_\sigma \overline{p}^2_{0\sigma} + \overline{d}_0^2 - 1\right)\\ &-\sum_\sigma \lambda^{(2)}_\sigma\big(\overline{p}^2_{0\sigma} + \overline{d}_0^2\big),\end{aligned} \tag{19}$$

where N_s is the number of sites, ϵ_k is the Fourier transformation of t_{ij}, and the mean-field quasiparticle renormalization $\zeta_{0\sigma}$ is given as

$$\zeta_{0\sigma} = [g_{1\sigma}(\overline{p}_{0\sigma}\overline{e}_0 + \overline{d}_0\overline{p}_{0\overline{\sigma}})g_{2\sigma}]^2, \tag{20}$$

$$g_{1\sigma} = \big(1 - \overline{p}^2_{0\overline{\sigma}} - \overline{e}_0^2\big)^{p_1}, \tag{21}$$

$$g_{2\sigma} = \big(1 - \overline{p}^2_{0\sigma} - \overline{d}_0^2\big)^{p_2}. \tag{22}$$

It is known that, with this drastic mean-field approximation, the noninteracting limit is correctly reproduced when we set p_1 and p_2 defined in Eqs. (9) and (10) as $p_1 = p_2 = -1/2$. For $U = 0$, the mean-field values of the density of bosons are given as $\overline{d}_0^2 = (n/2)^2$, $\overline{e}_0^2 = 1 - n + \overline{d}_0^2$, and $\overline{p}^2_{0\uparrow} = \overline{p}^2_{0\downarrow} = n/2 - \overline{d}_0^2$. Then, the mean-field renormalization factor turns out to be 1, correctly. Hereafter, the values of p_1 and p_2 are fixed as $p_1 = p_2 = -1/2$.

The saddle-point values for bosonic fields $\overline{e}_0$, $\overline{p}_{0\sigma}$, and $\overline{d}_0$ can easily be examined in some well-defined limits. In the strong-coupling limit $U/|t| \gg 1$, the density of doublon is suppressed $\overline{d}_0^2 \sim 0$ at any doping level. Then, the spin density is roughly proportional to the spin-dependent electron density $\overline{p}^2_{0\sigma} \sim n_\sigma$ (recall the mean-field constraint $n_\sigma = \overline{p}^2_{0\sigma} + \overline{d}_0^2$). In the hole-doping case, the density of empty site $\overline{e}_0^2$ is given by the doping concentration $\overline{e}_0^2 \sim x$.

In this limit, the doping dependences of $\zeta_{0\sigma}$ and LUSW are given as follows: By using the fact that $n_\sigma = n/2 = (1-x)/2$ in the paramagnetic phase, the mean-field renormalization factor ζ_0 is given as

$$\begin{aligned}\zeta_{0\sigma} = \zeta_0 &= \frac{\overline{p}_0^2(\overline{e}_0 + \overline{d}_0)^2}{\frac{n}{2}(1 - \frac{n}{2})}\\ &\simeq \frac{\frac{1-x}{2}\cdot x}{\frac{1-x}{2}\cdot\frac{1+x}{2}} = \frac{2x}{1+x}.\end{aligned} \tag{23}$$

By multiplying the number density of unoccupied states $1 + x$ with ζ_0, we obtain the LUSW as $2x$. This doping dependence of the LUSW is the same as that of the exact solution in the strong-coupling limit.

D. Previous studies on charge fluctuations

1. *Formation of upper and lower Hubbard bands*

The mean-field theory only accounts for the coherent quasiparticle excitations in the correlated electron systems. Of course, momentum-dependent renormalizations do not appear. For overall description of the energy spectrum including incoherent Hubbard bands, fluctuations of the charge bosons $\hat{e}$ and $\hat{d}$, which describe dynamics of holons and doublons, are known to be essential when one wishes to improve the slave-boson mean-field theory. For example, Castellani *et al.* have claimed that the Gaussian fluctuations of charge bosons around the saddle-point solution can reproduce the structure of incoherent Hubbard bands.[59] To take into account the fluctuations of bosonic fields around the saddle-point solutions, the Bogoliubov prescription[60] is used, in which the

boson operators are divided into condensate components and fluctuating components out of condensation as

$$\hat{e}_i^\dagger = \overline{e}_0 + \widetilde{e}_i^\dagger,\ \hat{e}_i = \overline{e}_0 + \widetilde{e}_i, \tag{24}$$

$$\hat{p}_{i\sigma}^\dagger = \overline{p}_{0\sigma} + \widetilde{p}_{i\sigma}^\dagger,\ \hat{p}_{i\sigma} = \overline{p}_{0\sigma} + \widetilde{p}_{i\sigma}, \tag{25}$$

$$\hat{d}_i^\dagger = \overline{d}_0 + \widetilde{d}_i^\dagger,\ \hat{d}_i = \overline{d}_0 + \widetilde{d}_i. \tag{26}$$

Here, the fluctuating components of the charge bosons $\hat{e}_i$, $\hat{d}_i$ and spin boson $\hat{p}_{i\sigma}$ ($\widetilde{e}_i$, $\widetilde{d}_i$, and $\widetilde{p}_{i\sigma}$, respectively) describe spin and charge fluctuations around the saddle-point solution.

2. *Bosonic propagators*

Propagators of the Gaussian fluctuations of the slave bosons are given by the action $S_{\rm B}$ with quadratic terms including only the bosonic fields $\widetilde{\boldsymbol{B}}^\dagger$, $\widetilde{\boldsymbol{B}}$ as

$$S_{\rm B} = S_{\rm B}^{(0)} + S_{\rm B}^{(1c)} + S_{\rm B}^{(1s)} + S_{\rm B}^{(1cs)}, \tag{27}$$

$$\begin{aligned} S_{\rm B}^{(0)} = \int_0^\beta d\tau \sum_i \Big\{ & \widetilde{e}_i^\dagger(\tau)\big[\partial_\tau + \lambda_i^{(1)}\big]\widetilde{e}_i(\tau) \\ & + \sum_\sigma \widetilde{p}_{i\sigma}^\dagger(\tau)\big[\partial_\tau + \lambda_i^{(1)} - \lambda_{i\sigma}^{(2)}\big]\widetilde{p}_{i\sigma}(\tau) \\ & + \widetilde{d}_i^\dagger(\tau)\left[\partial_\tau + U + \lambda_i^{(1)} - \sum_\sigma \lambda_{i\sigma}^{(2)}\right]\widetilde{d}_i(\tau)\Big\}, \end{aligned} \tag{28}$$

$$S_{\rm B}^{(1c)} = \int_0^\beta d\tau \sum_{ij} \boldsymbol{\beta}_i(\tau)\boldsymbol{L}_{ij}^{\rm c}\boldsymbol{\beta}_j^\dagger(\tau), \tag{29}$$

$$\begin{aligned} S_{\rm B}^{(1s)} &= \int_0^\beta d\tau \sum_{ij} \boldsymbol{\phi}_{i\uparrow}^{\dagger\,T}(\tau)\boldsymbol{L}_{ij}^{\rm s}\boldsymbol{\phi}_{j\uparrow}(\tau)^T \\ &= \int_0^\beta d\tau \sum_{ij} \boldsymbol{\phi}_{i\downarrow}^{\dagger\,T}(\tau)\boldsymbol{L}_{ij}^{\rm s}\boldsymbol{\phi}_{j\downarrow}(\tau)^T, \end{aligned} \tag{30}$$

$$\begin{aligned} S_{\rm B}^{(1cs)} = \int_0^\beta d\tau \sum_{ij\sigma} \Big[& \boldsymbol{\beta}_i(\tau)\boldsymbol{L}_{ij\sigma}^{\rm cs}\boldsymbol{\phi}_{j\uparrow}^{\ T}(\tau) \\ & + \boldsymbol{\phi}_{i\sigma}^{\dagger\,T}(\tau)\boldsymbol{L}_{ij\sigma}^{\rm cs\ T}\boldsymbol{\beta}_j^\dagger(\tau) + \delta_{ij}K_{i\sigma}^{\rm cs} \\ & \times \{\boldsymbol{\beta}_i(\tau)\cdot\boldsymbol{\phi}_{i\sigma}^\dagger(\tau) + \boldsymbol{\phi}_{i\sigma}(\tau)\cdot\boldsymbol{\beta}_i^\dagger(\tau)\}\Big], \end{aligned} \tag{31}$$

where we use vector notations $\boldsymbol{\beta}_i = (\widetilde{e}_i, \widetilde{d}_i^\dagger)$, $\boldsymbol{\phi}_{i\sigma} = (\widetilde{p}_{i\sigma}, \widetilde{p}_{i\overline{\sigma}}^\dagger)$, and coefficients defined as

$$\boldsymbol{L}_{ij}^{\rm c} = g_{ij}\begin{pmatrix} \overline{p}_{0\sigma}^2 + \overline{p}_{0\overline{\sigma}}^2 & 2\overline{p}_{0\sigma}\overline{p}_{0\overline{\sigma}} \\ 2\overline{p}_{0\sigma}\overline{p}_{0\overline{\sigma}} & \overline{p}_{0\sigma}^2 + \overline{p}_{0\overline{\sigma}}^2 \end{pmatrix}, \tag{32}$$

$$\boldsymbol{L}_{ij}^{\rm s} = g_{ij}\begin{pmatrix} \overline{e}_0^2 + \overline{d}_0^2 & 2\overline{e}_0\overline{d}_0 \\ 2\overline{e}_0\overline{d}_0 & \overline{e}_0^2 + \overline{d}_0^2 \end{pmatrix}, \tag{33}$$

$$\boldsymbol{L}_{ij\sigma}^{\rm cs} = g_{ij}\begin{pmatrix} \overline{e}_0\overline{p}_{0\sigma} & \overline{e}_0\overline{p}_{0\overline{\sigma}} \\ \overline{d}_0\overline{p}_{0\sigma} & \overline{d}_0\overline{p}_{0\overline{\sigma}} \end{pmatrix}, \tag{34}$$

$$K_{i\sigma}^{\rm cs} = \sum_j g_{ij}(\overline{e}_0\overline{p}_{0\sigma} + \overline{d}_0\overline{p}_{0\overline{\sigma}}). \tag{35}$$

Here, the hopping parameter for bosons is given as $g_{ij} = t_{ij}g_{1\sigma}^2 g_{2\sigma}^2\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle$, where static correlation functions for quasiparticles $\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle$ are introduced. The average $\langle\cdots\rangle$ is defined as

$$\langle X\rangle = \frac{\int \mathcal{D}[\hat{f}_\sigma^\dagger, \hat{f}_\sigma] X e^{-S_0 - S_{\rm B}}}{\int \mathcal{D}[\hat{f}_\sigma^\dagger, \hat{f}_\sigma] e^{-S_0 - S_{\rm B}}}, \tag{36}$$

where $S_0 + S_{\rm B}$ is the approximate action used in this section. The term $\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle$ represents the mean field that stands for kinetic motions of quasiparticles surrounding bosons. This mean field seems to be self-consistently determined through Eq. (36). However, by using the approximate action $S_0 + S_{\rm B}$, we obtain $\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle$ as

$$\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle = \frac{\int \mathcal{D}[\hat{f}_\sigma^\dagger, \hat{f}_\sigma]\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}e^{-S_0}}{\int \mathcal{D}[\hat{f}_\sigma^\dagger, \hat{f}_\sigma]e^{-S_0}} \tag{37}$$

because $S_{\rm B}$ does not contain $\hat{f}_{i\sigma}$ and/or $\hat{f}_{i\sigma}^\dagger$. This mean field $\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle$ is obtained through decoupling the "interaction" term $\hat{z}_{i\sigma}\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\hat{z}_{j\sigma}^\dagger$ as

$$\begin{aligned} \hat{z}_{i\sigma}\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\hat{z}_{j\sigma}^\dagger \simeq\ & \hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle + \langle\hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger\rangle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma} \\ & - \langle\hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger\rangle\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle \end{aligned} \tag{38}$$

by neglecting fluctuations

$$(\hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger - \langle\hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger\rangle)(\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma} - \langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle).$$

Averages such as $\langle\hat{f}_{i\sigma}^\dagger\hat{f}_{j\sigma}\rangle$ and $\langle\hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger\rangle$ are taken by using the resultant action. However, we should note that, in the previous studies,[57,59] the average $\langle\hat{z}_{i\sigma}\hat{z}_{j\sigma}^\dagger\rangle$ was treated as ζ_0 and the contribution from static correlation functions of bosons such as $\langle\widetilde{d}_i^\dagger\widetilde{e}_j^\dagger\rangle$ was dropped. We also note that, in the mean-field level, the dispersion of spin bosons for the paramagnetic phase vanishes as $x \to 0$, because of simultaneously vanishing condensation of charge bosons: $\overline{e}_0,\ \overline{d}_0 \to 0$ in Eqs. (30) and (31).

In the close proximity to the Mott insulating states, where $\overline{e}_0, \overline{d}_0 \ll 1$, propagators for the Gaussian fluctuations of charge bosons $\boldsymbol{\beta}_i = (\beta_i^1, \beta_i^2) = (\widetilde{e}_i, \widetilde{d}_i^\dagger)$ are approximately determined by the action $S_{\rm B}^{(0)} + S_{\rm B}^{(1c)}$ as

$$-\langle\beta^a{}_Q\beta^{b\dagger}{}_Q\rangle = \frac{Z_+^{ab}(Q)}{i\omega_m - |\Lambda_Q|} - \frac{Z_-^{ab}(Q)}{i\omega_m + |\lambda_Q|}, \tag{39}$$

where the coefficients $Z_\pm^{ab}(Q)$ are given by

$$\begin{aligned} \begin{pmatrix} Z_\pm^{11}(Q) & Z_\pm^{12}(Q) \\ Z_\pm^{21}(Q) & Z_\pm^{22}(Q) \end{pmatrix} = & \frac{\delta\lambda + \delta U/2}{2\sigma_Q}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ & \pm\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} - \frac{\overline{p}_0^2|\epsilon|\epsilon_Q}{2\sigma_Q}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \end{aligned} \tag{40}$$

Parameters used in the above equations are given as

$$\delta\lambda = \lambda^{(1)} - \lambda^{(2)}, \tag{41}$$

$$\sigma_Q = \sqrt{\left(\lambda^{(1)} + \frac{\delta U}{2} - \overline{p}_0^2|\epsilon|\epsilon_Q\right)^2 - \overline{p}_0^4|\epsilon|^2\epsilon_Q^2}, \tag{42}$$

$$\Lambda_Q = \frac{\delta U}{2} + \sigma_Q, \tag{43}$$

$$\lambda_Q = -\frac{\delta U}{2} + \sigma_Q, \tag{44}$$

where $\delta U = U - 2\lambda^{(2)}$ gives the amplitude of the Hubbard gap at half-filling $n = 1$, and

$$|\epsilon| = \left| \frac{1}{N_s} \sum_k \epsilon_k / \left[1 + e^{+\beta(q\epsilon_k - \mu + \lambda^{(2)})}\right] \right|.$$

Raimondi and Castellani[57] introduced the following approximate form of the single-particle Green's function as

$$\begin{aligned} \mathcal{G}_{ij\sigma}(\tau) &= -\langle T\hat{c}_{i\sigma}(\tau)\hat{c}^\dagger_{j\sigma}(0)\rangle \\ &= -\langle T\hat{z}^\dagger_{i\sigma}(\tau)\hat{z}_{j\sigma}(0)\hat{f}_{i\sigma}(\tau)\hat{f}^\dagger_{j\sigma}(0)\rangle \\ &\simeq -\langle T\hat{z}^\dagger_{i\sigma}(\tau)\hat{z}_{j\sigma}(0)\rangle\langle T\hat{f}_{i\sigma}(\tau)\hat{f}^\dagger_{j\sigma}(0)\rangle, \end{aligned} \tag{45}$$

where vertex corrections are dropped. A further approximation was introduced: By focusing only on the charge dynamics, contributions from the fluctuations of spin bosons were neglected as

$$\begin{aligned} \langle T\hat{z}^\dagger_{i\sigma}(\tau)\hat{z}_{j\sigma}(0)\rangle \simeq \zeta_{0\sigma} &+ \overline{p}_0^2 g_{1\sigma}^2 g_{2\sigma}^2 \\ &\times \langle T\widetilde{b}_{i\sigma}(\tau)\widetilde{b}^\dagger_{j\sigma}(0)\rangle, \end{aligned} \tag{46}$$

where $\widetilde{b}_{i\sigma}(\tau) = \widetilde{e}^\dagger_{i\sigma}(\tau) + \widetilde{d}_{i\sigma}(\tau)$. As a result, the approximate Green's function is given as

$$\begin{aligned} \mathcal{G}_{ij\sigma}(\tau) \simeq &-\zeta_{0\sigma}\langle T\hat{f}_{i\sigma}(\tau)\hat{f}^\dagger_{j\sigma}(0)\rangle \\ &-\overline{p}_0^2 g_{1\sigma}^2 g_{2\sigma}^2 \langle T\widetilde{b}_{i\sigma}(\tau)\widetilde{b}^\dagger_{j\sigma}(0)\rangle \\ &\times\langle T\hat{f}_{i\sigma}(\tau)\hat{f}^\dagger_{j\sigma}(0)\rangle. \end{aligned} \tag{47}$$

The first term of the right-hand side of Eq. (47) gives the coherent band and the second term gives the incoherent Hubbard bands.

When we include the fluctuations of slave bosons, we should take care of the constraint or the local conservation of the densities of the slave bosons [Eqs. (12) and (13)]. Although the previous theories, by including the fluctuations of charge bosons, reproduce the incoherent Hubbard bands, there exists a difficulty in conserving global spectral weight in the doped systems[57] because the simple decoupling of the fermionic and bosonic degrees of freedom in Eq. (47) can violate the local conservation of the boson density Eqs. (12) and (13).

