Nonperturbative approach to full Mott behavior

Tudor D. Stanescu
Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, New Jersey 08854-8019, USA

Philip Phillips
Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, 1100 West Green Street, Urbana, Illinois 61801-3080, USA

(Received 15 January 2003; revised manuscript received 11 July 2003; published 15 June 2004)

Though most fermionic Mott insulators order at low temperatures, ordering is ancillary to their insulating behavior. Our emphasis here is on disentangling ordering from the intrinsic strongly correlated physics of a doped half-filled band. To this end, we focus on the two-dimensional Hubbard model. Because the charge gap arises from on-site correlations, we have been refining the nonperturbative approach of Matsumoto and Mancini [Phys. Rev. B **55**, 2095 (1997)] which incorporates local physics. Crucial to this method is a self-consistent two-site dynamical cluster expansion which builds in the nearest-neighbor energy scale J. At half-filling, we find that the spectral function possesses a gap of order U and is devoid of any coherent quasiparticle peaks although ordering or charge fractionalization are absent. At low temperatures, local antiferromagnetic correlations emerge. In the doped case, we find that the Fermi surface exceeds the Luttinger volume. The breakdown of Luttinger's theorem in the underdoped regime is traced both to the dynamically generated Mott gap as well to a nonvanishing of the imaginary part of the self-energy at the Fermi level. Spectral weight transfer across the Mott gap also emerges as a ubiquitous feature of a doped Mott insulator and suggests that high- and low-energy scales are inseparable. Additionally in the underdoped regime, we find that a pseudogap exists in the single-particle density of states as well as in the heat capacity. The pseudogap (which is set by the energy scale t^2/U) is argued to be a ubiquitous feature of a lightly doped Mott state and simply represents the fact that hole transport involves double occupancy. In analogy with the Mott gap and antiferromagnetism, we propose that ordering may also accompany the formation of a pseudogap. We suggest a current pattern within a one-band model that preserves translational but breaks time-reversal symmetry along the canonical x and y axes but not along $x=\pm y$ that is consistent with the experimental observations. Finally, we show that the Hall coefficient in a doped Mott insulator must change sign at a doping level $x<1/3$. The sign change is tied to a termination of strong correlation physics in the doped Mott state.

DOI: 10.1103/PhysRevB.69.245104 PACS number(s): 71.10.Pm

I. INTRODUCTION

Electronic systems with an odd number of particles per unit cell are typically metallic at zero temperature. However, Mott[1] proposed that such systems in the presence of strong interactions can insulate at zero temperature without any accompanying symmetry breaking process, such as antiferromagnetism or charge ordering, which necessarily double the unit cell. Strictly then, a Mott insulator (should it exist) is a paramagnetic state with an odd number of particles per unit cell. Insulating behavior arises from the charge gap generated by the projective mismatch between the sublattices which have zero or some finite fraction of doubly occupied sites. When the overlap between such sublattices is sufficiently small, no transport obtains. In bosonic systems with a single boson per site, a true zero-temperature Mott insulating state is realized in the standard quantum rotor model or Bose-Hubbard model when the on-site charging energy exceeds a critical value such that phase coherence is destroyed.[2–4] In fact, the recent observation[5] that a Bose condensate in an optical lattice can be tuned between a superfluid and a Mott insulator simply by changing the intensity of the laser light places the bosonic Mott insulator on firm experimental footing.

However, for fermionic systems, the inherent problem with the Mott insulating state of a half-filled band in $d>1$ is its proclivity to order at zero temperature. Consequently, it is tempting to equate the Mott insulator with the ordered state or to assert categorically that the Mott insulator as an entity distinct from a symmetry-broken state never existed in the first place. In the context of high T_c in the cuprates, which are all antiferromagnetic Mott insulators, both views[6] have been strongly expressed. In fact, a large body of work on the cuprates has focused primarily on models that capture low-energy spin physics[7–9] at or near an antiferromagnet or a charge-ordered state,[10–18] or a classification[19,20] of the various charge-ordered states that ensue in a half-filled band. Alternatively, low-energy spin liquid models[21–28] (that is, models with spin translation and spin rotation symmetry) have been proposed as candidates to describe the Mott insulator. In such approaches, the high-energy scale associated with the charge gap is argued to be irrelevant, hence the focus exclusively on the spin sector to characterize the Mott insulator. Of course such an approach presupposes that the high- and low-energy degrees of freedom can be disentangled.

Should the insulating state in a half-filled band prove to be nothing other than a mean-field broken-symmetry state, then fermionic Mott insulators do not exist. Hence, a relevant question for the Mott insulating state is as follows: If we subtract the fact that ordering obtains at zero temperature, is there anything left over that is not explained by ordering?

Equivalently, does ordering provide an exhaustive explanation of the state proposed by Mott? We refer to whatever might be left over once we subtract the fact that ordering has occurred as *Mottness*. That something might be left over is immediately evident from the fact that charge- and spin-ordered states all result in a doubling of the unit cell and hence are adiabatically connected to an insulator with an even number of electrons per unit cell. A typical example of such a system is a band insulator. On the contrary, Mott insulators which have an odd number of electrons per unit cell are not. Additionally, spin- and charge-ordered states can be described at the level of Hartree-Fock simply by constructing the correct broken-symmetry ansatz. However the Mott state has no weak coupling or Hartree-Fock counterpart. Hence, although the Mott state might be unstable to ordering at low temperature, some features should be left over which represent the fingerprint of the nonadiabaticity with a band insulator and the fact that this state arises fundamentally from strong electron correlation. For example, experimentally,[29,30] it is clear that above any temperature associated with ordering in both the electron and hole-doped cuprates, a charge gap of order 2 eV is present in the optical conductivity and oxygen *K*-edge photoemission.[31] Hence, the vanishing of the low-energy (<1 eV) spectral weight at high temperature is not linked to magnetism or ordering of any kind. Further, the electronic bands below and above the charge gap are not rigid as would be the case in a band insulator. To illustrate, as a function of doping, in both hole and electron-doped cuprates,[29,30] the low-energy spectral weight increases at the expense of the high-energy (>2 eV) spectral weight such that the total integrated optical conductivity remains constant up to 4 eV. The same massive reshuffling of spectral weight from 2 eV above the Fermi energy is also observed in one-particle probes such as oxygen *K*-edge photoemission[31] and angle-resolved photoelectron spectroscopy (ARPES).[32,33] Such spectral weight transfer indicates that the total number of low-energy degrees of freedom in the normal state of the cuprates cannot be decoupled from the high-energy scales. What is surprising about the cuprates is that even when superconductivity obtains, the low- and high-energy degrees of freedom are still coupled. For example, Rübhaussen *et al.*[34] have shown that changes in the optical conductivity occur at energies 3 eV (roughly 100Δ, where Δ is the maximum superconducting gap) away from the Fermi energy at T_c, and Bolegräf *et al.*[35] have seen an acceleration in the depletion of the high-energy spectral weight accompanied with a compensating increase in the low-energy spectral weight at and below the superconducting transition. Similarly, Bontemps *et al.*[36] have directly observed that in underdoped (but not overdoped) BSCO, the Glover-Ferrel-Tinkham sum rule is violated and the optical conductivity must be integrated to 20 000 cm^{-1} to recover the spectral weight lost upon condensation into the superconducting state. In a standard BCS superconductor, condensation leads to loss of spectral weight at energy scales no more than ten times the pairing energy. The fact that pair condensation perturbs the optical conductivity on energy scales as large as 100Δ suggests that there is a direct link between the high-energy Mott scale and superconductivity. Further, the persistence of spectral weight transfer through the superconducting transition indicates that the properties of the Mott state remain intact even in the presence of ordering. While spin-density wave antiferromagnets also possess two bands with a gap, such a state is insufficient to explain the origin of the spectral weight transfer in the cuprates. The reason is simple: spectral weight transfer persists well above T_N and at a doping level $\gg 2\%$, where antiferromagnetism is absent.