III. COFERMION THEORY

A. Perspective

In the above sections, we reviewed the mean-field KR theory and how fluctuations of charge bosons induce the incoherent bands. These previous theories have drawbacks, in spite of an advantage in the simplicity. For example, the momentum-independent quasiparticle renormalization in the previous theories can not account for the "Fermi arc" observed in ARPES measurements of the cuprate superconductors.[9] Since significant momentum-dependent quasiparticle renormalizations have been captured in numerical works through the cluster-type extension of the dynamical mean-field theory,[20–25] it is desirable to construct a theory that is physically transparent and can examine the observed singular self-energy, which results in, for example, the pseudogap and the Fermi arc. We assume that the experimentally observed arclike Fermi surface, in the hole-underdoped cuprates, is a consequence of Fermi-surface reconstruction caused by the divergence of the quasiparticle self-energy or emergence of zeros of the Green's function. The divergence means the breakdown of the perturbation theory and, hence, the breakdown of the Fermi-liquid theory as well. Here, we extend the previous theories to understand this unconventional feature. Our goal is to acquire a simple framework and an intuitive understanding.

The Fermi-surface reconstruction itself is not an unconventional phenomenon as we have discussed in Sec. I in the example of the antiferromagnetic order in the ordinary Slater's mean-field description. In an antiferromagnetic metal on square lattices with the ordering vector $Q_0 = (\pi,\pi)$, the up-spin electrons with momentum k hybridize with the down-spin electrons with momentum $k + Q_0$ through the term $\Delta_{\rm AFM}[\hat{c}^\dagger_{k\uparrow}\hat{c}_{k+Q_0\downarrow} + \text{H.c.}]$ in the presence of the mean field $\Delta_{\rm AFM}$.

When the bare band dispersion of electron $\hat{c}^\dagger_{k\sigma}$ is given by

$$\begin{aligned} \xi_k &= \epsilon_{1k} + \epsilon_{2k} - \mu, \\ \epsilon_{1k} &= -2t(\cos k_x + \cos k_y), \\ \epsilon_{2k} &= 4t' \cos k_x \cos k_y, \end{aligned}$$

the Green's function of the electron $\hat{c}^\dagger_{k\sigma}$ that hybridizes with the electron $\hat{c}^\dagger_{k+Q_0\overline{\sigma}}$ is obtained as

$$G(k,\omega) = \frac{1}{\omega - \xi_k - \dfrac{\Delta^2_{\rm AFM}}{\omega - \xi_{k+Q_0}}} \tag{48}$$

$$= \frac{\dfrac{1}{2} - \dfrac{\epsilon_{1k}}{2\sqrt{\epsilon^2_{1k} + \Delta^2_{\rm AFM}}}}{\omega - \epsilon_{2k} + \mu - \sqrt{\epsilon^2_{1k} + \Delta^2_{\rm AFM}}} + \frac{\dfrac{1}{2} + \dfrac{\epsilon_{1k}}{2\sqrt{\epsilon^2_{1k} + \Delta^2_{\rm AFM}}}}{\omega - \epsilon_{2k} + \mu + \sqrt{\epsilon^2_{1k} + \Delta^2_{\rm AFM}}}. \tag{49}$$

The Green's function written in Eq. (48) shows that $G(k,\omega)$ vanishes at $\omega - \xi_{k+Q_0} = 0$. In other words, $G(k,\omega)$ has zero at $\omega - \xi_{k+Q_0} = 0$. This emergence of zeros is nothing but the divergence of the self-energy of the electron $\hat{c}^\dagger_{k\sigma}$, $\Sigma(k,\omega) = \Delta^2_{\rm AFM}/(\omega - \xi_{k+Q_0})$ at $\omega - \xi_{k+Q_0} = 0$, as is seen in Eq. (48). As a result, the zero surface defined by $\omega = \xi_{k+Q_0}$ splits the pole surface defined by $\omega = \xi_k$ into the two pole surfaces $\omega = \epsilon_{2k} - \mu \pm \sqrt{\epsilon^2_{1k} + \Delta^2_{\rm AFM}}$, as is seen in Eq. (49). Then, the Fermi surface is disconnected into pockets at the gap edge. As an example, a zero surface, reconstructed band dispersion, and Fermi surface are depicted in Fig. 2 for $t' = 0.25t$, $\mu = -0.75t$, and $\Delta_{\rm AFM}$. This has, to some extent, a qualitative similarity to what is observed in the cuprates. Another example is the case of the BCS superconductivity where the quasiparticle $\hat{c}^\dagger_{k\uparrow}$ has a "particle-particle hybridization" proportional to $\hat{c}^\dagger_{k\uparrow}\hat{c}^\dagger_{-k\downarrow}$ with $\hat{c}^\dagger_{-k\downarrow}$.[61]

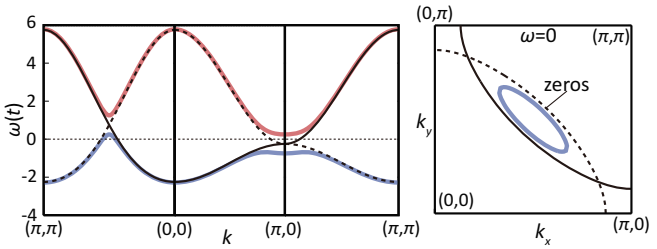

FIG. 2. (Color online) The left panel shows band dispersions and zero surface along lines running from (π,π) to (0,0), from (0,0) to $(\pi,0)$, and from $(\pi,0)$ to (0,0). The right panel shows bare and reconstructed Fermi surface and zero surface at $\omega = 0$. The thin (black) solid curve stands for bare band dispersion $\omega = \xi_k$, and the thin dashed (black) curve stands for zero surface $\omega = \xi_{k+Q_0}$. Bold (red and blue) solid curves stand for reconstructed bands.

However, as is already remarked in Sec. I, symmetry breakings such as the antiferromagnetic orders have not been universally observed in the hole-underdoped cuprates where the "Fermi arc" or the truncated Fermi surface has been experimentally suggested. Therefore, we need a mechanism for the emergence of zeros without any symmetry breakings. In this context, the truncated Fermi-surface scenario reviewed in Ref. 29 gives an interesting example inspired by a result of the renormalization-group methods, which attributes the emergence of zeros to umklapp scatterings without the symmetry breakings.[62] The pseudogap in this framework is the precursor of the Hubbard gap. We employ an alternative framework where a zero surface with a gap formation naturally emerges by a hybridization with some additional fermionic excitations, as is discussed in the above example of the antiferromagnetic ordered phase, although our additional fermions are different from the quasiparticle at $k + Q_0$ in the antiferromagnetic case. Since our pseudogap (the hybridization gap) will turn out to be different from the remnant of the Hubbard gap, our framework will yield results qualitatively different from the scenario by the umklapp scattering as we see in this paper.

Are there such fermionic excitations in proximity to Mott insulators in the absence of symmetry breakings? Or how do such fermionic excitations emerge? A hint for the existence of such additional fermionic excitations comes from examinations of the LUSW illustrated in Fig. 1 in the doped Mott insulators. Hereafter, we restrict our discussion to the hole-doped Mott insulator as in Fig. 1, with the hole-underdoped cuprates in mind. The LUSW is defined as the spectral weight above μ within the coherent band near the top of the LHB.

We first recall the origin of the LUSW discussed in the literature.[36,63] First, we begin with the atomic limit, where $t = 0$ and $U \neq 0$. Then, the LUSW of the hole-doped Mott insulator consists only of states created by adding an σ-spin electron to an empty site, namely, a holon site, to avoid creating a doubly occupied site, doublon, and to avoid the cost of the on-site Coulomb repulsion U. An electron added to an empty site is confined tightly in this limit. If the electron escapes from the holon sites, it inevitably creates a doublon and costs U. In other words, an electron created at a holon site can contribute to the LUSW, although this electron does not propagate coherently.

When t/U becomes nonzero, we have tightly bound doublons and holons even in the Mott insulating phase. Therefore, holons in the hole-doped systems consist of both of the doped holons and the preexisting holons already present in the insulators. Originally, the binding energy of a doublon and a holon in the Mott insulator is the order of U. However, for a nonzero t and a nonzero doping x, quantum fluctuations induced by coherent carriers dramatically weaken bindings between a doublon and a holon. Then, an electron added to this holon only weakly bound to the doublon requires only small energy and merges into the excitation of an added electron to a doped holon site. In the hole-doped systems, these two types of holons should not be distinguished and should constitute the same object. Then, an electron added to these weakly bound holons should also constitute LUSW. This electron added to these holons may constitute a novel composite particle distinguished from the original electron and offers a hint for the additional fermionic excitations, which bring about zeros to the quasiparticle Green's functions, if this composite particle hybridizes with an original quasiparticle. Actually, remnants of doublon-holon pairs, namely, weakly bound doublon-holon pairs, are known to play an important role in correlated electron systems, especially in the context of variational wave-function theories.[64,65]

In the KR formalism, creation (annihilation) operators for the electron at the holon site, or the tightly bound doublon-holon pair, are given by composite *fermion* operators $\hat{e}_i \hat{f}^\dagger_{i\sigma}$ $(\hat{f}_{i\sigma}\hat{e}^\dagger_i)$. This is just a σ-spin electron in the original Hubbard model and is definitely fermionic. A part of the tightly bound pair $\overline{e}_0 \hat{f}^\dagger_{i\sigma}$ has already been taken into account in the mean-field level in the previous theories (note $\hat{e}_i \hat{f}^\dagger_{i\sigma} = \overline{e}_0 \hat{f}^\dagger_{i\sigma} + \widetilde{e}_i \hat{f}^\dagger_{i\sigma}$). However, it represents nothing but the renormalized quasiparticle, which propagates in homogeneous mean fields and has nothing to do with the above composite particle. To substantiate our picture, we need to take into account the composite operator including bosonic fluctuations $\widetilde{e}_i \hat{f}^\dagger_{i\sigma}$. If we treat such a composite operator $\widetilde{e}_i \hat{f}^\dagger_{i\sigma}$ as a fermion, overlaps between tightly bound doublon-holon pairs and quasiparticle states cause a hybridization of fermions between two types, where one is the quasiparticle $\hat{f}^\dagger_{i\sigma}$ and the other is the "composite fermion" or "cofermion" $\widetilde{e}_i \hat{f}^\dagger_{i\sigma}$. From such a hybridization between the cofermion and quasiparticle, weakly bound pairs discussed above will naturally be born as a result.

If such a hybridization between the quasiparticle and the composite fermion really exists, this hybridization would contribute to the self-energy of quasiparticles. Depending on the dynamics of the composite fermions, such self-energy would have a strong momentum dependence and possibly have poles near the zero energy. In such cases, the hybridization between the quasiparticles and the composite particles causes a hybridization gap near the Fermi level. This gap is expected to account for the pseudogap phenomena. As is discussed below, we construct an action containing both of the quasiparticles and the composite fermions.

In the following sections, we concretely give our theoretical treatment by introducing the composite fermions discussed above. First, we recall shortcomings of the previous theories

and present our guiding principle to overcome the previous KR mean-field theory.

As is mentioned in the above section, the KR formalism gives exact results if we thoroughly treat the fluctuations of slave bosons and keep the constraint equations (12) and (13) (hereafter, we call this constraint the local conservation). When we include the fluctuations of slave bosons, we should take care of the constraint or the local conservation of the densities of the slave bosons, and avoid including unphysical processes.

When we impose the local constraints more strictly for fluctuating bosons beyond the mean-field level, as discussed below, it turns out that the term

$$\hat{\zeta}^{(1)}_{ij\sigma} = g^2_{1\sigma} g^2_{2\sigma} (\tilde{p}^\dagger_{i\sigma} \tilde{e}_i + \tilde{d}^\dagger_i \tilde{p}_{i\bar{\sigma}})(\tilde{e}^\dagger_j \tilde{p}_{j\sigma} + \tilde{p}^\dagger_{j\bar{\sigma}} \tilde{d}_j) \tag{50}$$

represented by the diagram in Fig. 3(a) is dominating among all the possible diagrams for the kinetic term in the action, Eq. (17), namely,

$$\int_0^\beta d\tau \sum_{ij} \hat{f}^\dagger_{i\sigma}(\tau) t_{ij} \hat{\zeta}_{ij\sigma}(\tau) \hat{f}_{j\sigma}(\tau). \tag{51}$$

Here, we employ $g^2_{1\sigma} = (1 - \overline{p}^2_{0\bar{\sigma}} - \overline{e}^2_0)^{-1}$ and $g^2_{1\sigma} = (1 - \overline{p}^2_{0\sigma} - \overline{d}^2_0)^{-1}$ by following Ref. 56.

To elucidate why we retain $\hat{\zeta}^{(1)}_{ij\sigma}$, we classify the diagrams illustrated in Fig. 3 into four types, categorized by the frequency dependence of quasiparticles and the local conservation of the boson densities:

T-1. Diagrams containing external propagators of quasiparticles in addition to bosonic propagators violating the local conservation of boson densities [Figs. 3(b)–3(d) and 3(h)]. Here, the violation means that before and after the interactions (represented by hexagons), the number of bosons expressed by external boson propagators is not the same.

T-2. Diagrams that contain time dependence of quasiparticles, but that do not violate the local conservation [Figs. 3(a) and 3(e)].

T-3. Diagrams that do not contain frequency dependence of quasiparticles but do violate the local conservation [Fig. 3(f)]. Here the quasiparticles contribute by forming a fermionic loop without momentum and frequency transfers between quasiparticles and slave bosons. In this process, quasiparticles do not show frequency dependence and just behave as frequency-independent mean fields acting on the bosons.

T-4. Diagrams that neither include time dependence of quasiparticles nor violate the local conservation [Fig. 3(g)].

Here we present our guiding principle to take account of boson fluctuations: We exclude (T-1) because it violates the local conservation when the quasiparticles dynamically fluctuate. On the other hand, we retain diagrams belonging to the categories (T-2), (T-3), and (T-4). The reason to retain these diagrams is as follows. The diagrams in the category (T-2) do not violate the local conservation when bosons fluctuate. Therefore, we take the diagrams in this category into account. On the other hand, the diagrams in the category (T-3) do violate the local conservation. However, in these diagrams, quasiparticles enter as fermionic loops, which behave as frequency-independent mean fields. The real violation of the local conservation occurs only when a dynamical quasiparticle process is induced by fluctuating boson hoppings with the momentum and/or frequency transfers. On the contrary, the real violation does not occur when the quasiparticles emerge as the static mean fields as in the case of (T-3). This is the reason to retain the diagrams in the category (T-3). Since (T-4) does not violate local conservation, we retain it.

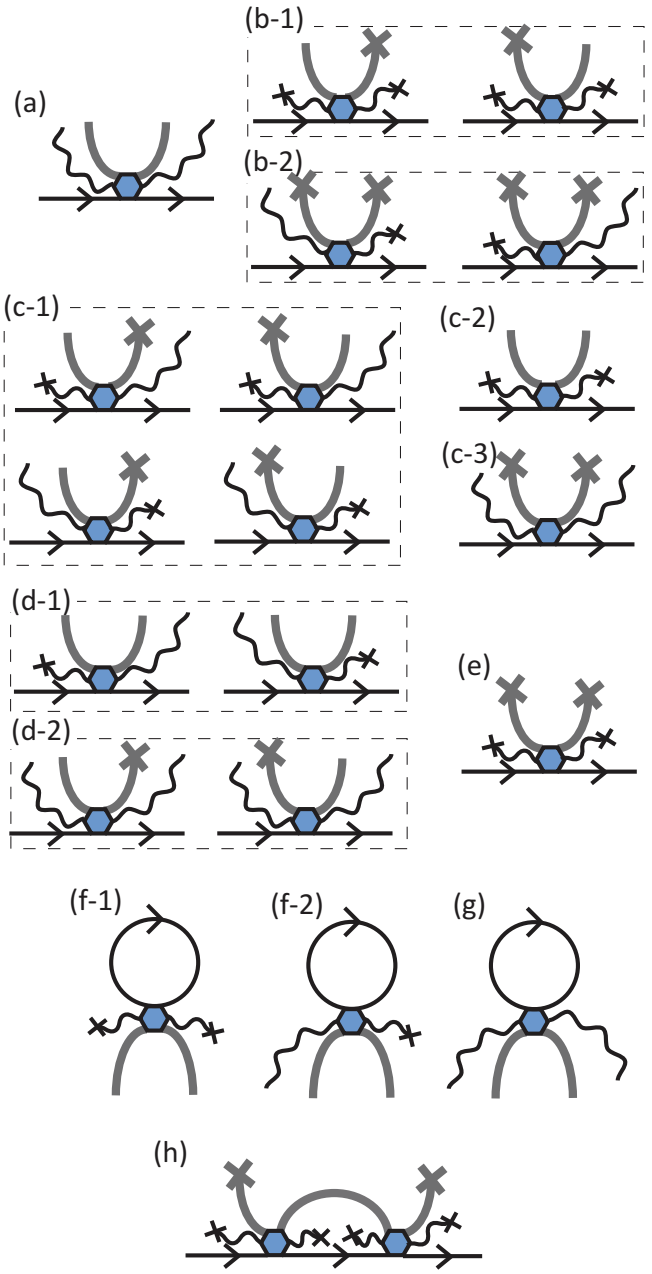

FIG. 3. (Color online) (a)–(e): Diagrams representing terms in $\hat{f}^\dagger_{i\sigma} \hat{\zeta}_{ij} \hat{f}_{j\sigma}$. (f)–(h): Examples of various diagrams on the one-loop level generated from the coupling (polygons) and derived from the terms (a)–(e). Solid lines with arrows represent propagators of the quasiparticles. Wavy lines stand for the charge bosons and bold solid lines are the spin bosons. Condensations of bosons are represented by lines terminated at crosses. Coupling constant $g^2_{1\sigma} g^2_{2\sigma} t_{ij}$ is represented by filled polygons. Here, we do not distinguish holons and doublons. Spins are also not distinguished in the diagram, for simplicity.

For the slave-particle formalism of correlated fermion systems, it is well known that fluctuations of gauge fields

play an important role in reinforcing the local constraint imposed on slave particles.[66] It was pointed out by Jolicoeur and Le Guillou that the Kotliar-Ruckenstein formalism has the U(1)×U(1)×U(1) gauge symmetry.[67] It comes from the phase symmetry of the slave-bosonic particles, namely, $\hat{e}_i$, $\hat{p}_{i\sigma}$, and $\hat{d}_i$.

In our theory, we treat fluctuations of such phases together with fluctuations of the amplitude of the condensation fraction of these slave particles by using the Bogoliubov prescription. Therefore, the phase fluctuations are taken into account, although the U(1)×U(1)×U(1) gauge structure is not strictly conserved.

B. Stratonovich-Hubbard transformation

We introduce Grassmannian variables (or fermionic fields) $\hat{\boldsymbol{\Upsilon}}_{i\sigma} = (\hat{\psi}_{i\sigma}, \hat{\chi}_{i\sigma})^T$ that stand for the cofermions as are discussed conceptually in Sec. III A by using the following identity:

$$\int \prod_{i\sigma} d\hat{\boldsymbol{\Upsilon}}^\dagger_{i\sigma} d\hat{\boldsymbol{\Upsilon}}_{i\sigma} e^{\mathcal{A}} = \det[\widetilde{\boldsymbol{T}}_\uparrow \widetilde{\boldsymbol{T}}_\downarrow], \tag{52}$$

where matrices $\widetilde{\boldsymbol{T}}_\sigma$ are defined as

$$(\widetilde{\boldsymbol{T}}_\sigma)_{ij} = g^2_{1\sigma} g^2_{2\sigma} t_{ij} \begin{bmatrix} \widetilde{p}^\dagger_{i\sigma} \widetilde{p}_{j\sigma} & \widetilde{p}^\dagger_{i\sigma} \widetilde{p}^\dagger_{j\bar{\sigma}} \\ \widetilde{p}_{i\bar{\sigma}} \widetilde{p}_{j\sigma} & \widetilde{p}_{i\bar{\sigma}} \widetilde{p}^\dagger_{j\bar{\sigma}} \end{bmatrix}, \tag{53}$$

and

$$\mathcal{A} = \int_0^\beta d\tau \sum_{ij\sigma} [(\hat{\boldsymbol{\Upsilon}}^\dagger_{i\sigma}(\tau) - \hat{\boldsymbol{C}}^\dagger_{i\sigma}(\tau)) \times (\widetilde{\boldsymbol{T}}_\sigma)_{ij} (\hat{\boldsymbol{\Upsilon}}_{j\sigma}(\tau) - \hat{\boldsymbol{C}}_{j\sigma}(\tau))]. \tag{54}$$

Here, we use vector notations as

$$\hat{\boldsymbol{C}}^\dagger_{i\sigma} = (\widetilde{e}_i, \widetilde{d}^\dagger_i) \hat{f}^\dagger_{i\sigma}, \quad \hat{\boldsymbol{C}}_{i\sigma} = \hat{f}_{i\sigma} (\widetilde{e}^\dagger_i, \widetilde{d}_i)^T. \tag{55}$$

The identity Eq. (52) gives the transformation for a coupling term of the quasiparticles and fluctuating bosons depicted in Fig. 3(a),

$$\mathcal{L}_{\mathrm{a}} = \sum_{ij\sigma} \hat{\boldsymbol{C}}^\dagger_{i\sigma}(\tau) (\widetilde{\boldsymbol{T}}_\sigma)_{ij} \hat{\boldsymbol{C}}_{j\sigma}(\tau), \tag{56}$$

as

$$\exp\left[- \int_0^\beta d\tau \mathcal{L}_{\mathrm{a}} \right] = \frac{\int \prod_{i\sigma} d\hat{\boldsymbol{\Upsilon}}^\dagger_{i\sigma} d\hat{\boldsymbol{\Upsilon}}_{i\sigma} e^{-\int_0^\beta d\tau \mathcal{L}'_a}}{\det[\widetilde{\boldsymbol{T}}_\uparrow \widetilde{\boldsymbol{T}}_\downarrow]}, \tag{57}$$

where

$$\mathcal{L}'_a = \mathcal{L}'_{a1} + \mathcal{L}'_{a2}, \tag{58}$$

$$\mathcal{L}'_{a1} = \sum_{ij\sigma} \hat{\boldsymbol{\Upsilon}}^\dagger_{i\sigma}(\tau) (\widetilde{\boldsymbol{T}}_\sigma)_{ij} \hat{\boldsymbol{\Upsilon}}_{j\sigma}(\tau), \tag{59}$$

$$\mathcal{L}'_{a2} = - \sum_{ij\sigma} \{ \hat{\boldsymbol{C}}^\dagger_{i\sigma}(\tau) (\widetilde{\boldsymbol{T}}_\sigma)_{ij} \hat{\boldsymbol{\Upsilon}}_{j\sigma}(\tau) + \hat{\boldsymbol{\Upsilon}}^\dagger_{i\sigma}(\tau) \widetilde{\boldsymbol{T}}_{ij} \hat{\boldsymbol{C}}_{j\sigma}(\tau) \}. \tag{60}$$

These transformed Lagrangians $\mathcal{L}'_{a1}$ [Fig. 4(a)] and $\mathcal{L}'_{a2}$ [Fig. 4(b)] lead to the cofermions' self-energy and the hybridization between the quasiparticles and cofermions, respectively, after integrating out the fluctuating bosonic degrees of freedom. It will be discussed below by using a set of the Dyson equations.

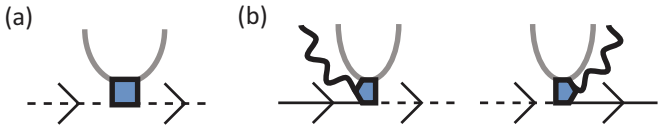

FIG. 4. (Color online) Diagrams for transformed Lagrangians defined in Eqs. (58)–(60). The solid and dashed lines with arrows represent propagators of the quasiparticles and cofermions, respectively. The wavy lines represent the charge bosons and bold solid lines are the spin bosons. (a) The diagram represents the Lagrangian $\mathcal{L}'_{a1}$ [Eq. (59)]. (b) The diagrams stand for terms in the Lagrangian $\mathcal{L}'_{a2}$ [Eq. (60)].

C. Prescription for self-consistent procedure and Green's functions

Here, we construct approximated Green's functions for the Gaussian fluctuations of the bosons, quasiparticles, and cofermions by using a set of Dyson equations as is depicted in Fig. 5: Wavy lines and bold gray lines stand for the Green's functions of the charge bosons $\mathcal{A}^{ab}(r,\tau) = -\langle T \beta^a_i(\tau) \beta^{b\dagger}_j(0) \rangle$

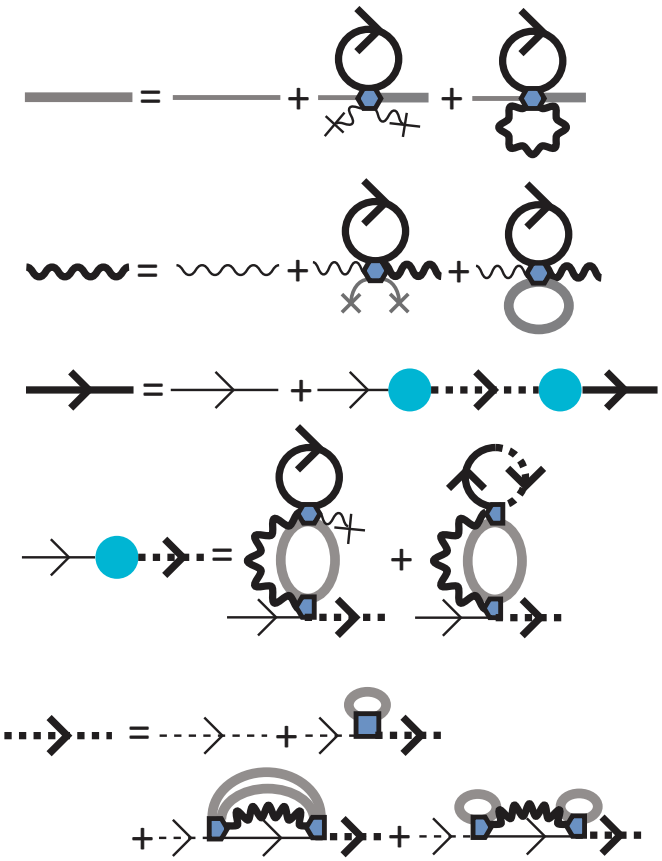

FIG. 5. (Color online) Diagrams for Dyson equations. The solid and dashed lines with arrows represent propagators of the quasiparticles and cofermions, respectively. Other notations are the same as Fig. 3. The bold (thin) lines are for renormalized (bare) propagators. Filled (blue) circles are amplitudes of the hybridization between quasiparticles and cofermions.

and the spin bosons $\mathcal{C}^{ab}(r,\tau) = -\langle T\phi_i^a(\tau)\phi_j^{b\dagger}(0)\rangle$, respectively, where $a,b = 1,2$, $r = i-j$, $(\beta_i^1,\beta_i^2) = (\widetilde{e}_i,\widetilde{d}_i)$, and $(\phi_i^1,\phi_i^2) = (\widetilde{p}_{i\sigma},\widetilde{p}_{i\bar{\sigma}}^\dagger)$. Bold lines with arrows represent the quasiparticle propagator $\mathcal{G}_\sigma^{(f)}(r,\tau)$. On the other hand, thin wavy lines and thin lines represent bare propagators of the charge bosons $\mathcal{A}_0^{ab}(r,\tau)$ and the spin bosons $\mathcal{C}_0^{ab}(r,\tau)$, respectively, determined by $\hat{\mathcal{L}}_{\rm B}^{(0)}$, in which self-energy effects are not taken into account. Thin lines with arrows stand for bare propagators of the quasiparticles $\mathcal{G}_{0\sigma}^{(f)}(r,\tau)$ determined by

$$\hat{\mathcal{L}}_0 = \sum_{ij} \hat{f}_{i\sigma}^\dagger(\tau)[(\partial_\tau - \mu)\delta_{ij} + \zeta_{0\sigma}t_{ij}]\hat{f}_{j\sigma}(\tau), \tag{61}$$

where $\zeta_{0\sigma} = g_{1\sigma}^2 g_{2\sigma}^2(\overline{p}_{0\sigma}\overline{e}_0 + \overline{d}_0\overline{p}_{0\bar{\sigma}})^2$. The Lagrangian $\hat{\mathcal{L}}_0$ is obtained by decoupling the fluctuating bosons from the KR action [see Eq. (17)]. Bold and thin dashed lines stand for the cofermions' propagators $\mathcal{F}^{ab}$ and bare propagators $\mathcal{F}_0^{ab} = \delta_{a,b}/\epsilon$ ($\epsilon \to 0$), respectively.

In the set of Dyson equations (Fig. 5, we neglect the coupling between charge and spin bosons described by propagators such as $\langle\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i\rangle$, at the Gaussian level, since these coupling terms are higher-order contributions. Below, we explain that the coupling gives higher-order contributions with respect to the hole-doping rate x, in proximity to Mott insulating states: Since operators including both charge and spin such as $\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i$ do not conserve the electric charge, propagators such as $\langle\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i\rangle$ should vanish in the Mott insulating phase, where the charge can not fluctuate. Therefore, the charge and spin excitations are well separated in the Mott insulating phase.

When hole carriers are doped, $\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i$ can have a nonzero expectation value, at most, scaled by the condensate fraction of holons $\overline{e}_0$, which gives a rough estimate of the amplitude of charge fluctuations. From a relation $\overline{e}_0^2 \propto x$ held in the KR theory for the hole-doped case, we obtain $\langle\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i\rangle \propto \sqrt{x}$. Furthermore, there is an additional constraint for the coupling terms such as $\langle\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i\rangle$: they do not appear alone in calculations of physical quantities. To conserve charge and spin on average, $\langle\widetilde{p}_{i\sigma}^\dagger\widetilde{e}_i\rangle$ appears with $\langle\widetilde{e}_i^\dagger\widetilde{p}_{i\sigma}\rangle$ in pair, for example. Therefore, the contribution of the coupling between charge and spin bosons to physical quantities is scaled by $(\sqrt{x})^2$. It concludes that the coupling between the charge and spin bosons gives contributions as a higher order in terms of x in physical quantities.

By solving the set of Dyson equations, we obtain the propagators for the quasiparticles and cofermions. Here, the bosonic degrees of freedom are taken into account in a self-consistent fashion through the cofermion self-energy $\mathbf{\Sigma}_\sigma^{(\rm cf)}(r,\tau)$ and the amplitude $\mathbf{\Delta}_{ij}$ of hybridization between the quasiparticles and cofermions, each of which we detail below.

The Lagrangian for the cofermions is given by

$$\hat{\mathcal{L}}_{\rm cf} = -\sum_{ij\sigma} \hat{\mathbf{\Upsilon}}_{i\sigma}^\dagger(\tau)\left[\mathbf{\Sigma}_\sigma^{(\rm cf)}(r,\tau)\right]\hat{\mathbf{\Upsilon}}_{j\sigma}(\tau), \tag{62}$$

where $\hat{\mathbf{\Upsilon}}_{i\sigma}^\dagger = (\hat{\psi}_{i\sigma}^\dagger,\hat{\chi}_{i\sigma}^\dagger)$ is a vector notation for the cofermions, and $r = i-j$. The cofermion self-energy $\mathbf{\Sigma}_\sigma^{(\rm cf)}(r,\tau)$ is a 2×2 symmetric matrix

$$\mathbf{\Sigma}_\sigma^{(\rm cf)} = \begin{bmatrix} \Sigma_\sigma^{11} & \Sigma_\sigma' \\ \Sigma_\sigma' & \Sigma_\sigma^{22} \end{bmatrix}. \tag{63}$$

The details for the self-energy matrix are given in Appendix A. On the other hand, the hybridization between the quasiparticles and cofermions is described by

$$\hat{\mathcal{L}}_{\rm hyb} = \sum_{i,j,\sigma}[\hat{\mathbf{\Upsilon}}_{i\sigma}^\dagger(\tau)\mathbf{\Delta}_{ij}\hat{f}_{j\sigma}(\tau) + \hat{f}_{i\sigma}^\dagger(\tau)\mathbf{\Delta}_{ij}^T\hat{\mathbf{\Upsilon}}_{j\sigma}(\tau)], \tag{64}$$

where $\mathbf{\Delta}_{ij}^T = (\Delta_{ij}^{(\psi)},\Delta_{ij}^{(\chi)})$. For details for $\mathbf{\Delta}_{ij}$, see Appendix B.

As a result, the effective Lagrangian for the quasiparticles and cofermions $\hat{\mathcal{L}}_{\rm eff}$ is given as

$$\hat{\mathcal{L}}_{\rm eff} = \hat{\mathcal{L}}_0 + \hat{\mathcal{L}}_{\rm cf} + \hat{\mathcal{L}}_{\rm hyb}. \tag{65}$$

When the charge gap is relatively small, $\Sigma_\sigma^{11} \simeq \Sigma_\sigma^{22}$ and $\Delta_{ij}^{(\psi)} \simeq \Delta_{ij}^{(\chi)}$ hold approximately. When the charge gap collapses, $\Sigma_\sigma^{11} = \Sigma_\sigma^{22}$ and $\Delta_{ij}^{(\psi)} = \Delta_{ij}^{(\chi)}$ hold exactly. In our results, we employ approximate relations $\Sigma = \Sigma_\sigma^{11} \simeq \Sigma_\sigma^{22}$ and $\Delta_{ij} = \Delta_{ij}^{(\psi)} \simeq \Delta_{ij}^{(\chi)}$. Then, a cofermion mode $(\hat{\psi}_{k\sigma} + \hat{\chi}_{k\sigma})/\sqrt{2}$ hybridizes with quasiparticles through the amplitude $\Delta(k)$, which is depicted in Fig. 5 as closed (blue) circles, where k is a momentum. The inverse of cofermion propagator (namely, the cofermion self-energy) for $(\hat{\psi}_{k\sigma} + \hat{\chi}_{k\sigma})/\sqrt{2}$ is given as

$$\tfrac{-1}{2}[\Sigma_\sigma(k,i\varepsilon_n) + \Sigma_\sigma'(k,i\varepsilon_n)] = \gamma_k i\varepsilon_n - \alpha_k + O\left(\varepsilon_n^2\right), \tag{66}$$

where ε_n is a fermionic Matsubara frequency.

Then, the Fourier transformation of the Green's function for the quasiparticles $\mathcal{G}_{ij\sigma}^{(f)}(\tau) = -\langle T\hat{f}_{i\sigma}(\tau)\hat{f}_{j\sigma}^\dagger(0)\rangle$ is given as

$$\begin{aligned}\mathcal{G}_\sigma^{(f)}(k,i\varepsilon_n \to \omega + i\delta) &= G_\sigma^{(f)}(k,\omega) \\ &\simeq \left[\omega + i\delta - \zeta_{0\sigma}\epsilon_k + \mu - \frac{\Delta(k)^2}{\gamma_k(\omega+i\delta) - \alpha_k}\right]^{-1},\end{aligned} \tag{67}$$

where ϵ_k is the Fourier transformation of t_{ij} and μ is the chemical potential. Here, we note that the weights of the two quasiparticle bands split by the zero surface defined by $\omega = \alpha_k/\gamma_k$ are not the same in our theory.

In our calculations, we define the doping rate x by using the quasiparticle Green's function as

$$1 - x = \lim_{T\to0+}\frac{T}{N_s}\sum_{k,i\varepsilon_n,\sigma}\mathcal{G}_\sigma^{(f)}(k,i\varepsilon_n), \tag{68}$$

where T stands for temperature and N_s is the number of sites.