Concretely, what the optical conductivity and oxygen *k*-edge photoemission experiments on the cuprates lay plain is that regardless of whether ordering obtains at sufficiently low temperatures, the charge degrees of freedom remain characterized by a distinctly different state than that of any ordered state or a band insulator. To understand what physics is entailed by this state, consider the Mott mechanism for generating an insulating state in the one-band Hubbard model. At half-filling, the chemical potential lies in the middle of the gap separating the lower and upper Hubbard bands, which are dynamically split by the on-site energy for double occupancy. However, half the spectral weight resides in each of the bands. Consequently, to satisfy the sum rule that each state in the first Brillouin Zone (FBZ) carries unit spectral weight, the spectral function must be integrated across the charge gap not simply up to the chemical potential. Hence, the half-filled state is characterized by the Fermi energy lying in a gap but partially occupied states exist. It is this seemingly contradictory state of affairs that is at the heart of Mottness. Spectral weight transfer cannot obtain without it. For example, if each state below the chemical potential had unit spectral weight, no state would be available for spectral weight transfer from high energies. As a consequence, adding or removing an electron cannot be done without affecting both high- and low-energy scales. Consequently, at any doping level, the electronic states describing the charge carriers can be written as linear combinations of excitations living in both the lower Hubbard band (LHB) and upper Hubbard band (UHB) as will be detailed below. As a result, in the Mott state, the traditional notion that the chemical potential demarcates the boundary between zero and unit occupancy fails fundamentally. Of course, in Fermi liquids, the spectral function for each $\mathbf{k}$ state can also have an incoherent background which can extend to high energies. However, as long as a coherence peak exists, a sharp criterion exists for unit occupancy of each state, namely, whether or not the coherence peak crosses the Fermi level. In a Mott insulator, no such coherence peak exists and consequently, incoherence dominates the Mott state.

Alternatively, one can view the spectral weight transfer in real space by simply counting the number of available states for the photoemission and inverse photoemission spectra as demonstrated by Meinders, Eskes, and Sawatzky.[37] We recount the argument here as it is simple and instructive. Consider the half-filled one-dimensional chain of one-electron atoms shown in Fig. 1. Both the electron-removal (photoelectron) and electron-addition (inverse photoemission) spectral weights are equal to N because at half-filling there are N ways of adding or subtracting an electron from a site that is singly occupied. When a single hole is added, both the electron-removal and the electron-addition spectral decrease

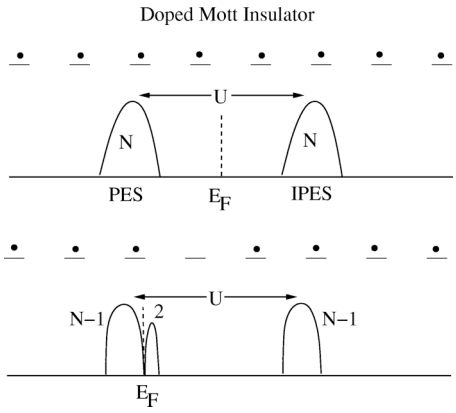

FIG. 1. Spectral weight transfer in a doped Mott insulator. The photoelectron spectrum (PES) denotes the electron removal states while the electron-addition states are located in the inverse photoelectron spectrum (IPES). The on-site charging energy is U. Removal of a single-electron results in the creation of two single-particle states at the top of the lower Hubbard band. By state conservation, one state comes from the lower and the other from the upper Hubbard band and hence spectral weight transfer across the Mott gap.

to $N-1$ as there are now $N-1$ ways to add or subtract an electron from sites that are already occupied. Hence, there are two less states. These states correspond to the spin-up and spin-down states of the empty site and hence belong to the LHB. Consequently, the low-energy spectral weight (LESW), $\Lambda(x)$, has increased by two states. One of the states must come from the UHB as the high-energy part now has a spectral weight of $N-1$. Hence, for a single hole, there is a net transfer of one state from high to low energy. This argument is simply the real-space restatement of the more general principle that in **k** space the LHB and UHB are not static but dynamic and hence necessarily give rise to spectral weight transfer. In general, simple state counting yields $2x$ for the growth of $\Lambda(x)$ and $1-x$ for the depletion of the high-energy sector. In actuality, the dynamical contribution to the LESW results in $\Lambda(x)>2x$. The dynamical LESW corresponds to virtual excitations to the UHB. Hence, in a strongly correlated system, the phase space available to add a single electron exceeds the nominal number of states initially present in the low-energy scale, leading thereby to an inseparability of the low- and high-energy scales.

The ubiquity of spectral weight transfer in the cuprates places extreme restrictions on which low-energy models are valid. Consider the standard t-J model.[21] As the t-J model projects onto the LHB, we can estimate the corresponding LESW exactly by counting the electron-removal states. Consequently, $\Lambda(x)=2x$ is exact in the t-J model. However, in the actual Hubbard model, $\Lambda(x)>2x$. Consequently, the standard t-J model does not have the correct number of low-energy degrees of freedom to describe the low-energy physics. This problem can be corrected,[38] however, by using the full strong-coupling Hamiltonian that results from the t^2/U expansion and replacing all the electron operators by their projected counterparts. However, the price one pays is that the new projected electron operators do not obey the standard fermion commutation relations. That a correct low-energy theory must give up on either standard fermion commutation relations or particle conservation has already been pointed out by Meinders, Eskes, and Sawatzky.[37] Hence, all standard low-energy fermionic theories do not have the correct physics to describe the cuprates primarily because the low-energy degrees of freedom are not fermionic.

Given the ubiquity of spectral weight transfer over a large part of the phase diagram of the cuprates, it is imperative that any theory of high T_c incorporate the high-energy scale associated with the charge gap. Further, because ordering does not seem to be a requirement for spectral weight transfer, our focus is on an accurate description of the high-temperature charge vacuum of the doped Mott insulating state. It is crucial in such an approach that the hierarchy of energy scales, U, t, t^2/U, for example, emerge. Consequently, we have been refining the nonperturbative approach of Matsumoto and Mancini[39] to describe the interrelation between the energy scales in a doped Mott insulator. This approach is based on the experimental observation that the low-energy scales in doped Mott insulators are derived from high energies. As a result, the beginning point, namely, the Hubbard operators, is one that is well known to yield the Mott charge gap scale. Successively smaller energy scales are derived by treating local correlations on a small cluster. The self-energy of the lattice is then determined self-consistently from the local impurity problem. The Matsumoto-Mancini[39] approach is then in the spirit of the cellular dynamical mean-field treatment.[40] We report here the full details of this approach and catalogue the general results that follow from Mottness. Aside from a LESW that exceeds the nominal value obtained from state counting, we find that (1) Mottness gives rise to broad spectral features in the underdoped regime, (2) a violation of Luttinger's theorem in the underdoped regime, (3) holelike Fermi surface near half-filling, (4) a jump in the chemical potential upon doping, and (5) a pseudogap in the underdoped regime without invoking any symmetry breaking. In general, we find that the pseudogap is due strictly to near-neighbor correlations and can be thought of as the nearest-neighbor analogue of the on-site generated Mott gap.

II. METHOD

Many years ago, Hubbard[41] wrote the electron annihilation operator $c_{i\sigma}$ as a linear combination

$$c_{i\sigma}=\eta_{i\sigma}+\xi_{i\sigma} \tag{1}$$

of two composite excitations that reflect the energetically challenged landscape an electron must traverse in the presence of a large on-site Coulomb repulsion, U. Physically, the operators $\eta_{i\sigma}=c_{i\sigma}n_{i-\sigma}$ and $\xi_{i\sigma}=c_{i\sigma}(1-n_{i-\sigma})$ represent an electron moving on doubly and singly occupied sites, respectively. Because such sites are split by U, the Hubbard operators lead naturally to a gap at half-filling in a paramagnetic state, a result which thus far, only two other methods, dynamical mean-field theories (DMFT) (Ref. 42) and quantum Monte Carlo (QMC),[43,44] have been able to obtain. Regardless of this success, the Hubbard operators have been criticized extensively because untested approximations generally