The Green's function for the electrons, instead of the quasiparticles, is given as

$$\begin{aligned}\mathcal{G}_{ij\sigma}(\tau) &= -\langle T\hat{c}_{i\sigma}(\tau)\hat{c}_{j\sigma}^\dagger(0)\rangle \\ &\simeq -\langle T\hat{z}_{i\sigma}^\dagger(\tau)\hat{z}_{j\sigma}(0)\rangle \\ &\quad\times\langle T\hat{f}_{i\sigma}(\tau)\hat{f}_{j\sigma}^\dagger(0)\rangle \\ &= \langle T\hat{z}_{i\sigma}^\dagger(\tau)\hat{z}_{j\sigma}(0)\rangle\mathcal{G}_{ij\sigma}^{(f)}(\tau),\end{aligned} \tag{69}$$

where the bosonic and fermionic degrees of freedom are decoupled because the resultant action in our theory does not contain the hybridization between bosons and fermions. The bosonic part in Eq. (69) is given by

$$\langle T\hat{z}^{\dagger}_{i\sigma}(\tau)\hat{z}_{j\sigma}(0)\rangle \simeq g^2_{1\sigma}g^2_{2\sigma}\langle T[\hat{\boldsymbol{b}}^{\dagger}_i(\tau)\cdot\hat{\boldsymbol{p}}_{i\sigma}(\tau)] \times[\hat{\boldsymbol{p}}^{\dagger}_{j\sigma}(\tau)\cdot\hat{\boldsymbol{b}}_j(\tau)]\rangle, \tag{70}$$

where we use vector notation as $\hat{\boldsymbol{b}}^{\dagger}_i=(\hat{e}^{\dagger}_i,\hat{d}_i)$, $\hat{\boldsymbol{p}}^{\dagger}_{i\sigma}=(\hat{p}^{\dagger}_{i\sigma},\hat{p}_{i\bar{\sigma}})$. Because we adopt the boson dynamics in which charge and spin bosons are decoupled, this bosonic part of the Green's function is rewritten as

$$\begin{aligned}
&\langle T\hat{z}^{\dagger}_{i\sigma}(\tau)\hat{z}_{j\sigma}(0)\rangle \\
&\simeq g^2_{1\sigma}g^2_{2\sigma}\langle T\left[\overline{\boldsymbol{b}}_0\cdot\overline{\boldsymbol{p}}^T_{0\sigma}\right][\overline{\boldsymbol{p}}_{0\sigma}\cdot\overline{\boldsymbol{b}}^T_0]\rangle \\
&\quad + g^2_{1\sigma}g^2_{2\sigma}\langle T\left[\widetilde{\boldsymbol{b}}^{\dagger}_i(\tau)\cdot\overline{\boldsymbol{p}}^T_{0\sigma}\right][\overline{\boldsymbol{p}}_{0\sigma}\cdot\widetilde{\boldsymbol{b}}_j(\tau)]\rangle \\
&\quad + g^2_{1\sigma}g^2_{2\sigma}\langle T\left[\overline{\boldsymbol{b}}_0\cdot\widetilde{\boldsymbol{p}}_{i\sigma}(\tau)\right][\widetilde{\boldsymbol{p}}^{\dagger}_{j\sigma}(\tau)\cdot\overline{\boldsymbol{b}}^T_0]\rangle \\
&\quad + g^2_{1\sigma}g^2_{2\sigma}\langle T[\widetilde{\boldsymbol{b}}^{\dagger}_i(\tau)\cdot\widetilde{\boldsymbol{p}}_{i\sigma}(\tau)][\widetilde{\boldsymbol{p}}^{\dagger}_{j\sigma}(\tau)\cdot\widetilde{\boldsymbol{b}}_j(\tau)]\rangle,
\end{aligned} \tag{71}$$

where $\overline{\boldsymbol{b}}_0=(\overline{e}_0,\overline{d}_0)$ and $\overline{\boldsymbol{p}}_{0\sigma}=(\overline{p}_{0\sigma},\overline{p}_{0\bar{\sigma}})$. If we retain only the first and second lines of the right-hand side of Eq. (71), the electron Green's function is reduced to that already obtained in Ref. 57. The contribution of the fourth line in the right-hand side of Eq. (71) is small compared with these from other lines in Eq. (71) because it is the fourth order in the terms of the fluctuation denoted by tildes, and we ignore the fourth term.

The present theory inherits a difficulty of the previous theory[57] on the Gaussian fluctuations of the slave bosons (see Sec. II) and still violates conservation of the total spectral weight of the coherent and incoherent bands as the previous theories do. Unfortunately, the formation of the cofermions does not affect the spectrum of high-energy excitations covering both LHB and UHB, which is crucial in the conservation of the global spectral weight. A practical way to enforce the conservation is a self-consistent renormalization of the "coupling constant" $g^2_{1\sigma}g^2_{2\sigma}$ in Eq. (71), as is already discussed in Ref. 57.

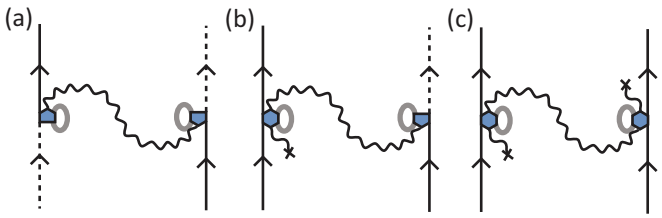

FIG. 6. (Color online) Interactions between quasiparticles and cofermions mediated by exchanging one charge boson.

D. Superconductivities mediated by cofermions and charge bosons

In this section, we examine how the present insight from the cofermion excitations offers the mechanism for the emergence of superconductivity. Here, we examine the singlet superconductivity induced by exchanging one charge boson (see Fig. 6).

We start from the mean-field action for the singlet superconductivity. Here, the effective interaction of quasiparticles and cofermions is induced by exchanging one charge boson. The mean-field action $S_{\rm SC}$ is defined as

$$S_{\rm SC}=\sum_{k,i\varepsilon_n}\boldsymbol{\Phi}^{\dagger}_k(i\varepsilon_n)\left[\boldsymbol{\mathcal{G}}(k,i\varepsilon_n)\right]^{-1}\boldsymbol{\Phi}_k(i\varepsilon_n), \tag{72}$$

where $\boldsymbol{\Phi}^{\dagger}_k(i\varepsilon_n)$ and $\boldsymbol{\Phi}_k(i\varepsilon_n)$ are defined as

$$\boldsymbol{\Phi}^{\dagger}_k(i\varepsilon_n)=[\hat{f}^{\dagger}_{k\uparrow}(i\varepsilon_n),\hat{\psi}^{\dagger}_{k\uparrow}(i\varepsilon_n)+\hat{\chi}^{\dagger}_{k\uparrow}(i\varepsilon_n),\hat{f}_{-k\downarrow}(-i\varepsilon_n),\hat{\psi}_{-k\downarrow}(-i\varepsilon_n)+\hat{\chi}_{-k\downarrow}(-i\varepsilon_n)], \tag{73}$$

$$\boldsymbol{\Phi}_k(i\varepsilon_n)=[\hat{f}_{k\uparrow}(i\varepsilon_n),\hat{\psi}_{k\uparrow}(i\varepsilon_n)+\hat{\chi}_{k\uparrow}(i\varepsilon_n),\hat{f}^{\dagger}_{-k\downarrow}(-i\varepsilon_n),\hat{\psi}^{\dagger}_{-k\downarrow}(-i\varepsilon_n)+\hat{\chi}^{\dagger}_{-k\downarrow}(-i\varepsilon_n)]^T. \tag{74}$$

The inverse of the matrix form of the Green's function is given by

$$\begin{aligned}
[\boldsymbol{\mathcal{G}}(k,i\varepsilon_n)]^{-1} &= \begin{bmatrix} i\varepsilon_n-\xi_k & \Delta_k & M_k & \Lambda_k \\ \Delta_k & \gamma_k i\varepsilon_n-\alpha_k & \Lambda_k & 0 \\ M_k & \Lambda_k & i\varepsilon_n+\xi_k & -\Delta_k \\ \Lambda_k & 0 & -\Delta_k & \gamma_k i\varepsilon_n+\alpha_k \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{\gamma_k} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \sqrt{\gamma_k} \end{bmatrix} \begin{bmatrix} i\varepsilon_n-\xi_k & \widetilde{\Delta}_k & M_k & \widetilde{\Lambda}_k \\ \widetilde{\Delta}_k & i\varepsilon_n-\widetilde{\alpha}_k & \widetilde{\Lambda}_k & 0 \\ M_k & \widetilde{\Lambda}_k & i\varepsilon_n+\xi_k & -\widetilde{\Delta}_k \\ \widetilde{\Lambda}_k & 0 & -\widetilde{\Delta}_k & i\varepsilon_n+\widetilde{\alpha}_k \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{\gamma_k} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \sqrt{\gamma_k} \end{bmatrix},
\end{aligned} \tag{75}$$

where $\xi_k=\zeta_0\epsilon_k-\mu$, and M_k and Λ_k are superconducting order parameters as is explicitly given in Eqs. (88) and (89). Here, the superconducting order parameters are given as

$$M_k(i\varepsilon_n)=\frac{T}{N_s}\sum_{i\omega_m,Q}\widetilde{t}_k\widetilde{t}_{k+Q}\langle\widetilde{b}^{\dagger}_Q(i\omega_m)\widetilde{b}_Q(i\omega_m)\rangle\left[\left(\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}\right)^2[\boldsymbol{\mathcal{G}}(k+Q,i\varepsilon_n+i\omega_m)]_{13}+\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}[\boldsymbol{\mathcal{G}}(k+Q,i\varepsilon_n+i\omega_m)]_{14}\right] \tag{76}$$

and

$$\Lambda_k(i\varepsilon_n) = \frac{T}{N_s}\sum_{i\omega_m,Q}\widetilde{t}_k\widetilde{t}_{k+Q}\langle\widetilde{b}_Q^\dagger(i\omega_m)\widetilde{b}_Q(i\omega_m)\rangle\left[[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{14}+\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}\left[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)\right]_{13}\right], \tag{77}$$

where $[\mathcal{G}(k+Q,i\varepsilon_n]_{ij}$ is the (i,j)th element of the matrix $\mathcal{G}$. We should note that there are quasiparticle-quasiparticle and cofermion-quasiparticle singlet pairings, which are described by the following anomalous Green's functions:

$$[\mathcal{G}(k,i\varepsilon_n)]_{13} = \frac{2\widetilde{\alpha}_k\widetilde{\Delta}_k\widetilde{\Lambda}_k - M_k[(i\varepsilon_n)^2-\widetilde{\alpha}_k^2]}{\left[(i\varepsilon_n)^2-\lambda_{k+}^2\right]\left[(i\varepsilon_n)^2-\lambda_{k-}^2\right]}, \tag{78}$$

$$[\mathcal{G}(k,i\varepsilon_n)]_{14} = \frac{1}{\sqrt{\gamma_k}}\frac{-M_k\widetilde{\Delta}_k(i\varepsilon_n-\widetilde{\alpha}_k)+\widetilde{\Lambda}_k\left[\widetilde{\Lambda}_k^2+\widetilde{\Delta}_k^2-(i\varepsilon_n-\widetilde{\alpha}_k)(i\varepsilon_n+\xi_k)\right]}{\left[(i\varepsilon_n)^2-\lambda_{k+}^2\right]\left[(i\varepsilon_n)^2-\lambda_{k-}^2\right]}, \tag{79}$$

where

$$\lambda_{k\pm} = \frac{\sqrt{X_k \pm \sqrt{X_k^2-4Y_k}}}{\sqrt{2}}, \tag{80}$$

$$X_k = \widetilde{\alpha}_k^2+\xi_k^2+2\widetilde{\Delta}_k^2+2\widetilde{\Lambda}_k^2+2M_k^2, \tag{81}$$

$$Y_k = (\widetilde{\alpha}_k M_k+2\widetilde{\Delta}\widetilde{\Lambda}_k)^2+\left[\widetilde{\alpha}_k\xi_k+\left(\widetilde{\Lambda}_k^2-\widetilde{\Delta}_k^2\right)\right]^2. \tag{82}$$

For simplicity, we introduce the cutoff frequency

$$\omega_c = \max\left\{\zeta_0 t,\sqrt{\widetilde{\Lambda}_k^2+M_k^2}\right\} \tag{83}$$

to the bosonic propagators for charge fluctuations $\widetilde{b}_i^\dagger = \widetilde{e}_i^\dagger + \widetilde{d}_i$ and apply a quasistatic approximation to solve the gap equations as

$$M_k(i\varepsilon_n)\left[\left(\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}\right)^2[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{13}+\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{14}\right] \tag{84}$$

$$\simeq \frac{1}{N_s}\sum_Q\widetilde{t}_k\widetilde{t}_{k+Q}\langle\widetilde{b}_Q^\dagger(\omega_c)\widetilde{b}_Q(\omega_c)\rangle T\sum_{i\omega_m}\left[\left(\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}\right)^2[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{13}+\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{14}\right] \tag{85}$$

and

$$\widetilde{\Lambda}_k(i\varepsilon_n)\left[[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{14}+\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{13}\right] \tag{86}$$

$$\simeq \frac{1}{\sqrt{\gamma_k}}\cdot\frac{1}{N_s}\sum_Q\widetilde{t}_k\widetilde{t}_{k+Q}\langle\widetilde{b}_Q^\dagger(\omega_c)\widetilde{b}_Q(\omega_c)\rangle \times T\sum_{i\omega_m}\left[[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{14}\frac{\overline{e}_0+\overline{d}_0}{\sqrt{2}}[\mathcal{G}(k+Q,i\varepsilon_n+i\omega_m)]_{13}\right]. \tag{87}$$

Then, we obtain the superconducting states induced with electron-electron and cofermion-electron pairs formed by exchanging one charge boson.

We solve the gap equations (85) and (87) with an assumption for the symmetry of the superconducting order parameters. Here, we assume a simple $d_{x^2-y^2}$ symmetry, which has experimentally been suggested for the hole-doped cuprates, for the singlet electron-electron and cofermion-electron pairs as

$$M_k(i\varepsilon_n) = \frac{M_{d_{x^2-y^2}}(i\varepsilon_n)}{2}(\cos k_x-\cos k_y), \tag{88}$$

$$\widetilde{\Lambda}_k(i\varepsilon_n) = \frac{\widetilde{\Lambda}_{d_{x^2-y^2}}(i\varepsilon_n)}{2}(\cos k_x-\cos k_y). \tag{89}$$

As a consequence of the quasistatic approximation, the gap amplitudes $M_{d_{x^2-y^2}}(i\varepsilon_n)$ and $\widetilde{\Lambda}_{d_{x^2-y^2}}(i\varepsilon_n)$ become constant independently of the Matsubara frequency. Of course, we expect that the amplitudes of $M_{d_{x^2-y^2}}(i\varepsilon_n)$ and $\widetilde{\Lambda}_{d_{x^2-y^2}}(i\varepsilon_n)$ will damp for $\omega_c \ll |i\varepsilon_n|$.

Here, we note how the cofermion-quasiparticle pairs, namely, $[\mathcal{G}]_{14}$, enhance the quasiparticle-quasiparticle pairing. First, it is clearly seen in Eqs. (76) or (85) that the cofermion-quasiparticle pairs contribute to the superconducting order parameters of the quasiparticle-quasiparticle channel $M_k(i\varepsilon_n)$.

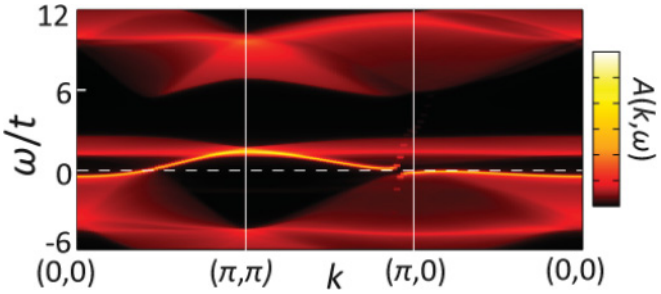

FIG. 7. (Color online) Spectral function $A(k,\omega)$ along lines running from (0,0) to (π,π), from (π,π) to $(\pi,0)$, and from $(\pi,0)$ to (0,0). Here, we use a finite broadening factor $\delta = 0.05t$.

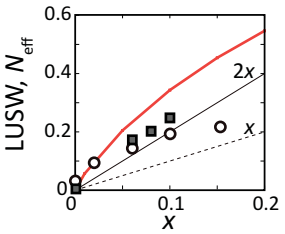

FIG. 8. (Color online) (The bold (red) solid curve shows doping dependence of low-energy spectral weight. The thin (black) solid line represents $2x$ and the thin (black) dashed line stands for x. Open circles and closed squares illustrate $N_{\rm eff}$ given in Ref. 37 for $La_{2-x}Sr_xCuO_4$ and in Ref. 68 for $Ca_{2-x}Na_xCuO_2Cl_2$, respectively.

This contribution becomes a constructive one only when the phases of the cofermion-quasiparticle and the quasiparticle-quasiparticle pairs are the same, namely, $M_k \cdot \Lambda_k > 0$. In this case, the interaction illustrated in Fig. 6(b) behaves as an attractive interaction between the cofermion-quasiparticle and the quasiparticle-quasiparticle pairs. Then, the formation of the cofermion-quasiparticle pairs enhances the quasiparticle-quasiparticle pairing through this attraction, and, indeed, enhances the pairing in our self-consistent solution of Eqs. (85) and (87).

IV. RESULTS

In this section, we show how our theory predicts physical properties of our interest. We estimate the quasiparticle and electron Green's function obtained in the above section (Sec. III C). We show the results for the Hubbard model defined in Eq. (1) on a square lattice at $U = 12t$ and $t' = 0.25t$ to get insight into the cuprate superconductors. We restrict the mean-field solutions to the homogeneous and paramagnetic ones. All the calculations are done at zero temperature. At this parameter and within the present calculation, the amplitude of the Hubbard gap δU at half-filling $n = 1$ is estimated to be $3.6t$.

First, we give the spectral functions calculated from the electron Green's function given in Eq. (69), and show the global structure of the spectra obtained from the electron Green's function. We define the retarded Green's function at the frequency ω as

$$G_\sigma(k,\omega) = \mathcal{G}_\sigma(k, i\varepsilon_n \to \omega + i\delta), \tag{90}$$

where $\delta \to +0$. Then, the spectral function is given by

$$A(k,\omega) = -\frac{1}{\pi}\mathrm{Im}\left[G_\sigma(k,\omega)\right]. \tag{91}$$

In Fig. 7, we show the result for $x = 0.05$. There are two main features. The first is the coherent band seen around the Fermi level, i.e., $\omega = 0$. The second is the incoherent spectrum induced by the charge- and spin-boson dynamics. As is already discussed in Ref. 57, the remnant of the upper Hubbard band is seen above $\omega \simeq 6t$, while the lower Hubbard band is seen between the coherent band and $\omega \simeq -8t$. The gap seen in the spectral function at $2t \lesssim \omega \lesssim 6t$ is nothing but the remnant of the Hubbard gap. Here, an additional incoherent band is seen just above the Fermi level up to $\omega \simeq 2t$. This additional incoherent band originates from the dynamics of the spin bosons. The spectral function shown in Fig. 7 is qualitatively consistent with the results of the quantum Monte Carlo simulations done by Preuss *et al.*,[42] although their simulations were done for $t'/t = 0$, $U/t = 8$, and $T \neq 0$. We note that the dispersion of the coherent band is given by poles of the quasiparticle Green's function. Below, we concentrate on the low-energy excitations within the coherent band. Therefore, we focus on the quasiparticle Green's functions instead of the electron Green's functions.