accompany their implementation, and they lead to a Fermi surface which does not preserve the Luttinger volume. However, methods that are deemed as reliable,[45–48] also find that in the lightly doped regime, Luttinger's theorem is violated as is seen experimentally.[32] Hence, a violation of Luttinger's theorem is not an *a priori* reason to dismiss the Hubbard operators. In fact, any violation of Luttinger's theorem must occur in the lightly doped or nonperturbative regime. The untested or uncontrolled approximations usually arise from truncations in the equations of motion. However, such problems can be circumvented by the following procedure. First, project all new operators that arise from the Heisenberg equations of motion of the Hubbard operators onto the Hubbard basis. Second, write the self-energy exactly in terms of the remaining operators which are now orthogonal to the Hubbard basis. Third, use local DMFT methods to calculate the resultant electron self-energy. The approximation introduced in the third step is that the self-energy for a finite cluster is used to determine the self-energy for the interacting lattice. However, such methods have been shown to be strongly convergent and in fact constitute the accepted methodology for treating strongly correlated systems. In principle, as the cluster size is extended to infinity, such a cluster procedure becomes exact. Hence, an implementation of the Hubbard operators coupled with DMFT-type technology places the limitations not on truncation in the equations of motion but on the accuracy of the impurity solver and the size of the finite cluster. It is such a procedure that we outline here. As many of the details have been left out in the previous presentation in the original paper by Mancini and Matsumoto[39] on the two-site cluster and subsequent implementations,[49] we will provide a complete derivation of the method so that anyone reading this paper can implement it immediately. The key features of this method are its ability to describe physics on the scale of the Mott gap U as well as on the the scale $J \approx t^2/U$.

A. Dynamical Green function

Our starting point is the on-site Hubbard model

$$H = -\sum_{i,j,\sigma} t_{ij} c^{\dagger}_{i\sigma} c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \tag{2}$$

with nearest-neighbor hopping, $t_{ij} = t\alpha_{ij}$. We also introduce the composite operator basis

$$\psi_{\sigma}(i) = \begin{pmatrix} \xi_{i\sigma} \\ \eta_{i\sigma} \end{pmatrix} \tag{3}$$

and its associated retarded Green function $S(i,j,t,t') = \langle\langle \psi_{i\sigma}; \psi^{\dagger}_{j\sigma} \rangle\rangle = \theta(t-t')\langle\{\psi_{i\sigma}(t), \psi^{\dagger}_{j\sigma}(t')\}\rangle$. Writing the equations of motion for the Hubbard basis and projecting with the Roth[50] projector $\mathcal{P}(O) = \Sigma_{ln}\langle\{O, \psi^{\dagger}_l\}\rangle I^{-1}_{ln} \psi_n$, we obtain for the "current" operator the expression

$$j_i(t) = i\frac{\partial}{\partial t}\psi_i = K\psi_i + \mathcal{P}(\delta j_i) + \delta J_i = E\psi_i + \delta J_i. \tag{4}$$

The formal solution for the Green function in Fourier space,

$$S(\mathbf{k},\omega) = \frac{1}{\omega - \mathbf{E}(\mathbf{k}) - \delta\mathbf{m}(\mathbf{k},\omega)\mathbf{I}^{-1}(\mathbf{k})}\mathbf{I}(\mathbf{k}), \tag{5}$$

contains the self-energy $\Sigma(\mathbf{k},\omega) = \delta\mathbf{m}(\mathbf{k},\omega)\mathbf{I}^{-1}$ with

$$\delta\mathbf{m}(\mathbf{k},\omega) = \mathrm{FT}\langle\theta(t-t')\{\delta J(t), \delta J^{\dagger}(t')\}\rangle_I \tag{6}$$

where the subscript I indicates the irreducible part and FT denotes the Fourier transform. We consider the paramagnetic case, for which the overlap matrix $\mathbf{I}$ is diagonal with $I_{11} \equiv I_1 = 1 - n/2$ and $I_{22} \equiv I_2 = n/2$.

The primary operational hurdle is the evaluation of the dynamical correction $\delta\mathbf{m}$. The first step is to write the explicit expressions for the operators δJ which are "orthogonal" to the basis ψ. Using the notation $\tilde{t} = 2dt$, d being the dimensionality of the system, and $\tilde{p} = p - n^2/4$, we obtain

$$\delta J_1(i) = -\tilde{t}\Big[\pi_{i\sigma} + \frac{n}{2}c^{\alpha}_{i\sigma} - e(\xi_{i\sigma}I_1^{-1} - \eta_{i\sigma}I_2^{-1}) - \tilde{p}(\xi^{\alpha}_{i\sigma}I_1^{-1} - \eta^{\alpha}_{i\sigma}I_2^{-1})\Big] \tag{7}$$

and $\delta J_2(i) = -\delta J_1(i)$, where the self-consistent parameters e and $p = \tilde{p} + n^2/4$ as well as the higher-order composite operator π_i are given by

$$e = \langle \xi^{\alpha}_i \xi^{\dagger}_i \rangle - \langle \eta^{\alpha}_i \eta^{\dagger}_i \rangle,$$

$$p = \langle n_{i\sigma} n^{\alpha}_{i\sigma} \rangle + \langle c^{\dagger}_{i\uparrow} c_{i\downarrow} (c^{\dagger}_{i\downarrow} c_{i\uparrow})^{\alpha} \rangle - \langle c_{i\uparrow} c_{i\downarrow} (c^{\dagger}_{i\downarrow} c^{\dagger}_{i\uparrow})^{\alpha} \rangle,$$

$$\tilde{t}\begin{pmatrix} \pi_{i\uparrow} \\ \pi_{i\downarrow} \end{pmatrix} = \sum_j t_{ij} \begin{pmatrix} -n_{i\downarrow}c_{j\uparrow} + c^{\dagger}_{i\downarrow}c_{i\uparrow}c_{j\downarrow} - c_{i\uparrow}c_{i\downarrow}c^{\dagger}_{j\downarrow} \\ -n_{i\uparrow}c_{j\downarrow} + c^{\dagger}_{i\uparrow}c_{i\downarrow}c_{j\uparrow} + c_{i\uparrow}c_{i\downarrow}c^{\dagger}_{j\uparrow} \end{pmatrix} \tag{8}$$

and the superscript α denotes the averaging over nearest-neighbor sites. Consequently, we can write the dynamical correction matrix $\delta\mathbf{m}$ in the form

$$\delta\mathbf{m}(\mathbf{k},\omega) = Dm(\mathbf{k},\omega)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{9}$$

and the problem reduces to the determination of the higher-order Green function $Dm(\mathbf{k}) = FT\langle\theta(t-t') \times \{\delta J_n(t)\, \delta J^{\dagger}_m(t)\}\rangle_I$. Because $Dm(\mathbf{k},\omega)$ cannot be evaluated exactly, we seek a systematic way of calculating the dynamical corrections. The simplest approach would be to consider the single-site approximation. Such an approximation is in the spirit of the $d = \infty$ (Ref. 42) methods, in which the self-energy is momentum independent. An improvement would be to consider the dynamics associated with two sites as proposed by Mancini and Matsumoto.[39] Evaluation of the self-energy over successively larger clusters would lead to an exact determination of the dynamical corrections. The essence of this approach is based on the fact that the physics of strongly correlated electrons emerges mainly from local correlations: on-site interactions generate the Mott gap, while nearest-neighbor interactions generate the scale, t^2/U. Successively larger clusters build in lower and lower energy scales. However, due to the fact that transfer of spectral

weight from high to low energies is characteristic of Mott insulators, a separation of energy scales in strongly correlated problems is not possible. Therefore, it is crucial that all the excitations be treated on equal footing and, consequently, local correlations must be included. Following Mancini and Matsumoto,[39] we write the dynamical corrections as a series

$$Dm(x,x')=\delta_{x,x'}Dm_0(x,x')+\sum_a \delta_{x+a,x'}Dm_1(x,x')+\cdots \tag{10}$$

in increasing cluster size. Here, x and x' are two representative sites and a indexes all nearest-neighbor sites. In the two-site approximation, the series is truncated at the level of on-site, Dm_0, and nearest-neighbor, Dm_1 contributions. In Fourier space, the dynamical corrections can be written as

$$Dm(\mathbf{k},\omega)\approx Dm_0(\omega)+\alpha(\mathbf{k})Dm_1(\omega). \tag{11}$$

Here Dm_0 and Dm_1 involve Green functions of operators defined on nearest-neighbor sites [see Eq. (7)]. Consequently, these Green functions contain operators defined on up to four sites. Further simplifications can be made if we assume that the dominant contributions arise from terms involving at most two nearest-neighbor sites x and x'. In this case, the superscript α in Eq. (7) corresponds to a particular neighboring site instead of an average over all the nearest-neighbor sites. With this assumption, we obtain

$$Dm_0(\omega)=\frac{1}{2d}\,\mathrm{FT}\langle\theta(t-t')\{\delta J(t),\delta J^{\dagger}(t')\}\rangle_I,$$

$$Dm_1(\omega)=\frac{1}{2d}\,\mathrm{FT}\langle\theta(t-t')\{\delta J(t),\delta J'^{\dagger}(t')\}\rangle_I, \tag{12}$$

where the factor of $1/2d$ arises from the coordination number, $\alpha(\mathbf{k})=(1/d)\Sigma_l\cos(k_l)$, and δJ and $\delta J'$ are centered on two nearest-neighbor sites x and x', respectively. It is straightforward now to express the Green function in terms of these quantities,

$$S(\mathbf{k},\omega)=\frac{1}{S_0^{-1}(\omega)+\tilde{t}\,\alpha(\mathbf{k})V(\omega)}, \tag{13}$$

where the on-site Green function S_0 is given by

$$S_0^{-1}(\omega)=\mathbf{I}^{-1}\begin{pmatrix}(\omega+\mu)I_1+\tilde{t}e-Dm_0(\omega) & -\tilde{t}e+Dm_0(\omega)\\ -\tilde{t}e+Dm_0(\omega) & (\omega+\mu)I_1+\tilde{t}e-Dm_0(\omega)\end{pmatrix}\mathbf{I}^{-1}, \tag{14}$$

and the nearest-neighbor contribution is

$$V(\omega)=\begin{pmatrix}1+[\tilde{p}-\tilde{t}^{-1}Dm_1(\omega)]I_1^{-2} & 1-[\tilde{p}-\tilde{t}^{-1}Dm_1(\omega)]I_1^{-1}I_2^{-1}\\ 1-[\tilde{p}-\tilde{t}^{-1}Dm_1(\omega)]I_1^{-1}I_2^{-1} & 1+[\tilde{p}-\tilde{t}^{-1}Dm_1(\omega)]I_1^{-2}\end{pmatrix}. \tag{15}$$

B. The two-site problem: Level operators and resolvents

At this stage, solving our problem entails a calculation of the functions $Dm_0(\omega)$, $Dm_1(\omega)$ and the parameter $\tilde{p}$. To this end, we express these quantities[39] in terms of correlation functions for the level operators associated with a two-site problem. Let us introduce first the single-site level operators $B(i)$, $F_\sigma(i)$, and $D(i)$ which annihilate empty, singly occupied (with spin σ) and doubly occupied states, respectively. In terms of these operators, the original Hubbard operators can be written as $\xi_\sigma=B^{\dagger}F_\sigma$ and $\eta_\sigma=\sigma F^{\dagger}_{-\sigma}D$. As the system can be at a given time in one of the possible four states, the level operators satisfy the condition

$$Q(i)\equiv B^{\dagger}(i)B(i)+\sum_\sigma F^{\dagger}_\sigma(i)F_\sigma(i)+D^{\dagger}(i)D(i)=1. \tag{16}$$

This restriction can be introduced by adding a Lagrange multiplier term of the form, $\varepsilon_0\Sigma_i Q(i)$, to the original Hamiltonian. The level operators for the two-site states Φ_n are obtained by taking all the combinations of single-site level operators. It is convenient to use the following symmetric or antisymmetric combinations corresponding to eight fermionic-type states j:

$$FB_S^\sigma\equiv\frac{1}{\sqrt{2}}[F_\sigma(x)B(x')+B(x)F_\sigma(x')],$$

$$FD_S^\sigma\equiv\frac{1}{\sqrt{2}}[F_\sigma(x)D(x')+D(x)F_\sigma(x')],$$

$$FB_A^\sigma\equiv\frac{1}{\sqrt{2}}[F_\sigma(x)B(x')-B(x)F_\sigma(x')],$$

$$FD_A^\sigma\equiv\frac{1}{\sqrt{2}}[F_\sigma(x)D(x')-D(x)F_\sigma(x')], \tag{17}$$

and eight bosonic-type states,

$$BB\equiv B(x)B(x'),\quad DD\equiv D(x)D(x'),$$

$$FF^\sigma\equiv F_\sigma(x)F_\sigma(x'),$$

$$FF_S \equiv \frac{1}{\sqrt{2}}[F_\uparrow(x)F_\downarrow(x') + F_\downarrow(x)F_\uparrow(x')],$$

$$FF_A \equiv \frac{1}{\sqrt{2}}[F_\uparrow(x)F_\downarrow(x') - F_\downarrow(x)F_\uparrow(x')],$$

$$DB_S \equiv \frac{1}{\sqrt{2}}[D(x)B(x') + B(x)D(x')],$$

$$DB_A \equiv \frac{1}{\sqrt{2}}[D(x)B(x') - B(x)D(x')], \tag{18}$$

where x and x' are the positions of the two-sites and the spin index σ takes two values, $(+1=\uparrow)$ and $(-1=\downarrow)$. Note that FF^σ and FF_S correspond to the spin-triplet states, while FF_A corresponds to the spin-singlet state.

We are interested in solving a two-site problem for a cluster embedded in a reservoir, constituted by the rest of the system. Formally, the total Hamiltonian can be divided into three parts[39] describing the two-site subsystem, H_0 the reservoir H_R, and their interaction H_{0R},

$$H = H_0 + H_R + H_{0R}. \tag{19}$$

To describe the properties of the two-site system, we introduce the resolvent

$$R_{nm}(t-t') = \theta(t-t')\frac{\mathrm{Tr}_R[\langle 0|\Phi_n(t)\Phi_m(t')^\dagger|0\rangle e^{-\beta H_R}]}{\mathrm{Tr}_R[e^{-\beta H_R}]}, \tag{20}$$

where the trace is over the degrees of freedom of the reservoir, $|0\rangle$ denotes the vacuum for the two-site problem, and, as usual, $\beta = 1/k_B T$. Note that $\Phi_n|0\rangle = 0$ and, consequently, $H_{0R}|0\rangle = 0$. The Fourier transform of the resolvent can be expressed using the spectral function, $\sigma_{nm}(\omega) = -(1/\pi)\mathrm{Im}\, R_{nm}$,

$$R_{nm}(\omega) = \int dx \frac{\sigma_{nm}(\omega)}{\omega - x + i\delta}. \tag{21}$$

We also introduce the auxiliary function $\bar{\sigma}_{nm}(\omega) = e^{-\beta\omega}\sigma_{nm}(\omega)$. Once we know the resolvents, we can express any average of operators of the type $\Phi_{nm} = \Phi_m^\dagger \Phi_n$ as

$$\langle \Phi_{nm}\rangle = \frac{1}{Z}\int d\omega \bar{\sigma}_{nm}(\omega) \tag{22}$$

with $Z = \Sigma_n \int d\omega \bar{\sigma}_{nn}(\omega)$.