In Fig. 8, we show the LUSW defined by $(1+x)\zeta_0$ as a function of x. The quick increase of the LUSW larger than $2x$ is clearly seen in Fig. 8. We refer to the data on the effective electron number $N_{\rm eff}$ estimated from the optical conductivity measurement,[37,68] which shows nice agreement with the present LUSW at the small doping as is expected. For larger doping, the LUSW and $N_{\rm eff}$ become trivially different because $N_{\rm eff}$ is also proportional to the population of the occupied states and decreases when the number of occupied states becomes small, while the LUSW expresses only the population of unoccupied states and should always be larger than $2x$. LUSW and $N_{\rm eff}$ become identical in the small doping asymptotically. The quick increase of LUSW indicates that the Mott gap collapses and is interrupted by the quick emergence of the low-energy unoccupied states. This LUSW comes from the quasiparticle band hybridized with the cofermions. The quick increase of the LUSW is also related with a positive feedback of the doping, where the hole doping introduces additional screening of carriers leading to a further increase of unbound holon and doublon. Although the Mott transition is continuous in the low-energy limit at the Fermi level as the ground-state properties, this positive feedback gives a character close to the first-order transition in the energy scale of the LUSW, namely, in the low-energy excitation spectra. It may be related to the tendency for the phase separation universally suggested in the doped Mott insulators (including the cuprate superconductors), which has been discussed from several different viewpoints.[69]

A. Fermi-surface topology

Now we show how the reconstruction of the Fermi surface occurs in our theory. The quasiparticle Green's function

$$G_\sigma^{(f)}(k,\omega) = \mathcal{G}_\sigma^{(f)}(k, i\varepsilon_n \to \omega + i\delta)$$

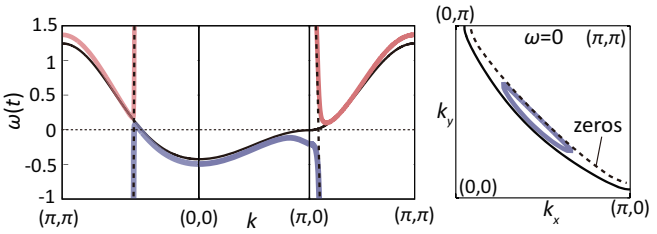

FIG. 9. (Color online) The left panel shows band dispersions and zero surface calculated for $x = 0.05$ along lines running from (π,π) to (0,0), from (0,0) to $(\pi,0)$, and from $(\pi,0)$ to (0,0). The right panel shows bare and reconstructed Fermi surface, and zero surface at $\omega = 0$. The thin solid (black) curve gives the bare band dispersion $\omega = \zeta_{0\sigma}\epsilon_k - \mu$, and thin dashed (black) curve represents the zero surface $\gamma_k\omega = \alpha_k$. The thick (blue and red) solid curves stand for reconstructed bands.

is given as

$$G_\sigma^{(f)}(k,\omega) = \left[\omega + i\delta - \zeta_{0\sigma}\epsilon_k + \mu - \frac{\Delta_k^2}{\gamma_k(\omega+i\delta)-\alpha_k}\right]^{-1} \tag{92}$$

$$= \left[\frac{1}{2} + \frac{d_k^-}{\sqrt{(d_k^-)^2 + \Delta_k^2/\gamma_k}}\right] \frac{1}{\omega + i\delta - d_k^+ - \sqrt{(d_k^-)^2 + \Delta_k^2/\gamma_k}} + \left[\frac{1}{2} - \frac{d_k^-}{\sqrt{(d_k^-)^2 + \Delta_k^2/\gamma_k}}\right] \frac{1}{\omega + i\delta - d_k^+ + \sqrt{(d_k^-)^2 + \Delta_k^2/\gamma_k}}, \tag{93}$$

where $d_k^\pm = \frac{1}{2}(\zeta_0\epsilon_k - \mu \pm \alpha_k/\gamma_k)$.

The Green's function given in Eq. (92) shows the divergence of the quasiparticle self-energy given by $\Delta_k^2/(\gamma_k\omega - \alpha_k)$ at $\gamma_k\omega - \alpha_k = 0$. In other words, the zero surface defined by $\gamma_k\omega = \alpha_k$ emerges. Then, the zero surface splits the band dispersion defined by $\omega = \zeta_{0\sigma}\epsilon_k - \mu$ into two bands as $\omega = d_k^+ \pm \sqrt{(d_k^-)^2 + \Delta_k^2/\gamma_k}$, as is depicted in Fig. 9. For small doping such as $x = 0.05$, our theory predicts that the reconstructed Fermi surface becomes a small pocket, as is seen in the right panel of Fig. 9.

To show changes in the Fermi-surface topology with increasing doping, the single-particle spectral function at $\omega = 0$ is given for the hole-doping rate $x = 0.05, 0.15, 0.20$ in Figs. 10(a)–10(c), with $\delta = 0.05t$. The topological transitions occur at $x \simeq 0.13$ and 0.18, as is depicted in Fig. 10(e). In the region $0.13 \lesssim x$, only small Fermi pockets exist, where the zero surface smears out the outer part of the pocket yielding an arc structure as we see in Fig. 10(a) (note that the damping is large near the zero surface because of an enhanced self-energy), consistently with the experimental signature in Fig. 10(d) as we discuss below. For $0.13 \gtrsim x$, a completely different topology with large Fermi surfaces appears instead of Fermi pockets. For $0.13 \lesssim x \lesssim 0.18$, a holelike surface centered at (π,π) [depicted by the solid curve in Fig. 10(b)] and an electronlike one centered at (π,π) [depicted by the dashed curve in Fig. 10(b)] coexist. These two surfaces together figure out an enclosed hole (electron unoccupied) strip between the solid and dashed curves. However, the electronlike surface (the dashesd curve) is hardly seen because the zero surface exists near this electronlike surface. On the other hand, for $0.18 \gtrsim x$, there exists an electronlike surface centered at (0,0) [depicted by the solid curve in Fig. 10(c) and an electronlike one centered at (π,π) [depicted by the dashed curve in Fig. 10(c)]. The electronlike surface centered at (π,π) becomes less visible for the used broadening factor $\delta = 0.05t$ than the electronlike surface centered at (π,π) for $x \lesssim 0.18$.

The quantum transition at $x \simeq 0.18$ is a trivial one expected from the single-particle picture. The topology of the Fermi surface at this quantum transition point changes from a holelike surface centered at (π,π) depicted for $x = 0.15$ in Fig. 10 by the solid curve to an electronlike surface centered at (0,0) depicted for $x = 0.20$ in Fig. 10 by the solid curve. This topology change is essentially understood from the noninteracting picture, where the band dispersion $\epsilon_k = -2t(\cos k_x + \cos k_y) + 4t' \cos k_x \cos k_y$ with $t'/t = 0.25$ shows the transition from the electronlike to the holelike Fermi surfaces with the increasing electron concentrations. When the Fermi level is shifted by doping and touches the saddle points at $(\pm\pi,0)$ and $(0, \pm\pi)$, such topological changes occur.

However, our solution shows another nontrivial topological transition at $x \simeq 0.13$. The topology in the phase $x < 0.13$ is highly nontrivial, and is not adiabatically connected with the conventional Fermi liquid. Such a topological change never occurs without the zeros of the Green's function.

For comparison with experiments, we refer to the ARPES spectrum of $La_{2-x}Sr_xCuO_4$ (Ref. 10) in Fig. 10(d). Arclike Fermi surfaces observed for $x = 0.03$ and 0.07 show overall consistency with our result for $x = 0.05$. On the other hand, so-called nodal metallic behaviors, i.e., large amplitude of the spectral weight allowed only around the nodal direction [line running from (0,0) to (π,π)] are not seen in our results. Instead, in our theory, the weight is rather larger near the endpoint of the arc than the nodal point as we see in Fig. 10(a). This subtle discrepancy is likely to originate from additional self-energy effects, which are not included in the present theory. For example, fluctuations of d-wave superconductivity

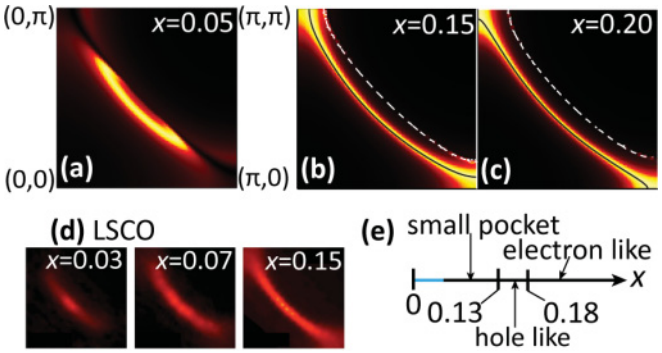

FIG. 10. (Color online) (a)–(c) Single-particle spectral function at $\omega = 0$. Solid and dashed lines illustrate the poles of quasiparticles. (d) Fermi surface observed by ARPES (Ref. 10). (e) Phase diagram of Fermi-surface topology in our theory. Topological phase boundaries exist at the doping $x = 0.13$ and 0.18.

and/or short-range antiferromagnetic fluctuations are possible candidates. However, we note that these fluctuations with finite correlation length by themselves never induce changes in the Fermi-surface topology.

When the large Fermi surface appears for $x \gtrsim 0.13$, the Fermi surface always extends across the Brillouin zone boundary or the lines, which connect the Γ point, (0,0), and the antinodal points $(\pm\pi,0)$ or $(0,\pm\pi)$. However, the single-particle excitation from the Fermi level now has a gap due to the hybridization gap Δ_k in Eq. (92) in the quantum phase characterized by the nontrivial Fermi-surface topology for $x \lesssim 0.13$. In Fig. 11, the single-particle spectral function is depicted along the symmetry lines in the Brillouin zone. A gap measured from the Fermi level emerges, which corresponds to the pseudogap in the ARPES measurements. Hereafter, we define the amplitude of the pseudogap in the single-particle spectrum Δ as the gap amplitude between the Fermi level μ and the maximum of the single-particle dispersion below μ along the line running from (0,0) to $(\pi,0)$ and the line running from $(\pi,0)$ to (π,π). We also reproduce the ARPES dispersion observed in $La_{2-x}Sr_xCuO_4$ for $x = 0.10$ (Ref. 70) for comparison in Fig. 11 (indicated by bold dashed curve). Here, we employ a widely accepted parameter $t =$ 400 meV. Indeed, the dispersion has an excellent similarity. The quasiparticle dispersion does not touch the Fermi level along the Brillouin zone boundary, here the line connecting $(\pi,0)$ and (π,π), in both our result and the ARPES measurement. In other words, the pseudogap amplitude Δ is finite for both cases. However, the quasiparticle state at $(\pi,0)$ in the ARPES data has lower energy than our result and the discrepancy is up to 20 meV. We expect some additional factors to push the dispersion down around the antinodal points $(\pm\pi,0)$ and $(0,\pm\pi)$. As is discussed above, fluctuations of d-wave superconductivity and/or short-range antiferromagnetic fluctuations are possible candidates of the origin of this discrepancy.

B. Pseudogap

The pseudogap in single-particle excitations defined in the above section characterizes a quantum phase with the nontrivial Fermi surface because a nonzero pseudogap induces the breakdown of the trivial large Fermi surface. In our theory, the pseudogap is induced by the hybridization gap between the cofermions and quasiparticles. Therefore, the amplitude of the pseudogap is roughly determined by

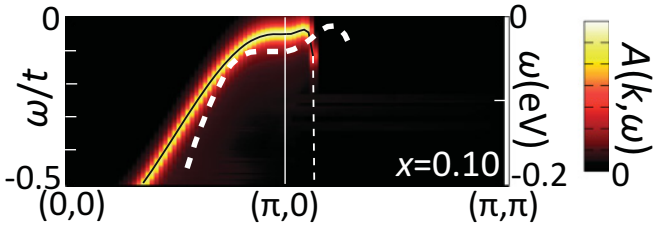

FIG. 11. (Color online) Single-particle spectral function near antinodal point for $x = 0.10$. The thin solid and dashed lines illustrate poles of quasiparticles. The white bold dashed line represents the ARPES spectrum observed in $La_{2-x}Sr_xCuO_4$ for $x = 0.10$ (Ref. 70).

the hybridization gap $\Delta_k/\sqrt{\gamma_k}$. The explicit evaluation of the amplitude of Δ_k is done by using Eq. (B3) in Appendix B, in the lowest order of t_{ij}. The amplitude of Δ_k is roughly estimated as t^2/W multiplied by a numerical factor, where W is the bandwidth of original electrons. Here we note that the denominators in Eq. (B3) include the dispersion of the charge and spin bosons. The bandwidth of the charge and spin bosons is roughly equal to the bandwidth of original electrons $W \simeq 8t$, and the bandwidth of the spin wave $J \sim 4t^2/U$, respectively. For $U = 12t$, the bandwidth of the charge bosons dominates that of the spin bosons, namely, $W > J$. Therefore, the denominators appearing in Eq. (B3) are roughly scaled by W. On the other hand, the amplitude of γ_k is roughly estimated as t^2/W^2 multiplied by a numerical factor through Eq. (A17). Therefore, the amplitude of the hybridization gap $\Delta_k/\sqrt{\gamma_k}$ is roughly estimated as $(t^2/W)/\sqrt{t^2/W^2} = t$ multiplied by a numerical factor. The present result for the energy scale of the pseudogap is consistent with previous studies based on the cluster perturbation theory[22] and, later, cellular dynamical mean-field theory (DMFT).[71]

We discussed the pseudogap Δ along the definition often used in ARPES measurements in Sec. IV A. Here, we discuss the behaviors of the hybridization gap between two bands given in Eq. (93) and illustrate them in Fig. 9. First, the hybridization gap Δ_k/γ_k in our theory does not have nodes and is s-wave-like in the momentum space, although, as is illustrated in Fig. 12, the gap seems to be smaller in the nodal direction [the line running from (0,0) to (π,π)] than around the antinodal points $(\pm\pi,0)$ and $(0,\pm\pi)$.

Below the transition point $x \simeq 0.13$, the density of states (DOS) of the original electrons $\hat{c}^\dagger_{k\sigma}$ at the Fermi level ρ_F, defined as

$$\rho_F = \zeta_0 \frac{1}{N_s} \sum_{k\sigma} \left[-\frac{1}{\pi} \mathrm{Im} G_{f\sigma}(k,\omega=0) \right], \qquad (94)$$

is clearly suppressed, as is illustrated in Fig. 13. We compare ρ_F with the specific heat measured for $La_{2-x}Sr_xCuO_4$.[72,73] The specific-heat coefficient γ is expected to be proportional to the DOS. In the Sommerfeld's free electron model, the specific-heat coefficient γ at $T = 0$ is given by $\gamma = \pi^2\rho_F/3$. Here, we use this relation and $t = 400$ meV to calculate the expected values for γ with our theory.

We note that γ in Fig. 13 indicated by closed (blue) circles is obtained through a linear extrapolation from low-temperature data in Ref. 72, and shows the possible lower limit for γ. From the data in Ref. 72, γ in the underdoped region shows three characteristic temperature ranges. For example, the data for $x = 0.08$ (see the inset of Fig. 13) show (i) for 100 K $\lesssim T \lesssim$ 200 K, a linear-temperature dependence, (ii) for 50 K $\lesssim T \lesssim$ 100 K, another linear-temperature dependence but with different slope from (i), and (iii) for $T \lesssim$ 50 K, the BCS-type jump. We linearly extrapolate γ for 50 K $\lesssim T \lesssim$ 100 K and obtain γ at $T = 0$. On the other hand, data for γ indicated by closed (red) squares[73] in Fig. 13 may set an upper limit for γ because to exclude the influences from superconductivities, the authors of Ref. 73 used extrapolation from Zn-doped samples, which tend to show larger γ than samples without Zn doping.

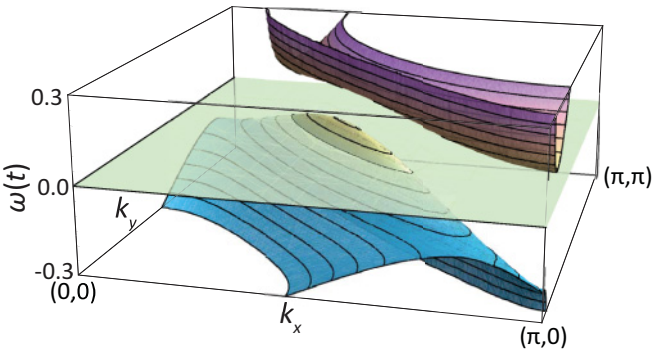

FIG. 12. (Color online) Band dispersion of quasiparticles for $x = 0.05$. A full gap structure is clear.

Our result for γ is consistent with the experiments, although we do not choose parameters specific to $La_{2-x}Sr_xCuO_4$. On the other hand, the extrapolated γ [closed (blue) circles in Fig. 13] indicates further reduction of the density of states at the Fermi level. It suggests possible roles of fluctuations from the d-wave superconductivity and/or antiferromagnetism in the lower energy scale, which is left for future studies beyond the scope of this paper.

C. Superconductivities

In studies on high-T_c superconducting cuprates, the mechanism of the superconductivity is of course the central issue.[74] However, the consensus on the mechanism has not been reached after more than 20 years of the discovery. Here, we show how the novel cofermion mechanism for superconductivities developed in Sec. III D works and reproduces experimental observations on the high-T_c superconducting cuprates.