A formal solution of the resolvents[39] can be obtained using the equation of motion method. We can write

$$R_{nm}(\omega) = \left(\frac{1}{\omega - E - \Sigma(\omega)}\right)_{nm}, \tag{23}$$

where the energy matrix E is determined by the levels of an isolated two-site system and the self-energy Σ is a measure of the effects of the reservoir. Explicitly, we have

$$E_{nm} = \langle 0|i\frac{\partial}{\partial t}\Phi_n(t)\Phi_m^\dagger(t)|0\rangle_R \tag{24}$$

and

$$\Sigma_{nm}(\omega) = \mathrm{FT}\langle 0|\theta(t-t')\delta J_{\Phi_n}(t)\delta J^\dagger_{\Phi_n}(t')|0\rangle_{RI}, \tag{25}$$

where R indicates that the trace over the reservoir degrees of freedom has been taken, I indicates the irreducible part, and $\delta J_{\Phi_n}(t) = i(\partial/\partial t)\Phi_n(t) - \Sigma_m E_{nm}\Phi_m(t)$. To proceed, we determine the equations of motion for the level operators. It is convenient to write first the equations for the single-site operators,

$$i\frac{\partial}{\partial t}B = \varepsilon_0 B - \tilde{t}\sum_\sigma c_\sigma^{\dagger\alpha} F_\sigma,$$

$$i\frac{\partial}{\partial t}F_\sigma = (\varepsilon_0 - \mu)F_\sigma = \tilde{t}Bc_\sigma^\alpha - \tilde{t}\sigma c_{-\sigma}^{\dagger\alpha} D,$$

$$i\frac{\partial}{\partial t}D = (\varepsilon_0 - 2\mu + U)D - \tilde{t}\sum_\sigma \sigma F_{-\sigma} c_\sigma^\alpha, \tag{26}$$

where the arbitrary reference energy ε_0 will be set to $\varepsilon_0 = -\mu$. In fact, the lattice and the two-site cluster had differing chemical potentials. However, equilibrium between the two-site system and the lattice requires that both have the same chemical potential. The equations of motion for the two-site level operators can be determined directly using the equations of motion, Eq. (26) and Eqs. (17) and (18). Explicitly, these equations are given in Appendix A. From the equations of motion we can extract the energies E_{nm} for the resolvents. Selecting the terms that do not contain c_σ^α-type operators, that is, the terms that do not depend on the degrees of freedom of the reservoir, we obtain

$$E_{FB_S} = 2\varepsilon_0 - \mu - \frac{\tilde{t}}{2d}, \quad E_{FD_S} = 2\varepsilon_0 - 3\mu + U + \frac{\tilde{t}}{2d},$$

$$E_{FB_A} = 2\varepsilon_0 - \mu + \frac{\tilde{t}}{2d}, \quad E_{FD_A} = 2\varepsilon_0 - 3\mu + U - \frac{\tilde{t}}{2d},$$

$$E_{BB} = 2\varepsilon_0, \quad E_{DD} = 2(\varepsilon_0 - 2\mu + U),$$

$$E_{FF_S} = E_{FF^\sigma} = E_{FF_A} = 2(\varepsilon_0 - \mu),$$

$$E_{DB_S} = E_{DB_A} = 2\varepsilon_0 - 2\mu + U, \quad E_{DB_S FF_A} = E_{FF_A DB_S} = \frac{\tilde{t}}{d}. \tag{27}$$

Employing the standard noncrossing approximation[51] and using the time derivatives of the Φ_n operators from Appendix A, we compute the self-energies Σ_{nm}, Eq. (25), of the resolvents. The expressions for the self-energies are given in Appendix B.

It is known[52] that intersite spin fluctuations, which are ignored in the noncrossing approximation, are in fact important and cannot be neglected at energy scales on the order of

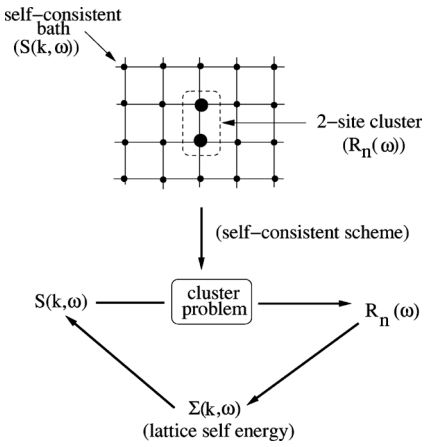

FIG. 2. Self-consistent scheme for the computation of the self-energy.

t^2/U. Since our attempt is to develop a top-down approach in which on-site physics as well as nearest-neighbor correlations are described accurately, we must include spin fluctuations. To overcome this problem, we include the effects of spin fluctuations as a higher-order correction to the self-energies of the resolvents. Physically, we can understand this correction by observing that the solution for singlet and triplet states, FF_A and FF_S, is sharply peaked at energies separated by

$$J=\int_{-\infty}^{\infty}\omega(\sigma_{FF_S}-\sigma_{FF_A})d\omega. \tag{28}$$

In the strong coupling limit, J is of order t^2/U and, consequently, singlet-triplet mixing cannot be ignored. We consider J to be the coupling constant of an effective antiferromagnetic interaction which is responsible for the spin fluctuations. The corrections to the self-energies given by spin fluctuations are given in Appendix C. The energies from Eq. (27), together with the self-energies given in Appendix B and the corrections from Appendix C constitute the equations necessary for the evaluation of the two-site resolvents.

C. Self-consistent procedure

The goal of introducing the two-site resolvents is to express the dynamical corrections Dm_0 and Dm_1, as well as the parameter $\tilde{p}$ in terms of quantities associated with the two-site problem as depicted in Fig. 2. To this end, let us introduce the quantity $Z_{\phi_n}=Z\langle\Phi_n^\dagger\Phi_n\rangle$ by

$$Z_{\Phi_n}=\int d\omega\overline{\sigma}_{\Phi_n}(\omega). \tag{29}$$

The two-site occupation numbers for each of the 16 states can be written directly in terms of Z_{Φ_n}. For example, for the singlet and triplet states, we have $n_{FF_A}=Z_{FF_A}/Z$ and $n_{FF_S}=3Z_{FF_S}/Z$, respectively. Single-site occupation numbers can also be expressed in terms of Z_{Φ_n} using Eq. (16) and writing the identity operator on a neighboring site as $\mathbf{1}'=Q(x')$. For example,

$$D^\dagger D\equiv D^\dagger D\mathbf{1}'=(DB')^\dagger DB'+\sum_\sigma (DF'_\sigma)^\dagger DF'_\sigma + (DD')^\dagger DD', \tag{30}$$

and as a result

$$n_D\equiv\langle D^\dagger D\rangle=Z^{-1}[Z_{FD_S}+Z_{FD_A}+\tfrac{1}{2}Z_{DB_S}+\tfrac{1}{2}Z_{DB_A}+Z_{DD}]. \tag{31}$$

Similarly, we obtain for the average single occupancy, $\Sigma_\sigma\langle F_\sigma^\dagger F_\sigma\rangle$,

$$n_F=Z^{-1}[3Z_{FF_S}+Z_{FF_A}+Z_{FB_S}+Z_{FB_A}+Z_{FD_S}+Z_{FD_A}]. \tag{32}$$

The electron filling is given by the sum of single occupancy and (twice) the double occupancy,

$$n=n_F+2n_D. \tag{33}$$

As all the terms in $\tilde{p}$ can be expressed in terms of products of two-site operators, it is straightforward to write this parameter as

$$\tilde{p}=Z^{-1}[Z_{FD_S}+Z_{FD_A}+Z_{DD}+\tfrac{3}{2}Z_{FF_S}-\tfrac{1}{2}Z_{FF_A}-\tfrac{1}{2}Z_{DB_S}+\tfrac{1}{2}Z_{DB_A}]-\frac{n^2}{4}. \tag{34}$$

This represents the self-consistency condition for the parameter $\tilde{p}$ and obviates the need to employ a decoupling scheme required in the static approximation.[41,53]

The next step is to determine the dynamical corrections. To this end, we express[39] the currents δJ_σ in terms of two-site level operators. From Eq. (7) and $x_i\to x$ and $x_I^\alpha\to x'$, we find that

$$\delta J_\sigma=-\tilde{t}\sum_{m,n}a_{nm}\Phi_n^\dagger\Phi_m,$$

$$\delta J'_\sigma=-\tilde{t}\sum_{m,n}a'_{nm}\Phi_n^\dagger\Phi_m, \tag{35}$$

where Φ_n represents the complete set of two-sites level operators and the coefficients a_{nm} are given in Appendix D. As $\delta J'$ is obtained from δJ by exchanging the positions x and x', the coefficients a'_{nm} will be identical with a_{nm} up to a sign that depends on the symmetry properties of the states Φ_m and Φ_n under the exchange of x and x'. Let us denote by $\mathcal{E}$ the operator that produces the exchange,

$$\mathcal{E}O(x,x')=O(x',x), \tag{36}$$

where $O(x,x')$ is an arbitrary operator defined on the two-site cluster. The symmetry properties of the level operators are given by

$$\mathcal{E}FB_S=FB_S,\quad \mathcal{E}FB_A=-FB_A,$$

$$\mathcal{E}FD_S=FD_S,\quad \mathcal{E}FD_A=-FD_A,$$

$$\mathcal{E}FF_S = -FF_S, \quad \mathcal{E}FF^\sigma = -FF^\sigma,$$

$$\mathcal{E}FF_A = FF_A,$$

$$\mathcal{E}DB_S = DB_S, \quad \mathcal{E}DB_A = -DB_A,$$

$$\mathcal{E}BB = BB, \quad \mathcal{E}DD = DD. \tag{37}$$

Consequently, $a'_{nm} = a_{nm}$ if Φ_n and Φ_m have the same symmetry and $a'_{nm} = -a_{nm}$ otherwise. The self-energy contributions can be calculated only approximately. We will use the noncrossing approximation[51] which has been proven to be effective in solving problems with local correlations. Defining $\Phi_{mn} = \Phi_n^\dagger \Phi_m$, we have

$$\begin{aligned}
&\mathrm{Tr}_R\langle 0|\Phi_{mn}(t)\Phi^\dagger_{m'n'}(t')e^{-\beta H}|0\rangle \\
&= \mathrm{Tr}_R[\langle 0|\Phi_m(t)\Phi^\dagger_{m'}(t')|0\rangle\langle 0|\Phi_{n'}(t')e^{-\beta H}\Phi^\dagger_n(t)|0\rangle] \\
&\approx \mathrm{Tr}_R\langle 0|\Phi_m(t)\Phi^\dagger_{m'}(t')|0\rangle \mathrm{Tr}_R\langle 0|\Phi_{n'}(t')\Phi^\dagger_n(t+i\beta)|0\rangle.
\end{aligned} \tag{38}$$

Consequently, we obtain

$$\begin{aligned}
\langle \Phi_{mn}(t)\Phi^\dagger_{m'n'}(t')\rangle \approx \frac{1}{Z}\int d\omega e^{-i\omega(t-t')} \\
\times \int dx \sigma_{mm'}(\omega + x)\overline{\sigma}_{n'n}(x),
\end{aligned} \tag{39}$$

which is the formulation of the noncrossing approximation that we will use systematically in our calculations. Introducing the expansion Eq. (35) in the expression for the dynamical corrections, Eq. (12), and using the noncrossing approximation formula [Eq. (39)], we obtain

$$\begin{aligned}
Dm_0(\omega) = \frac{1}{2dZ}\int dx dx' \sum_{n,m,n',m'} a_{nm} a^\star_{n'm'} \\
\times \frac{\sigma_{mm'}(x)\overline{\sigma}_{n'n}(x') + \overline{\sigma}_{mm'}(x)\sigma_{n'n}(x')}{\omega - x + x' + i\delta}
\end{aligned} \tag{40}$$

and

$$\begin{aligned}
Dm_1(\omega) = \frac{1}{2dZ}\int dx dx' \sum_{n,m,n',m'} a_{nm} a'^\star_{n'm'} \\
\times \frac{\sigma_{mm'}(x)\overline{\sigma}_{n'n}(x') + \overline{\sigma}_{mm'}(x)\sigma_{n'n}(x')}{\omega - x + x' + i\delta}.
\end{aligned} \tag{41}$$

Detailed expressions for Dm_0 and Dm_1 are given in Appendix D.

This concludes the process of writing the components of the Green function, Eq. (5), which cannot be expressed in terms of the Hubbard basis, in terms of spectral functions for the two-site system. However, for a fully self-consistent calculation, we need to determine the reservoir spectral functions $\rho_\pm(\omega)$ that enter into the formulas for the self-energies of the resolvents (see Appendix B). The reservoir consists of the full system from which the two sites x and x' have been excluded. Our goal is to determine the Green functions $\overline{G} = \langle\langle c^\alpha_\sigma, c^{\dagger\alpha}_\sigma\rangle\rangle$ and $\overline{G}' = \langle\langle c^\alpha_\sigma, c^{\dagger\alpha'}_\sigma\rangle\rangle$. We introduce the two-site full propagator $[S] = \langle\langle \Psi, \Psi^\dagger\rangle\rangle$ with $\Psi^\dagger = [\psi^\dagger(x)\psi^\dagger(x')]$ and the irreducible propagator

$$[\overline{S}] = \begin{pmatrix} \overline{S} & \overline{S}' \\ \overline{S}' & \overline{S} \end{pmatrix}, \tag{42}$$

where $\overline{S} = \langle\langle \psi^\alpha, \psi^{\dagger\alpha}\rangle\rangle$ and $\overline{S}' = \langle\langle \psi^\alpha, \psi^{\dagger\alpha'}\rangle\rangle$. From standard scattering theory, we have that

$$[S] = [S_0] + [S_0][V][\overline{S}][V][S] \tag{43}$$

where

$$[V] = \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix} \tag{44}$$

and

$$[S_0]^{-1} = \begin{pmatrix} S_0^{-1} & tV \\ tV & S_0^{-1} \end{pmatrix}. \tag{45}$$

The solution of these equations is

$$\begin{aligned}
\overline{S} &= V^{-1}[S_0^{-1} - (S - S'S^{-1}S')^{-1}]V^{-1}, \\
\overline{S}' &= V^{-1}[tV + S^{-1}S'(S - S'S^{-1}S')^{-1}]V^{-1}.
\end{aligned} \tag{46}$$

The irreducible Green functions used in the evaluation of the self-energies of the resolvents will be given by

$$\begin{aligned}
\overline{G} &= \overline{S}_{11} + \overline{S}_{12} + \overline{S}_{21} + \overline{S}_{22}, \\
\overline{G}' &= \overline{S}'_{11} + \overline{S}'_{12} + \overline{S}'_{21} + \overline{S}'_{22},
\end{aligned} \tag{47}$$

and the corresponding spectral functions are

$$\begin{aligned}
\rho_+(\omega) &= -\frac{2}{\pi}\mathrm{Im}[\overline{G}(\omega) + \overline{G}'(\omega)], \\
\rho_-(\omega) &= -\frac{2}{\pi}\mathrm{Im}[\overline{G}(\omega) - \overline{G}'(\omega)].
\end{aligned} \tag{48}$$

III. RESULTS

We now have all the ingredients necessary for the implementation of the self-consistent procedure shown in Fig. 2. Starting with an initial guess for the spectral functions $\rho_\pm$ describing the properties of the environment, we solve the two-site problem iteratively using the expression Eq. (23) for the resolvents, the energies Eq. (27) and the self-energies from Appendix B with the spin fluctuation corrections from Appendix C. While in principle the cluster and the lattice can have different chemical potentials as in the previous work on the two-site cluster,[39] we have used the more physical re-

straint that both the cluster and the lattice must have the same chemical potential. However, as a result of Eq. (33), the filling in the cluster and the lattice will be different. By symmetry, however, at half-filling both the cluster and the lattice will have the same filling. All the frequency-dependent functions are discretized on a grid of $N=8192$ points from $\omega_{min}=-20t$ to $\omega_{max}=20t$. To increase the computational speed, we performed all the convolutions involved in the calculation of the self-energies using a fast Fourier transform algorithm. The procedure converges for temperatures above $T=0.02t$ at finite doping and $T=0.08t$ at half-filling, although convergence problems occurred below $T=0.1t$. Once we computed the resolvents, the mean-field parameter $\tilde{p}$ can be determined using Eq. (34), as well as the dynamical corrections Dm_0 and Dm_1 from Appendix D. The mean-field parameter $e=\langle \xi^{\alpha}\xi^{\dagger}\rangle-\langle \eta^{\alpha}\eta^{\dagger}\rangle$ can be expressed in terms of the Green function using the general self-consistency condition,

$$\langle \psi_m(i)\psi_n^{\dagger}(j)\rangle=\frac{\Omega}{(2\pi)^2}\int d^2k\,d\omega e^{i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)}[1-f(\omega)] \times\left(\frac{-1}{\pi}\right)\mathrm{Im}\,S_{mn}(\mathbf{k},\omega). \tag{49}$$

Within the grand-canonical ensemble, the chemical potential μ is determined by the self-consistent solution to

$$n=2(\langle \xi\xi^{\dagger}+2\xi\eta^{\dagger}+\eta\eta^{\dagger}\rangle). \tag{50}$$

We imposed the constraint that the chemical from Eq. (50) also equal that for the cluster. We then determined the full Green function $S(\mathbf{k},\omega)$ using Eqs. (13)–(15). Next, new spectral functions $\rho_{\pm}$ are determined using Eq. (48) and the whole procedure [see Fig. 2] is repeated until full convergence is reached.