We examine the doping dependence of the single-particle excitation gap in the superconducting state based on the formalism developed in Sec. III D. The result is shown in Fig. 14 (bold dashed line). We adopt, as a definition of the amplitude

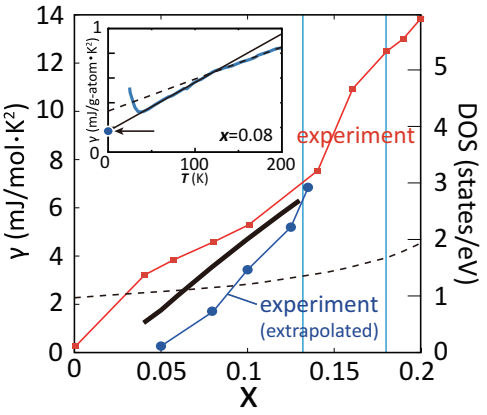

FIG. 13. (Color online) Density of states (DOS) of electrons given as a function of hole-doping rate x. The thin dashed curve represents the DOS of the noninteracting case. The closed (blue) circles show γ extrapolated from experimental data of $La_{2-x}Sr_xCuO_4$ in Ref. 72. The closed (red) squares represent experimental data of $La_{2-x}Sr_xCuO_4$ in Ref. 73. The inset shows an example for the extrapolation used here [see the text below Eq. (94) for details].

of the gap in the superconducting state, the minimum of the single-particle excitation gap along the Brillouin zone boundary and the line connecting the Γ point (0,0) and the antinodal points $(\pm\pi,0)$ and $(0,\pm\pi)$, which is the same definition for the pseudogap in the normal state. This definition of the amplitude of the gap is consistent with the data analysis of the ARPES measurements. As is shown in Fig. 14 (bold dashed line), the single-particle excitation gap is still finite around $x = 0.3$ for the d-wave superconducting state in the present treatment, which is quantitatively inconsistent with the experimental fact that the superconductivity disappears for $x \gtrsim 0.26$ in the high-T_c cuprates. To highlight a possible origin of this inconsistency, we also show a result for the single-particle excitation gap in Fig. 14 (thick solid line) calculated by rescaling the quasistatic boson propagator as

$$\langle \tilde{b}_Q^\dagger(\omega_c)\tilde{b}_Q(\omega_c)\rangle \to (1-\zeta_0)\,\langle \tilde{b}_Q^\dagger(\omega_c)\tilde{b}_Q(\omega_c)\rangle.$$

By the rescaling, the single-particle excitation gap is clearly suppressed around $x = 0.3$. The reason why we consider such a rescaling is the following: In our treatment, the charge fluctuations, especially holon fluctuations, remain finite even in the dilute limit $x \to 1$, where holons fill all the sites and should be localized as hard-core bosons. Holons are completely localized and form the "Mott insulating" state for $x = 1$. The fluidity of the holons for $x \lesssim 1$ originates from finite electron density,$1 - x \neq 0$ because the finite electron density introduces "vacancies" in the holon's Mott insulating state, and causes the fluidity of the holons. On the other hand, doublons vanish at the limit $x \to 1$. Near this limit, the density of the doublons is scaled by $(1-x)^2$. Therefore, for large doping x, the weight of fluctuating charge bosons should vanish as a function of $1-x$. We note that, for $x \simeq 1$, $1-\zeta_0 \simeq 1-x$. On the other hand, for $x \simeq 0$, the charge fluctuating bosons have the weight $1-\zeta_0$ because the

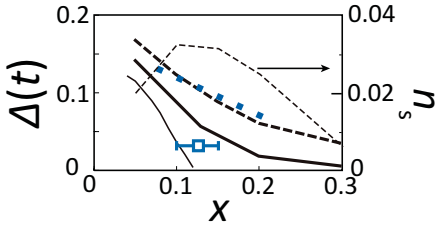

FIG. 14. (Color online) Doping dependence of excitation gaps. The thin (black) solid line represents the pseudogap amplitude in the normal state. This amplitude shows the contribution that has nothing to do with the superconductivity but rather has to do with the precursor of the Mott gap. The bold (black) dashed line represents the minimum of the single-particle excitation gap in the superconducting state along the Brillouin zone boundary or the diagonal line connecting the Γ point and the antinodal points. The bold (black) solid line shows a refined estimate of the single-particle excitation gap in the superconducting state obtained from the pair potential mediated by exchanging one charge boson, where the potential is multiplied by a scale factor $(1-\zeta_0)$. The x dependence of excitation gaps observed by ARPES for $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Ref. 75) is indicated by the thick (blue) dotted line. The open square represents the superconducting gap of optimally doped $La_{2-x}Sr_xCuO_4$ estimated by ARPES measurements (Ref. 76). The thin (black) dashed curve represents the density of superconducting electrons n_s.

incoherent band induced by fluctuating bosons shares the total spectral weight with the coherent band, which has the weight ζ_0. Therefore, for all the doping range, the rescaling of the boson propagators is naturally interpolated by $1-\zeta_0$.

Then we compare our result with the experimentally observed gap Δ. The result of Δ without rescaling of the propagators given by the thick dashed curve in Fig. 14 shows a rough qualitative agreement with the ARPES result obtained from the single-particle excitation gap at low temperatures ($T = 14$ K) for $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Ref. 75) [dotted (blue) line in Fig. 14]. However, our result without the rescaling overestimates Δ around $x = 0.3$, where the superconducting phase is not observed in any of the cuprates. On the other hand, the single-particle excitation gap calculated by using the rescaled boson propagator shows a suppression of Δ around $x = 0.3$. The single-particle excitation gap Δ for $x \sim 0.1$ is consistent with the superconducting gap of optimally doped $La_{2-x}Sr_xCuO_4$ estimated by ARPES measurements[76] (open square in Fig. 14). On the other hand, the rescaled boson propagator appears to be still overestimated for the overdoped region $x \sim 0.3$. The origin of such overestimated fluctuating charge bosons in our theory may be ascribed to the approximation on the bosonic propagators given in the Dyson equations shown in Fig. 5: Lifetime of the coherent propagation of the fluctuating charge bosons is not taken into account in our theory. There exist two possible origins of the finite lifetime. The first is damping caused by single-particle excitations, namely, the Landau damping. When the doping increases, the motion of the coherent carriers may disturb the propagation of the fluctuating charge bosons seriously. The second is the separation of charge and spin fluctuations, which becomes a good approximation for $x \simeq 0$. Such separation of charge and spin fluctuations may become worse for large x because the motion of the coherent carriers mixes the charge and spin modes. The mixing of charge and spin fluctuations also disturbs the propagation of the fluctuating charge bosons and gives a finite lifetime. Therefore, these factors neglected in our theory will suppress the charge fluctuations and, consequently, the superconductivities.

The single-particle gap Δ decreases monotonically as the doping x increases. It is consistent with ARPES measurements. On the other hand, the critical temperature of the superconductivity T_c shows a peak around the $x \sim 0.15$ in $La_{2-x}Sr_xCuO_4$ and exhibits a domelike structure as a function of x. To estimate the tendency of T_c from our results for $T = 0$, we show the x dependences of the density of superconducting electrons n_s in Fig. 14 (thin dashed curve) defined as

$$n_s = \left| \zeta_0 \frac{T}{N_s} \sum_{k,i\varepsilon_n} (\cos k_x - \cos k_y) \left[\mathcal{G}(k, i\varepsilon_n) \right]_{13} \right|. \tag{95}$$

The pairing mechanism proposed here includes the contribution of the cofermion-quasiparticle pairs in addition to the quasiparticle-quasiparticle pairs. This contribution of the cofermion-quasiparticle pairs itself is an unexplored perspective introduced in this paper. Furthermore, it offers an insight into one of the most interesting issues in the physics of the cuprate superconductors, namely, relationship between the pseudogap formation and the high-T_c superconductivity. The cofermions induce the hybridization gap around the Fermi level and, as a result, the pseudogap. The pseudogap formation itself reduces the DOS around the Fermi level and is harmful for the high-T_c superconductivity. However, the cofermions support the high-T_c superconductivity through the cofermion-quasiparticle pairing simultaneously. Therefore, the pseudogap structure does not necessarily destroy the superconducting pairing. The pseudogap formation is the other side of the coin of the cofermion pairing contributing to the superconductivity. This dual character offers an insight into the recent controversy called dichotomy observed in ARPES measurements.[38,39] There exist two major pictures for the relationship between the pseudogap and high-T_c superconductivity: One tells us that the pseudogap is a precursor of the Mott gap[62] and coexists with the superconducting gap; the other claims that preformed pairs of strong-coupling superconductivities induce the pseudogap.[77] The present theory offers an alternate route of understanding that reconciles the dichotomy.

By exchanging one charge boson, attractive interactions are induced for forward scattering channels. Even though there is the strong on-site repulsion U, such attractive interactions may favor anisotropic superconductivities, including not only the $d_{x^2-y^2}$-wave superconductivity but also, for example, d_{xy}-wave and extended s-wave superconductivities. In the present theory, however, a simple extended s-wave superconducting parameter proportional to $\cos k_x + \cos k_y$ does not develop.

D. Metal-insulator transitions

We propose a scenario for the filling control metal-Mott-insulator transitions based on our theory. Before going into our scenario, we note that there is a difficulty in our theory in the small doping limit $x \to +0$. In this limit, we have Fermi pockets with small but finite volume depending on parameters such as t'/t and U/t. It is in contrast to the zero-doping limit of the ordered insulators. In our theory, the weights of the two bands split by the zero surface illustrated in Fig. 9 are not the same. From Eq. (93), these weights are given as

$$\frac{1}{N_s} \sum_k \left[\frac{1}{2} \pm \frac{d_k^-}{\sqrt{(d_k^-)^2 + \Delta_k^2/\gamma_k}} \right]. \tag{96}$$

In the case of the antiferromagnetic ordered state, the counterparts of these weights are given by

$$\frac{1}{N_s} \sum_k \left[\frac{1}{2} \pm \frac{\epsilon_{1k}}{2\sqrt{\epsilon_{1k}^2 + \Delta_{\mathrm{AFM}}^2}} \right] \tag{97}$$

from Eq. (49) and are the same because of

$$\frac{1}{N_s} \sum_k \frac{\epsilon_{1k}}{2\sqrt{\epsilon_{1k}^2 + \Delta_{\mathrm{AFM}}^2}} = 0. \tag{98}$$

The weight compensation between these two bands split by the zero surface is needed to reproduce ordered insulating phases. On the other hand, our theory does not offer this weight compensation. Therefore, our theory does not necessarily predict that the small hole pockets seen in Fig. 10(a) shrink to points at the small doping limit $x \to +0$. In addition, in the limit $x \to +0$, the imaginary part of the quasiparticle and cofermion self-energy, which is not taken into account

in the present theory, may affect the weight compensation. Higher-order corrections to the cofermion dynamics will also affect the structure of the zero surface in this limit.

Despite these limitations, our result still suggests a possible scenario for the filling control metal-Mott-insulator transitions. If the small Fermi pockets seen in Fig. 10(a) shrink to the Fermi points, the filling control metal-insulator transition occurs as a topological transition. On the other hand, if the small Fermi pockets do not shrink to the Fermi points, the quasiparticle weight of the Fermi pockets becomes zero, in spite of the finite volume of pockets. Then, the insulating state appears because of the fading out of the coherent band. In both of the two cases, a small but finite DOS is expected in the limit $x \to +0$. It should be noted that, anyway, a nontrivial topological transition is expected to occur at a finite doping before the system becomes insulating.

Here, we note that the order of the three quantum phase transitions, namely, the Mott transition at $x = 0$ and the two topological transitions at $x \sim 0.13$ and 0.18 are all continuous within the present calculated results, although in general, the first-order transition is not *a priori* excluded. If a topological quantum phase transition takes place as a continuous one at $T = 0$, it becomes just a crossover at nonzero temperature because the Fermi surface blurs at $T > 0$ and the topological character of the Fermi surface loses well-defined meaning at $T > 0$ anymore. It has been stressed in the literature[31–34] that the sharp change between the overdoped and underdoped regions is associated with some kind of quantum criticality triggered by a quantum phase transition of the symmetry-breaking order. In this conventional quantum criticality, the phase transition usually extends to nonzero temperature as the border of the symmetry-broken phase. However, the present quantum criticality is completely different from this category, where not the symmetry breaking but the topology change drives the transition.[78–81]

V. SUMMARY AND DISCUSSIONS

In this paper, we have proposed a microscopic mechanism for the Fermi-surface reconstruction and non-Fermi-liquid behaviors emerging in proximity to the Mott insulators, motivated by puzzles experimentally observed in the underdoped cuprates. We construct our theory starting from one of the simplest theories for correlated metals, namely, the slave-boson mean-field theory for the Hubbard model by Kotliar and Ruckenstein (KR). Then, a special emphasis is placed on the role of extra charge dynamics on low-energy spectra. We find that hidden cofermionic particles constructed from composite fermions play a crucial role in addition to the quasiparticles described by the previous KR theory. The additional cofermions called *holo-electrons* and *doublo-holes* represent substantial low-energy parts of the charge dynamics.

Thus, these introduced cofermions hybridize with the quasiparticles and cause a hybridization gap represented by zeros of the quasiparticle Green's functions. As a result, although the explicit symmetry breaking is absent, the gap identified as the pseudogap of the cuprates naturally comes out. Although the origin of the gap has an apparent similarity to the case of the symmetry-broken phases such as commensurate antiferromagnetic metals, in the sense that both of them can be ascribed to a hybridization gap, the mechanism of the hybridization and the partner of the hybridization are very different from the simple symmetry breaking. The origin of the gap is also different from the Mott gap itself. The large Mott gap in the insulating phase is quickly replaced with the present, much smaller, hybridization gap upon doping because the LUSW above the hybridization gap emerges. In the result of the cellular dynamical mean-field theory,[25] this hybridization gap coexists with the clear remnant of Mott gap structure in a much larger energy scale. This indicates that the two gaps are separated phenomena. Thus, the pseudogap or the hybridization gap in the underdoped region is not really the precursor of the Mott gap.

When the cofermion strongly hybridizes with the quasiparticles and the resultant hybridization gap splits the quasiparticle band around the Fermi level, Fermi-surface reconstruction occurs in a natural way. The hybridization gap indeed leads to a break-up of the Fermi surface into small hole pockets around the nodal direction, and a phase topologically different from the prediction of the weak-coupling picture emerges. This phase realized in the underdoped region is separated by a topological phase transition from the normal Fermi-liquid phase in the overdoped region. Our result clearly shows that an unconventional non-Fermi-liquid phase emerges in the underdoped region. Such topological changes and an emergence of a new phase are consistent with experimentally observed "Fermi-arc" formation in hole-underdoped cuprates.

The topological quantum phase transition between the conventional Fermi liquid and the non-Fermi liquid is a continuous one at $T = 0$ within the present approximation. This implies that it transforms to a crossover at nonzero temperature because the continuous topological transition is well defined only at zero temperature. In principle, it does not exclude a possibility of the first-order transition of the topological change, which extends to nonzero temperatures with the critical end point at a nonzero temperature.[80] The present results do not support the existence of such a first-order transition in agreement with the absence of the clear indication of the first-order transition in the experiments.

Our formulation offers a concept for the "pseudogap" phenomena observed in hole-underdoped cuprates. It proposes that the pseudogap in the single-particle spectrum results from the hybridization gap between the cofermions and the quasiparticles. The part below the Fermi level has the maximum in the antinodal region and has a structure similar to the $d_{x^2-y^2}$ symmetry, in agreement with the experimental observation. The main part of the gap lies, however, above the Fermi level. The total hybridization gap has the structure of the s-wave gap.

Consequently, the characteristic energy scale of the pseudogap is identified as that of the hybridization gap amplitude and, thus, basically scaled by the bandwidth (kinetic energy) of the bare electrons t multiplied by a numerical factor. Therefore, the energy scale of the pseudogap has nothing to do either with the Hubbard gap, the on-site Coulomb repulsion U, or with the superexchange interaction $J \sim t^2/U$.

We have also numerically studied the relevance of our theory in the realistic condition. For the on-site Coulomb repulsion U large enough to create the Hubbard gap $U \gtrsim 10t$ and small amount of hole doping $x \lesssim 0.1$, regarded as a

relevant parameter for the high-T_c cuprates, the structure of the momentum dependence and the amplitude of the pseudogap below the Fermi level has a quantitative consistency between the ARPES measurements for the hole-doped cuprates and the present results. Doping dependence of the pseudogap amplitude is also consistent with the experimental observation.

Reduction of the electronic density of states induced by the pseudogap formation gives another consistency with the experiments as observed as specific-heat coefficients. The pseudogap formation in our theory also induces asymmetry of the DOS around the Fermi level, which naturally explains the asymmetric STM spectra.[82,83] The overall consistency supports that the cofermions, holo-electrons, and doublo-holes are indeed a relevant object to be considered in the physics of the cuprate superconductors.

The present theory has further consequences and predictions for experiments. The hybridization gap basically has an s-wave-like symmetry and a major part of the gap lies above the Fermi level for the model of the hole-doped cuprates. Although the present resolution limit of the inverse photoemission does not allow determination of the detailed structure of the unoccupied electronic states, we propose that our prediction of this s-wave-like gap structure accompanied by the main part of the LUSW lying above the hybridization gap can be tested experimentally if some high-resolution measurements of the unoccupied spectra are provided. The optical conductivity $\sigma(\omega)$, especially mid-infrared peak and a long tail of $\sigma(\omega)$ observed in the underdoped cuprates, indeed supports the existence of such LUSW.[37,68]

So far, direct measurement of the cofermions appears to be difficult because the dispersion of the cofermions corresponds to the zeros of the quasiparticles, while the zeros are in general hard to detect. In addition, the cofermion does not have a charge and does not allow an electromagnetic detection. However, it contributes to the entropy and thermal transport. Since the electric conduction is contributed only from the quasiparticle, while the thermal transport can arise from the cofermion as well, we expect a serious breakdown of the Wiedeman-Franz (WF) law. The WF law predicts that the ratio of the thermal conductivity κ to the electric conductivity σ is equal to L_0T, namely, $L_0 = \kappa/T\sigma$, where L_0 is a universal constant given by $L_0 = (\pi^2/3)\cdot(k_B/e)^2$, namely, the Lorenz number. On the other hand, in our theory, the ratio $L = \kappa/T\sigma$ is predicted to be larger than L_0 in proximity to Mott insulators because the cofermions carry additional energy and contribute to κ. Indeed, the breakdown of the Wiedeman-Franz law reported recently in hole-underdoped cuprates[84] supports the present prediction.

We propose a scenario for the filling control metal-Mott-insulator transitions in two dimensions based on the introduced cofermions. We predict a pseudogap formation and consequently a nontrivial change in the Fermi-surface topology in the underdoped region. Then, the criticality of the small Fermi pockets determines the nature of the Mott transition if the transition is continuous. Two scenarios remain possible: On the verge of the Mott transition, it may either shrink to points before vanishing by keeping the quasiparticle weight nonzero, or the quasiparticle weight decreases to vanish by keeping a finite volume of the Fermi surface. In the former case, the filling control metal-Mott-insulator transitions occur as a topological one. The density of states remains nonzero and finite on the verge of the transition because of the two dimensionality. In the latter case, the effective mass diverges. These two possibilities are essentially identical to the two types discussed in the literature.[85] Although the present numerical accuracy is not sufficient for determining the ultimate criticality, the shrinkage of the pocket and arc overall supports the former scenario, while the reduction of the DOS suggests only a part of the pocket contributes to low-energy excitations as the arc. This agrees with the experimental results. More detailed and accurate determination of the criticality is left for future studies.

We also propose a mechanism for high-temperature superconductivity driven by the cofermion-quasiparticle pairing. It offers a new insight into the relationship between the pseudogap formation and the high-T_c superconductivity. The cofermions introduced in this paper induce the hybridization gap around the Fermi level and, as a result, the pseudogap. This pseudogap formation itself reduces the DOS around the Fermi level and destroys superconductivity. On the other hand, the cofermions enhance the high-T_c superconductivity through the cofermion-quasiparticle pairing, simultaneously. At the present level of approximation, in the overdoped region, a process exchanging one charge boson overestimates the charge fluctuations and predicts much larger single-particle gap in superconducting phases than in experimentally observed gaps. It is left for future studies to correct such an overestimate of charge fluctuations. In addition to one-charge-boson exchange processes, cofermion polarization will help the pairing between quasiparticles, which is not taken into account as a higher-order contribution in this paper. However, the superconducting gap obtained from the present approximation already reproduces the domelike structure as a function of the doping concentration with a right order of magnitude if we compare with the cuprates.

Although we focus on the hole-doped cuprates in this paper, electron-underdoped cuprates are also interesting from the viewpoint of the present cofermion theory. In the electron-underdoped systems, our theory will predict the emergence of two electron pockets centered at $(\pi,0)$ and $(0,\pi)$ even in the absence of antiferromagnetic long-range orders. Such a Fermi-surface topology may reduce instabilities toward antiferromagnetic orders, and will make $d_{x^2-y^2}$-wave superconductivities stable: The electron pocket formation prevents the nesting of the Fermi surface, and allows an emergence of full-gapped superconductors, but with the $d_{x^2-y^2}$-wave symmetry, because the node lines run the momenta where the original electron pockets are absent.

Here, we make some remarks in regard to the relationship with other theoretical approaches. In contrast to our approach, the high-energy charge degrees of freedom to do with the UHB have often been integrated out to extract the low-energy physics from the Hubbard model. The t-J model is a typical effective model derived from such a treatment and has been studied as a canonical model describing the doped Mott insulator. The gauge theory based on the t-J model has been intensively studied to explain the low-energy physics of cuprate superconductors as doped spin liquids.[66] Wen and Lee[86] proposed that a hole-doped spin-liquid state may exhibit hole pockets. According to this theory, the optical or direct gap

of this phase is d-wave-like: there is no optical gap along the nodal direction. Although the gapless excitation is present in the nodal direction by the electron-hole excitation through the hole pocket even in our theory, the gauge-theory scenario by Wen and Lee is, in contrast to our result, predicting that a relatively small but nonzero amplitude of the optical gap even along the nodal direction is superimposed, namely, a s-wave-like optical gap should be visible between the two bands separated by the hybridization gap. The s-wave-like pseudogap is also supported in the previous numerical studies by Stanescu and Kotliar[23] and Sakai *et al.*[25] in support of the present theory. An earlier exact diagonalization study without any bias has indicated that even the doped t-J model has a similar s-wave-like pseudogap structure in agreement with our theory.[87] As we mentioned already, it is desired to test against the two contradicting predictions of our theory and the gauge theory by inverse photoemission spectroscopies or by more sophisticated and high-resolution experimental tools to measure the unoccupied states in the future.

Another crucial difference between the present theory and the gauge theory can be tested by the breakdown of the Wiedeman-Franz law. In the gauge theory, the breakdown is due to the contribution from the spinons.[88] However, if the arc structure is observed, the d-wave-like gap needs to be developed by the flux fluctuations, which lead to the confinement of a holon and a spinon generating a quasiparticle in the nodal direction. It should recover the Wiedeman-Franz law in this region of the arc formation. Therefore, at low temperatures in the underdoped region, the Wiedeman-Franz law should eventually be followed. On the contrary, in the present case of the cofermions, the contribution continues even at low temperatures and the breakdown of the Wiedeman-Franz law is robust. Recent experimental results appear to support our prediction.[84,89]

Integrating out the "high-energy"-charge degrees of freedom represented by the existence of a doublon thoroughly leads to the ignorance and overlook of important aspects of the low-energy spectrum of the Hubbard model. There exist a variety of experimental facts that are not accounted for by the t-J physics; an example is the estimate of the LUSW by using $N_{\rm eff}(\omega)$ observed in the optical conductivity measurement.[37] Choy *et al.* have made a careful treatment to extract the low-energy effective action, with the failure of the t-J model kept in mind.[90] They claim the weight transfer among the coherent band, the LHB, and the UHB described by hidden $2e$ bosons. In spite of the illuminating proposal, their effective action could not be solved exactly. Approximate evaluations of their theory show a soft gap behavior and predict a semiconducting behavior in the "pseudogap phase," where resistivity ρ is expected to diverge as $\rho \propto T^{-1}$ with lowering temperatures. In the present theory, the Fermi pocket appears in the pseudogap phase in the underdoped region, say, for $x \lesssim 0.13$. Therefore, we predict metallic conduction in the pseudogap phase, in sharp contrast to the hidden $2e$-boson theory.

As the authors have already discussed in Ref. 27, the formation of the cofermions shares profound similarity with that of the excitions in semiconductors. In the so-called d-p model for the cuprates, which contains both d electrons on the copper sites and p electrons on the oxygen sites, it has been claimed that the excitionic effects due to d-p interactions strongly affects its excitation spectrum.[91] The relationship between excitonlike features of the cofermions and excitionic effects studied in the d-p model are also desired to be clarified in the future.

Our theory has been constructed to account for the charge dynamics more seriously than the literature and proposed the topological changes of the Fermi surface in proximity to the Mott insulators. The reconstructed Fermi surface also offers an unexplored avenue at smaller energy scale if it is combined with other possible fluctuations such as spin and superconducting fluctuations. Clarifying possible emergence of antiferromagnetic orders in the small doping region $x \lesssim 0.02$ by using a unified scheme is left for future studies. It is also left for future studies as to how the topological changes affect various possible symmetry breakings such as time-reversal symmetry breakings, stripe formation, and incommensurate charge orders including phase separations. Experimentally suggested nodal metallic behaviors should also be examined in more detail. Our theory will give a new insight into emergence of these competing orders, the high-$T_{\rm c}$ superconductivity, and the anomalous metallic state.

ACKNOWLEDGMENTS

The authors thank Yukitoshi Motome and Shiro Sakai for useful discussions. Y.Y. is supported by the Japan Society for the Promotion of Science.

APPENDIX A: SELF-ENERGY FOR THE COFERMIONS

Here, we calculate the self-energy for the cofermions in Dyson equations (see Fig. 5).

The self-energy matrix in Eq. (63) is calculated by

$$\Sigma^{cd}(k) = \frac{T^3}{N_{\rm s}^3} \sum_{P,Q,R} \sum_{a,b=1,2} (g_{1\sigma} g_{2\sigma})^4 \langle \beta^a_Q \beta^{b\dagger}_Q \rangle \Big[t_{\boldsymbol{k}+\boldsymbol{P}} t_{\boldsymbol{k}+\boldsymbol{R}} \langle \phi^{a\dagger}_{P\sigma} \phi^{c}_{P\sigma} \rangle \langle \phi^{b}_{R\sigma} \phi^{d\dagger}_{R\sigma} \rangle \langle \hat{f}_{k+Q\sigma} \hat{f}^{\dagger}_{k+Q\sigma} \rangle$$

$$+ \left(t_{\boldsymbol{k}+\boldsymbol{R}}\right)^2 \langle \phi^{a\dagger}_{P\sigma} \phi^{b}_{P\sigma} \rangle \langle \phi^{c}_{R\sigma} \phi^{d\dagger}_{R\sigma} \rangle \langle \hat{f}_{k+Q-P+R\sigma} \hat{f}^{\dagger}_{k+Q-P+R\sigma} \rangle \Big], \tag{A1}$$

where we use vector notations, $k = (i\varepsilon_n, \boldsymbol{k})$, $Q = (i\omega_\ell, \boldsymbol{P})$, $Q = (i\omega_m, \boldsymbol{Q})$, and $R = (i\omega_n, \boldsymbol{R})$. When we use the mean-field propagators for the quasiparticle $\langle \hat{f}_{k\sigma} \hat{f}^\dagger_{k\sigma} \rangle$, we obtain a simple analytic expression for the self-energy matrix as

$$\begin{aligned}
\Sigma^{cd}(k) = &-\frac{1}{N_s^3} \sum_{\boldsymbol{P},\boldsymbol{R}} \sum_{a,b=1,2} (g_{1\sigma} g_{2\sigma})^4 t_{\boldsymbol{k}+\boldsymbol{P}} t_{\boldsymbol{k}+\boldsymbol{R}} \langle {\phi^a}^\dagger_{\boldsymbol{P}\sigma} {\phi^c}_{\boldsymbol{P}\sigma} \rangle \langle \phi^b_{\boldsymbol{R}\sigma} {\phi^d}^\dagger_{\boldsymbol{R}\sigma} \rangle \\
&\times \sum_{\boldsymbol{Q}} \left[\frac{\theta(\epsilon_{\boldsymbol{k}+\boldsymbol{Q}} - \mu) Z^{ab}_-(\boldsymbol{Q})}{i\varepsilon_n - |\epsilon_{\boldsymbol{k}+\boldsymbol{Q}} - \mu| - |\lambda_{\boldsymbol{Q}}|} + \frac{\theta(\mu - \epsilon_{\boldsymbol{k}+\boldsymbol{Q}}) Z^{ab}_+(\boldsymbol{Q})}{i\varepsilon_n + |\epsilon_{\boldsymbol{k}+\boldsymbol{Q}} - \mu| + |\Lambda_{\boldsymbol{Q}}|} \right] \\
&+ \frac{1}{N_s^3} \sum_{\boldsymbol{P},\boldsymbol{Q},\boldsymbol{R}} \sum_{a,b=1,2} (g_{1\sigma} g_{2\sigma})^4 (t_{\boldsymbol{k}+\boldsymbol{R}})^2 \left[\frac{\theta(\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}} - \mu) Z^{ab}_-(\boldsymbol{Q}) W^{ab}_+(\boldsymbol{P}) W^{cd}_-(\boldsymbol{R})}{i\varepsilon_n - |\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}} - \mu| - |\lambda_{\boldsymbol{Q}}| - |\ell_{\boldsymbol{P}}| - |\ell_{\boldsymbol{R}}|} \right. \\
&\left. + \frac{\theta(\mu - \epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}}) Z^{ab}_+(\boldsymbol{Q}) W^{ab}_-(\boldsymbol{P}) W^{cd}_+(\boldsymbol{R})}{i\varepsilon_n + |\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}} - \mu| + |\Lambda_{\boldsymbol{Q}}| + |\ell_{\boldsymbol{P}}| + |\ell_{\boldsymbol{R}}|} \right],
\end{aligned} \tag{A2}$$

where the propagators for the spin bosons are given by

$$-\langle {\phi^a}^\dagger_Q \phi^b{}_Q \rangle = \frac{W^{ab}_+(\boldsymbol{Q})}{i\omega_m - |\ell_{\boldsymbol{Q}}|} - \frac{W^{ab}_-(\boldsymbol{Q})}{i\omega_m + |\ell_{\boldsymbol{Q}}|}, \tag{A3}$$

$$-\langle {\beta^a}_Q \beta^{b\dagger}_Q \rangle = \frac{Z^{ab}_+(\boldsymbol{Q})}{i\omega_m - |\Lambda_{\boldsymbol{Q}}|} - \frac{Z^{ab}_-(\boldsymbol{Q})}{i\omega_m + |\lambda_{\boldsymbol{Q}}|}, \tag{A4}$$

where the coefficient matrices $W^{ab}_\pm$ and $Z^{ab}_\pm$ are given as

$$\begin{pmatrix} W^{11}_\pm(\boldsymbol{Q}) & W^{12}_\pm(\boldsymbol{Q}) \\ W^{21}_\pm(\boldsymbol{Q}) & W^{22}_\pm(\boldsymbol{Q}) \end{pmatrix} = \pm\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} - \frac{\delta\lambda - a_1 \frac{|\epsilon|}{2} \epsilon_{\boldsymbol{Q}}}{2\ell_{\boldsymbol{Q}}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{b_1 \frac{|\epsilon|}{2} \epsilon_{\boldsymbol{Q}}}{2\ell_{\boldsymbol{Q}}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{A5}$$

$$\begin{pmatrix} Z^{11}_\pm(\boldsymbol{Q}) & Z^{12}_\pm(\boldsymbol{Q}) \\ Z^{21}_\pm(\boldsymbol{Q}) & Z^{22}_\pm(\boldsymbol{Q}) \end{pmatrix} = \frac{\delta\lambda + \delta U/2}{2\sigma_{\boldsymbol{Q}}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pm \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} - \frac{c_1 \frac{|\epsilon|}{2} \epsilon_{\boldsymbol{Q}}}{2\sigma_{\boldsymbol{Q}}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{d_1 \frac{|\epsilon|}{2} \epsilon_{\boldsymbol{Q}}}{2\sigma_{\boldsymbol{Q}}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{A6}$$

The parameters used in the above equations are given as $\delta\lambda = \lambda^{(1)} - \lambda^{(2)}$, $\delta U = U - 2\lambda^{(2)}$, $|\epsilon| = |\frac{T}{N_s} \sum_{\boldsymbol{k}, i\varepsilon_n} \epsilon_k \mathcal{G}^{(f)}_\sigma(\boldsymbol{k}, i\varepsilon_n)|$,

$$\ell_{\boldsymbol{Q}} = \sqrt{\left(\delta\lambda - a_1 \frac{|\epsilon|}{2} \epsilon_{\boldsymbol{Q}} \right)^2 - b_1^2 \frac{|\epsilon|^2}{4} \epsilon^2_{\boldsymbol{Q}}}, \tag{A7}$$

$$\sigma_{\boldsymbol{Q}} = \sqrt{\left(\lambda^{(1)} + \frac{\delta U}{2} - c_1 \frac{|\epsilon|}{2} \epsilon_{\boldsymbol{Q}} \right)^2 - d_1^2 \frac{|\epsilon|^2}{4} \epsilon^2_{\boldsymbol{Q}}}, \tag{A8}$$

$$\Lambda_{\boldsymbol{Q}} = \frac{\delta U}{2} + \sigma_{\boldsymbol{Q}}, \tag{A9}$$

$$\lambda_{\boldsymbol{Q}} = -\frac{\delta U}{2} + \sigma_{\boldsymbol{Q}}. \tag{A10}$$

The coefficients a_1, b_1, c_1, and d_1 are determined as,

$$a_1 = \overline{e}_0^2 + \overline{d}_0^2 + \langle \widetilde{e}_i \widetilde{e}^\dagger_j \rangle + \langle \widetilde{d}^\dagger_i \widetilde{d}_j \rangle, \tag{A11}$$

$$b_1 = 2\overline{e}_0 \overline{d}_0 + \langle \widetilde{e}_i \widetilde{d}_j \rangle + \langle \widetilde{d}^\dagger_i \widetilde{e}^\dagger_j \rangle, \tag{A12}$$

$$c_1 = 2\overline{p}_0^2 + \langle \widetilde{p}^\dagger_{i\sigma} \widetilde{p}_{j\sigma} \rangle + \langle \widetilde{p}_{i\bar{\sigma}} \widetilde{p}^\dagger_{j\bar{\sigma}} \rangle, \tag{A13}$$

$$d_1 = 2\overline{p}_0^2 + \langle \widetilde{p}_{i\sigma} \widetilde{p}_{j\bar{\sigma}} \rangle + \langle \widetilde{p}^\dagger_{i\bar{\sigma}} \widetilde{p}^\dagger_{j\sigma} \rangle, \tag{A14}$$

where we only take into account nearest-neighbor pairs for (i,j).