A. Spectral function at half-filling: Mott insulator

Before we analyze the doped case, we first review[49] the properties of the charge vacuum that determines the insulating behavior at half-filling. To reiterate, there are two distinct routes to the insulating state at half-filling. For $U\gg t$, the charge and spin degrees of freedom are decoupled and the system is an insulator for temperatures smaller than $T_0\sim U$. The spins are coupled due to the super-exchange interaction, $|J|\approx 4t^2/U$. It is this spin exchange interaction that gives rise to local antiferromagnetic fluctuations and eventually ordering at $T=0$. This is the antiferromagnetic Mott insulating state. In general, in the weak coupling regime, a metal-insulator transition occurs as a consequence of the Brillouin zone folding generated by magnetic or charge ordering and the corresponding gap is essentially related to antiferromagnetism or some type of charge density wave. This type of transition is referred to as a Slater transition and the corresponding insulating state should not be confused with the Mott insulator. Such a regime can be successfully described by conventional many-body approaches.

Shown in Fig. 3 is the total electron spectral function,

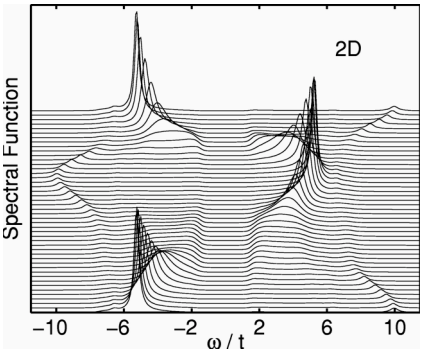

FIG. 3. Momentum and energy dependence of the electron spectral function for a half-filled 2D system with $U=8t$ and $T=0.15t$. From top to bottom, the curves correspond to $(k_x,k_y)=(0,0)\rightarrow(\pi,\pi)\rightarrow(\pi,0)\rightarrow(0,0)$.

$$A(\mathbf{k},\omega)=-(1/\pi)\mathrm{Im}\{S_{11}(\mathbf{k},\omega)+2S_{12}(\mathbf{k},\omega)+S_{22}(\mathbf{k},\omega)\} \tag{51}$$

for the 2D half-filled Hubbard model with $U=8t$ and $T=0.15t$. Clearly visible are the upper and lower Hubbard bands with an energy gap of order U and flatness of the band near the $(\pi,0)$ point. The chemical potential $(\omega=0)$ lies in the middle of the gap and hence the system is an insulator. However, no symmetry is broken as is evident because the periodicity is 2π rather than π as would be the case if the Brillouin zone had doubled. In addition, spin and charge are not fractionalized.[26] The insulating behavior arises because the charge gap has splintered the spectral weight of each $\mathbf{k}$ state into ''bonding'' and ''antibonding'' pieces. Consequently, there is a fundamental breakdown of what is meant by an electronic state. In fact, the electronic states themselves have fractionalized. To make contact with the real-space picture shown in Fig. 1, we note that the PES and IPES spectra are determined by the spectral weight in the lower and upper Hubbard bands, respectively. In general, the upper and lower Hubbard bands carry total spectral weight $n/2$ and $1-n/2$, respectively, which of course reduces to 1/2 at half-filling. However, $A(\mathbf{k},\omega)$ is strongly momentum dependent as illustrated in Fig. 3. For the lower Hubbard band, the maximum in the spectral weight is peaked at (0,0) and decreases as the $(\pi,0)$ point is reached and becomes vanishingly small at (π,π). In fact, the states at (π,π) in the lower Hubbard band carry almost none of the spectral weight. However, the decrease in the occupancy of each $\mathbf{k}$ state is a continuous function, as depicted in Fig. 4, rather than a discontinuous one as would be the case in a Fermi liquid with a well-defined Fermi surface. In the strict sense, the discontinuity in $n_{\mathbf{k}}$ in a Fermi liquid occurs at $T=0$. Although we cannot reach $T=0$ in our approach, we find no indication that a discontinuity develops in $n_{\mathbf{k}}$ as the temperature decreases. In fact, the continuous behavior we have obtained here is consistent with the exact result[38]

$$n_{\mathbf{k}}=\frac{1}{2}+2\frac{\epsilon_{\mathbf{k}}}{U}\left\langle \mathbf{S}_i\cdot\mathbf{S}_{i+\boldsymbol{\delta}}-\frac{1}{4}\right\rangle \tag{52}$$

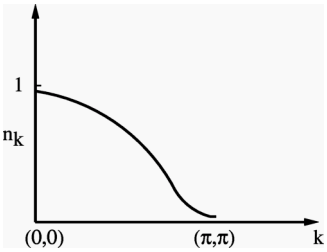

FIG. 4. Occupancy of each momentum state with $k_x=k_y$ for the spectral function shown in Fig. 3. The occupancy is peaked at (0,0) and decreases continuously to a minimum value at (π,π) without the discontinuity indicative of a Fermi liquid.

for the occupancy in each **k** state projected into the LHB of a half-filled Hubbard model. Note all the **k** dependence is determined by the single-particle energies,

$$\epsilon_{\mathbf{k}}=-t\sum_{\delta} e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_{i+\delta})}. \tag{53}$$

For a paramagnetic state, the term in the angle brackets is exactly, $-1/4$. Hence, $n_{\mathbf{k}}$ is a continuously decreasing function from (0,0) to (π,π) as found here for $U=8t$.

Although $n_{\mathbf{k}}<1$, each state satisfies the sum rule,

$$\int_{-\infty}^{\infty} A(\mathbf{k},\omega)d\omega=1. \tag{54}$$

In a band insulator the upper cutoff on the energy is simply the chemical potential. This does not mean that in a band insulator or in a Fermi liquid, a broad incoherent background cannot be present which extends to high energies. In fact, in a Fermi liquid, $n_{\mathbf{k}}$ can be less than unity. However, in Fermi liquids, a coherent quasiparticle peak always exists regardless of the momentum. Hence, the criterion for occupancy of a single-particle state is simply whether or not the coherent peak lies above or below the chemical potential. In a Mott insulator not only is the spectral weight split over an energy scale of U but there is an absence of coherent quasiparticles as evidenced by the broad spectral features. Hence, there is no sharp criterion for unit occupancy of a single-particle state. The broadness of the spectral features stems from the local correlations on neighboring sites not the Mott gap itself. Without the dynamical corrections, the spectral function would simply be a sum of δ-function peaks at the lower and upper Hubbard bands.