To derive Eq. (A2), we also use following relations:

$$T\sum_{i\omega_m}\frac{1}{i\varepsilon_n+i\omega_m-\xi}\frac{1}{i\omega_m-\lambda_1}=\frac{\theta(\xi)\theta(-\lambda_1)-\theta(-\xi)\theta(\lambda_1)}{i\varepsilon_n-\xi+\lambda} \tag{A15}$$

and

$$\begin{aligned}&T^3\sum_{i\omega_\ell}\sum_{i\omega_m}\sum_{i\omega_n}\frac{1}{i\varepsilon_n+i\omega_m-i\omega_\ell+i\omega_n-\xi}\frac{1}{i\omega_m-\lambda_1}\frac{1}{i\omega_\ell-\lambda_2}\frac{1}{i\omega_n-\lambda_3}\\&=\frac{\theta(\xi)\theta(-\lambda_1)\theta(\lambda_2)\theta(-\lambda_3)-\theta(-\xi)\theta(\lambda_1)\theta(-\lambda_2)\theta(\lambda_3)}{i\varepsilon_n-\xi+\lambda_1-\lambda_2+\lambda_3}.\end{aligned} \tag{A16}$$

Derivatives of $\Sigma^{cd}(k)$ with respect to $i\varepsilon_n$ give γ_k after taking a limit $i\varepsilon_n\to 0$, as is mentioned in Sec. III C. Here we give the derivatives as

$$\begin{aligned}\left.\frac{\partial\Sigma^{cd}}{\partial i\varepsilon_n}\right|_{i\varepsilon_n\to 0}&=\frac{1}{N_{\rm s}^3}\sum_{\boldsymbol{P},\boldsymbol{R}}\sum_{a,b=1,2}(g_{1\sigma}g_{2\sigma})^4\,t_{\boldsymbol{k}+\boldsymbol{P}}t_{\boldsymbol{k}+\boldsymbol{R}}\langle\phi^{a\dagger}{}_{\boldsymbol{P}\sigma}\phi^{c}{}_{\boldsymbol{P}\sigma}\rangle\langle\phi^{b}_{\boldsymbol{R}\sigma}\phi^{d\dagger}{}_{\boldsymbol{R}\sigma}\rangle\\&\times\sum_{\boldsymbol{Q}}\left[\frac{\theta(\epsilon_{\boldsymbol{k}+\boldsymbol{Q}}-\mu)Z_-^{ab}(\boldsymbol{Q})}{(|\epsilon_{\boldsymbol{k}+\boldsymbol{Q}}-\mu|+|\lambda_{\boldsymbol{Q}}|)^2}+\frac{\theta(\mu-\epsilon_{\boldsymbol{k}+\boldsymbol{Q}})Z_+^{ab}(\boldsymbol{Q})}{(|\epsilon_{\boldsymbol{k}+\boldsymbol{Q}}-\mu|+|\Lambda_{\boldsymbol{Q}}|)^2}\right]\\&-\frac{1}{N_{\rm s}^3}\sum_{\boldsymbol{P},\boldsymbol{Q},\boldsymbol{R}}\sum_{a,b=1,2}(g_{1\sigma}g_{2\sigma})^4\,(t_{\boldsymbol{k}+\boldsymbol{R}})^2\left[\frac{\theta(\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}}-\mu)Z_-^{ab}(\boldsymbol{Q})W_+^{ab}(\boldsymbol{P})W_-^{cd}(\boldsymbol{R})}{(|\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}}-\mu|+|\lambda_{\boldsymbol{Q}}|+|\ell_{\boldsymbol{P}}|+|\ell_{\boldsymbol{R}}|)^2}\right.\\&\left.+\frac{\theta(\mu-\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}})Z_+^{ab}(\boldsymbol{Q})W_-^{ab}(\boldsymbol{P})W_+^{cd}(\boldsymbol{R})}{(|\epsilon_{\boldsymbol{k}+\boldsymbol{Q}-\boldsymbol{P}+\boldsymbol{R}}-\mu|+|\Lambda_{\boldsymbol{Q}}|+|\ell_{\boldsymbol{P}}|+|\ell_{\boldsymbol{R}}|)^2}\right].\end{aligned} \tag{A17}$$

APPENDIX B: HYBRIDIZATION

The amplitudes of hybridization between quasiparticles and cofermions

$$\boldsymbol{\Delta}_{ij}^T=\left(\Delta_{ij}^{(\psi)},\Delta_{ij}^{(\chi)}\right)^T \tag{B1}$$

are given as $(\Delta_{\boldsymbol{k}}^1,\Delta_{\boldsymbol{k}}^2)=(\Delta_{\boldsymbol{k}}^{(\psi)},\Delta_{\boldsymbol{k}}^{(\chi)})$. Its Fourier transformation is given as

$$\Delta_{\boldsymbol{k}}^d=\frac{T^2}{N_{\rm s}^2}\sum_{P,R}\sum_{\boldsymbol{q}}\sum_{a,b,c=1,2}t_{\boldsymbol{k}+\boldsymbol{R}}\left(g_{1\sigma}^2g_{2\sigma}^2\right)^2\left\langle\beta^a_{P-R}\beta^{b\dagger}{}_{P-R}\right\rangle\left\langle\phi^{a\dagger}{}_{P\sigma}\phi^{b}{}_{P\sigma}\right\rangle\left\langle\phi^{d}_{R\sigma}\phi^{c\dagger}{}_{R\sigma}\right\rangle b_0^c t_{\boldsymbol{q}+\boldsymbol{R}}\langle\hat{f}^\dagger_{\boldsymbol{q}\sigma}\hat{f}_{\boldsymbol{q}\sigma}\rangle, \tag{B2}$$

where we use vector notations $\boldsymbol{b}_0=(b_0^1,b_0^2)=(\overline{e}_0,\overline{d}_0)$, $\boldsymbol{\beta}_i=(\beta_i^1,\beta_i^2)=(\widetilde{e}_i,\widetilde{d}_i^\dagger)$, and $\boldsymbol{\phi}_i=(\phi_i^1,\phi_i^2)=(\widetilde{p}_{i\sigma},\widetilde{p}_{i\overline{\sigma}}^\dagger)$. Here, if we use the mean-field propagators for the quasiparticle, we obtain

$$\begin{aligned}\Delta_{\boldsymbol{k}}^d=&-\frac{1}{N_{\rm s}^2}\sum_{\boldsymbol{P},\boldsymbol{R},\boldsymbol{q}}\sum_{a,b,c=1,2}t_{\boldsymbol{k}+\boldsymbol{R}}\left(g_{1\sigma}^2g_{2\sigma}^2\right)^2t_{\boldsymbol{q}+\boldsymbol{R}}n_{\boldsymbol{q}}\\&\times\left[\frac{Z_+^{ab}(\boldsymbol{P}-\boldsymbol{R})W_-^{ab}(\boldsymbol{P})W_+^{dc}(\boldsymbol{R})}{|\Lambda_{\boldsymbol{P}-\boldsymbol{R}}|+|\ell_{\boldsymbol{P}}|+|\ell_{\boldsymbol{R}}|}+\frac{Z_-^{ab}(\boldsymbol{P}-\boldsymbol{R})W_+^{ab}(\boldsymbol{P})W_-^{dc}(\boldsymbol{R})}{|\lambda_{\boldsymbol{P}-\boldsymbol{R}}|+|\ell_{\boldsymbol{P}}|+|\ell_{\boldsymbol{R}}|}\right],\end{aligned} \tag{B3}$$

where we use the following relation:

$$T^2\sum_{i\omega_\ell}\sum_{i\omega_n}\frac{1}{i\omega_\ell-i\omega_n-\lambda_1}\frac{1}{i\omega_\ell-\lambda_2}\frac{1}{i\omega_n-\lambda_3}=\frac{\theta(\lambda_1)\theta(-\lambda_2)\theta(\lambda_3)+\theta(-\lambda_1)\theta(\lambda_2)\theta(-\lambda_3)}{\lambda_1-\lambda_2+\lambda_3}. \tag{B4}$$

Equation (B3) will help us to estimate the amplitude of the pseudogap in our theory.

*Present address: Department of Physics, Rutgers University, Piscataway, New Jersey 08854, USA.
[1]N. P. Ong, Z. Z. Wang, J. Clayhold, J. M. Tarascon, L. H. Greene, and W. R. Mckinnon, Phys. Rev. B **35**, 8807 (1987).
[2]H. Takagi, T. Ido, S. Ishibashi, M. Uota, S. Uchida, and Y. Tokura, Phys. Rev. B **40**, 2254 (1989).
[3]H. Yasuoka, T. Imai, and T. Shimizu, in *Strong Correlation and Superconductivity*, edited by H. Fukuyama, S. Maekawa, and A. P. Malozemoff (Springer, Berlin, 1989), p. 254.
[4]J. Rossat-Mignod, L. P. Regnault, C. Vettier, P. Burlet, J. Y. Henry, and G. Lapertot, Phys. B (Amsterdam) **169**, 58 (1991).
[5]J. W. Loram, K. A. Mirza, J. R. Cooper, and W. Y. Liang, Phys. Rev. Lett. **71**, 1740 (1993).
[6]C. C. Homes, T. Timusk, R. Liang, D. A. Bonn, and W. N. Hardy, Phys. Rev. Lett. **71**, 1645 (1993).
[7]T. Nishikawa, J. Takeda, and M. Sato, J. Phys. Soc. Jpn. **62**, 2568 (1993).
[8]D. S. Marshall, D. S. Dessau, A. G. Loeser, C.-H. Park, A. Y. Matsuura, J. N. Eckstein, I. Bozovic, P. Fournier, A. Kapitulnik, W. E. Spicer *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).
[9]A. Damascelli, Z. Hussain, and Z.-X. Shen, Rev. Mod. Phys. **75**, 473 (2003).
[10]T. Yoshida, X. J. Zhou, K. Tanaka, W. L. Yang, Z. Hussain, Z.-X. Shen, A. Fujimori, S. Sahrakorpi, M. Lindroos, R. S. Markiewicz *et al.*, Phys. Rev. B **74**, 224510 (2006).
[11]R. Peierls, Proc. Phys. Soc., London, Sect. A **49**, 72 (1937).
[12]N. F. Mott, Proc. Phys. Soc., London, Sect. A **49**, 72 (1937).
[13]M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998).
[14]A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[15]J. Hubbard, Proc. R. Soc. London, Ser. A **281**, 401 (1964).
[16]W. F. Brinkman and T. M. Rice, Phys. Rev. B **2**, 4302 (1970).
[17]M. C. Gutzwiller, Phys. Rev. **137**, A1726 (1965).
[18]W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989).
[19] Müller-Hartmann, Z. Phys. B: Condens. Matter **74**, 507 (1989).
[20]M. Imada and S. Onoda, J. Phys. Chem. Solids **62**, 47 (2001).
[21]S. Onoda and M. Imada, J. Phys. Chem. Solids **63**, 2225 (2002).
[22]D. Sénéchal and A.-M. S. Tremblay, Phys. Rev. Lett. **92**, 126401 (2004).
[23]T. D. Stanescu and G. Kotliar, Phys. Rev. B **74**, 125110 (2006).
[24]Y. Z. Zhang and M. Imada, Phys. Rev. B **76**, 045108 (2007).
[25]S. Sakai, Y. Motome, and M. Imada, Phys. Rev. Lett. **102**, 056404 (2009).
[26]S. Sakai, Y. Motome, and M. Imada, Phys. Rev. B **82**, 134505 (2010).
[27]Y. Yamaji and M. Imada, Phys. Rev. Lett. **106**, 016404 (2011).
[28]I. Dzyaloshinskii, Phys. Rev. B **68**, 085113 (2003).
[29]T. M. Rice, Suppl. Prog. Theor. Phys. **160**, 39 (2005).
[30]J. Meng, G. Liu, W. Zhang, L. Zhao, H. Liu, X. Jia, D. Mu, S. Liu, X. Dong, J. Zhang *et al.*, Nature (London) **462**, 335 (2009).
[31]B. Fauqué, Y. Sidis, V. Hinkov, S. Pailhès, C. T. Lin, X. Chaud, and P. Bourges, Phys. Rev. Lett. **96**, 197001 (2006).
[32]J. Xia, E. Schemm, G. Deutscher, S. A. Kivelson, D. A. Bonn, W. N. Hardy, R. Liang, W. Siemons, G. Koster, M. M. Fejer *et al.*, Phys. Rev. Lett. **100**, 127002 (2008).
[33]Y. Li, V. Balédent, N. Barišić, Y. Cho, B. Fauqué, Y. Sidis, G. Yu, X. Zhao, P. Bourges, and M. Greven, Nature (London) **455**, 372 (2008).
[34]R. Daou, J. Chang, D. LeBoeuf, O. Cyr-Choinière, F. Laliberté, N. Doiron-Leyraud, B. J. Ramshaw, R. Liang, D. A. Bonn, W. N. Hardy *et al.*, Nature (London) **463**, 517 (2010).
[35]G. J. MacDougall, A. A. Aczel, J. P. Carlo, T. Ito, J. Rodriguez, P. L. Russo, Y. J. Uemura, S. Wakimoto, and G. M. Luke, Phys. Rev. Lett. **101**, 017001 (2008).
[36]M. B. J. Meinders, H. Eskes, and G. A. Sawatzky, Phys. Rev. B **48**, 3916 (1993).
[37]S. Uchida, T. Ido, H. Takagi, T. Arima, Y. Tokura, and S. Tajima, Phys. Rev. B **43**, 7942 (1991).
[38]K. Tanaka, W. S. Lee, D. H. Lu, A. Fujimori, T. Fujii, Risdiana, I. Terasaki, D. J. Scalapino, T. P. Devereaux, Z. Hussain *et al.*, Science **314**, 1910 (2006).
[39]H.-B. Yang, J. D. Rameau, P. D. Johnson, T. Valla, A. Tsvelik, and G. D. Gu, Nature (London) **456**, 77 (2008).
[40]J. Kanamori, Prog. Theor. Phys. **30**, 275 (1963).
[41]E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[42]R. Preuss, W. Hanke, C. Gröber, and H. G. Evertz, Phys. Rev. Lett. **79**, 1122 (1997).
[43]M. H. Hettler, A. N. Tahvildar-Zadeh, M. Jarrell, T. Pruschke, and H. R. Krishnamurthy, Phys. Rev. B **58**, R7475 (1998).
[44]G. Kotliar, S. Y. Savrasov, G. Palsson, and G. Biroli, Phys. Rev. Lett. **87**, 186401 (2001).
[45]C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, Phys. Rev. Lett. **98**, 110201 (2007).
[46]D. Tahara and M. Imada, J. Phys. Soc. Jpn. **77**, 114701 (2008).
[47]T. Aimi and M. Imada, J. Phys. Soc. Jpn. **76**, 084709 (2007).
[48]T. Aimi and M. Imada, J. Phys. Soc. Jpn. **76**, 113708 (2007).
[49]T. Kashima and M. Imada, J. Phys. Soc. Jpn. **70**, 2287 (2001).
[50]T. Kashima and M. Imada, J. Phys. Soc. Jpn. **70**, 3052 (2001).
[51]T. Mizusaki and M. Imada, Phys. Rev. B **74**, 014421 (2006).
[52]L. D. Landau, Zh. Eksperim. i Teor. Fiz **30**, 1058 (1956) [Sov. Phys. JETP **3**, 920 (1957)].
[53]N. F. Mott, Proc. Phys. Soc., London, Sect. A **62**, 416 (1949).
[54]N. F. Mott, Can. J. Phys. **34**, 1356 (1956).
[55]N. F. Mott, Philos. Mag. **6**, 287 (1961).
[56]G. Kotliar and A. E. Ruckenstein, Phys. Rev. Lett. **57**, 1362 (1986).
[57]R. Raimondi and C. Castellani, Phys. Rev. B **48**, 11453 (1993).
[58]D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984).
[59]C. Castellani, G. Kotliar, R. Raimondi, M. Grilli, Z. Wang, and M. Rozenberg, Phys. Rev. Lett. **69**, 2009 (1992).
[60]N. N. Bogoliubov, J. Phys. (USSR) **11**, 23 (1947).
[61]A. Abrikosov, L. Gorfkov, and I. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1975).
[62]N. Furukawa, T. M. Rice, and M. Salmhofer, Phys. Rev. Lett. **81**, 3195 (1998).
[63]P. Phillips, T.-P. Choy, and R. G. Leigh, Rep. Prog. Phys. **72**, 036501 (2009).
[64]T. A. Kaplan, P. Horsch, and P. Fulde, Phys. Rev. Lett. **49**, 889 (1982).
[65]H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **59**, 3669 (1990).
[66]P. A. Lee, N. Nagaosa, and X.-G. Wen, Rev. Mod. Phys. **78**, 17 (2006).
[67]Th. Jolicoeur and J. C. Le Guillou, Phys. Rev. B **44**, 2403 (1991).
[68]K. Waku, T. Katsufuji, Y. Kohsaka, T. Sasagawa, H. Takagi, H. Kishida, H. Okamoto, M. Azuma, and M. Takano, Phys. Rev. B **70**, 134501 (2004).

[69]V. J. Emery and S. A. Kivelson, Phys. C (Amsterdam) **209**, 597 (1993).
[70]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).
[71]B. Kyung, S. S. Kancharla, D. Sénéchal, A.-M. S. Tremblay, M. Civelli, and G. Kotliar, Phys. Rev. B **73**, 165114 (2006).
[72]J. Loram, J. Luoa, J. Coopera, W. Lianga, and J. Tallon, J. Phys. Chem. Solids **62**, 59 (2001).
[73]N. Momono, M. Ido, T. Nakano, M. Oda, Y. Okajima, and K. Yamaya, Phys. C (Amsterdam) **233**, 395 (1994).
[74]M. R. Norman, D. Pines, and C. Kallin, Adv. Phys. **54**, 715 (2005).
[75]H. Ding, J. R. Engelbrecht, Z. Wang, J. C. Campuzano, S.-C. Wang, H.-B. Yang, R. Rogan, T. Takahashi, K. Kadowaki, and D. G. Hinks, Phys. Rev. Lett. **87**, 227001 (2001).
[76]A. Ino, C. Kim, T. Mizokawa, Z.-X. Shen, A. Fujimori, M. Takaba, K. Tamasaku, H. Eisaki, and S. Uchida, J. Phys. Soc. Jpn. **68**, 1496 (1999).
[77]Y. Yanase, T. Jujo, T. Nomura, H. Ikeda, T. Hotta, and K. Yamada, Phys. Rep. **387**, 1 (2003).
[78]M. Imada, Phys. Rev. B **72**, 075113 (2005).
[79]T. Misawa, Y. Yamaji, and M. Imada, J. Phys. Soc. Jpn. **75**, 083705 (2006).
[80]Y. Yamaji, T. Misawa, and M. Imada, J. Phys. Soc. Jpn. **75**, 094719 (2006).
[81]T. Misawa and M. Imada, Phys. Rev. B **75**, 115121 (2007).
[82]C. Renner, B. Revaz, J.-Y. Genoud, K. Kadowaki, and Ø. Fischer, Phys. Rev. Lett. **80**, 149 (1998).
[83]T. Hanaguri, C. Lupien, Y. Kohsaka, D.-H. Lee, M. Azuma, M. Takano, H. Takagi, and J. C. Davis, Nature (London) **430**, 1001 (2004).
[84]C. Proust, K. Behnia, R. Bel, D. Maude, and S. I. Vedeneev, Phys. Rev. B **72**, 214511 (2005).
[85]M. Imada, J. Phys. Soc. Jpn. **62**, 1105 (1993).
[86]X.-G. Wen and P. A. Lee, Phys. Rev. Lett. **76**, 503 (1996).
[87]T. Tohyama, Phys. Rev. B **70**, 174517 (2004).
[88]K.-S. Kim and C. Pépin, Phys. Rev. Lett. **102**, 156404 (2009).
[89]R. W. Hill, C. Proust, L. Taillefer, P. Fournier, and R. L. Greene, Nature (London) **414**, 711 (2001).
[90]T.-P. Choy, R. G. Leigh, P. Phillips, and P. D. Powell, Phys. Rev. B **77**, 014512 (2008).
[91]J. Wagner, W. Hanke, and D. J. Scalapino, Phys. Rev. B **43**, 10517 (1991).