The bifurcation of the spectral weight of each **k** state above and below the charge gap can be modeled as follows. Consider, for example, $\gamma^{\dagger}_{\mathbf{k}\sigma}=u_{\mathbf{k}}\xi^{\dagger}_{\mathbf{k}\sigma}+v_{\mathbf{k}}\eta^{\dagger}_{\mathbf{k}\sigma}$, with coefficients $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ are determined by the projection of the spectral function onto the lower and upper Hubbard bands, respectively. Hence, the antisymmetrized state formed from such excitations

$$|\mathrm{MI}\rangle=\sum_{P}(-1)^{P}\prod_{\mathbf{k}\in \mathrm{FBZ}}\gamma^{\dagger}_{\mathbf{k}\uparrow}\gamma^{\dagger}_{\mathbf{k}\downarrow}|0\rangle \tag{55}$$

is a candidate for describing the elusive paramagnetic Mott insulator. Provided magnetic frustration is present so that ordering is preempted, Eq. (55) should be the $T=0$ Mott insulating state. The sum over all permutations P is necessary as the $\gamma_{\mathbf{k}\sigma}$ operators obey the nonfermionic commutation relations,

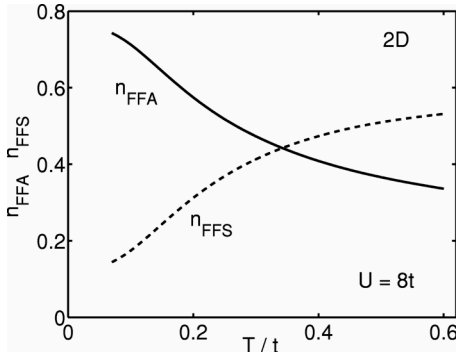

FIG. 5. Singlet (n_{FF_A}) and triplet (n_{FF_S}) occupation numbers as a function of temperature for $U=8t$ in two dimensions. The fact that $n_{FF_A}>n_{FF_S}$ as $T\to 0$ is consistent with local antiferromagnetic order.

$$\{\gamma^{\dagger}_{\mathbf{k}\uparrow},\gamma^{\dagger}_{\mathbf{q}\downarrow}\}=\sum_{i} e^{-i(k+q)r_i}(u_k v_q-u_q v_k)c^{\dagger}_{i\uparrow}c^{\dagger}_{i\downarrow} \tag{56}$$

and $\{\gamma^{\dagger}_{\mathbf{k}\sigma},\gamma^{\dagger}_{\mathbf{q}\sigma}\}=\{\gamma^{\dagger}_{\mathbf{k}\sigma},\gamma_{\mathbf{q}\sigma'}\}=0$. When $u_k=v_k$, $\gamma^{\dagger}_{\mathbf{k}\uparrow}\gamma^{\dagger}_{\mathbf{k}\downarrow}|0\rangle$ generates the completely doubly occupied state. However, because the right-hand side of the anticommutation relation in Eq. (56) is identically zero for $u_k=v_k$, the zero state results upon summation over all permutations. Consequently, Eq. (55) completely projects out the fully doubly occupied state.

1. Singlet formation: local antiferromagnetism

A crucial test of the correctness of the method we have used here is whether or not short-range antiferromagnetic correlations are present at low temperatures. Such correlations do not signal that long-range magnetic order obtains at $T=0$ but rather that the ground state at $T=0$ is a liquid of nearest-neighbor singlet states as in the resonating valence-bond (RVB) state proposed by Anderson.[21] We are able with our two-site formalism to probe the existence of local magnetic order by computing the nearest-neighbor singlet and triplet occupation numbers $n_{FF_A}=Z_{FF_A}/Z$ and $n_{FF_S}=3Z_{FF_S}/Z$, respectively. From Fig. 5 we find that, at high temperatures, triplet excitations dominate. However, this trend is reversed below some temperature and the singlet occupancy becomes of order unity. Hence, the low-temperature properties of the insulating state we have computed here are consistent with a liquid of nearest-neighbor singlet states as in the RVB state. In fact, at $T=0$ the liquid state we have found here persists because we have imposed a paramagnetic solution.

A final diagnostic of the insulating state we have found here is the behavior of the effective exchange interaction as a function of U. In the Mott state, a super-exchange interaction is self-generated which should scale as $1/U$. It is this exchange interaction that sets the scale for the Neél tempera-

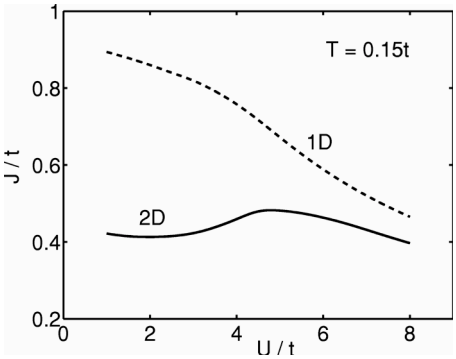

FIG. 6. Effective exchange interaction coupling constant J as a function of U/t for $T=0.15t$.

ture. Using Eq. (28), we computed the effective exchange interaction shown in Fig. 6 for both 1D and 2D. Note first that J is always positive as a consequence of the fact that the singlet state is lower in energy than the triplet. This is a further indication of antiferromagnetic correlations in the ground state. As expected, J is well approximated by $4t^2/U$ in the strong-coupling regime. However, as U decreases, deviations from this behavior are observed.

B. Doped Mott insulators

1. Chemical potential

Two scenarios are possible for the doping dependence of the chemical potential: (1) the chemical potential remains pinned and mid-gap states are generated by some physical mechanism, or (2) the chemical potential jumps to the top of the LHB or the bottom of the UHB upon hole or electron doping, respectively. Our results shown in Fig. 7 demonstrate that the chemical potential jumps upon hole or electron doping, indicating an absence of mid-gap states. The magnitude of the jump is set by the Mott gap which is fully developed at $T=0$. While at some finite temperature, the chemical potential may appear to evolve smoothly, at $T=0$, the chemical potential should jump discontinuously. That is, we find that doping a Mott insulator leads to a continuous depletion of the spectral weight in the first Brillouin zone. Our finding is consistent with the exact result for the 1D Hubbard model[54] as well as quantum Monte-Carlo simulations[43,44] in two dimensions. However, in $d=\infty$, the chemical potential exhibits[55] a jump but one that is smaller than the gap. Hence, mid-gap states are generated.[55] It would appear then that $d=\infty$ is vastly different from an actual finite-dimensional system and the 2D Hubbard model is quite similar to its 1D counterpart at least as far as the chemical potential is concerned. The possible source of the discrepancy is the form of the self-energy. A chemical potential jump requires a large imaginary part of the self-energy at the chemical potential, thereby indicating an absence of well-defined quasiparticles. Mid-gap states are resonance states and hence are reminiscent of the Brinkman/Rice[56] mechanism for the insulator-metal transition in the doped Mott state. The inset in Fig. 7 indicates that $\mathrm{Im}\,\Sigma$ in the underdoped regime is large and nonzero at the Fermi energy. Such behavior points to an absence of well-defined quasiparticles. In the overdoped regime, the characteristic ω^2 dependence appears, indicative of a Fermi liquid. Consequently, the method we use here is capable of recovering Fermi liquid theory in the overdoped regime.

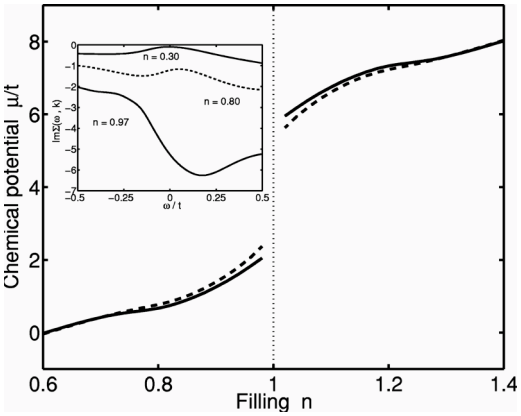

FIG. 7. Doping dependence of the chemical potential in the 2D Hubbard model computed using the local cluster approach for $T=0.15t$ (dashed line) and $T=0.07t$ (solid line). The inset shows the imaginary part of the self-energy evaluated at a Fermi momentum (0.3,2.10) for $n=0.97$, (0.3,1.84) for $n=0.8$ and (0.3,1.06) for $n=0.3$.

Experimentally, whether the chemical potential is pinned or moves upon doping appears to be cuprate dependent. For example, in $La_{2-x}Sr_xCuO_4$ (Ref. 57) (LSCO), the chemical potential remains pinned roughly at 0.4 eV above the top of the LHB, while for $Nd_{2-x}Ce_xCuO_4$ (NDCO),[58] $Bi_2Sr_2Ca_{1-x}R_xCu_2O_{8+y}$ (BSCO),[59–62] and Na-doped $Ca_2CuO_2Cl_2$ (Ref. 33) (CACLO) the chemical potential jumps upon doping by an amount in accordance with half the Mott gap and scales roughly as δ^2 as obtained here. Because stripes or macroscopic phase separation require the chemical potential to be pinned, they have been invoked[63] to explain the origin of mid-gap states in LSCO. The pseudogap in the underdoped cuprates has also been attributed[63] to stripes. However, because $\Delta\mu\neq 0$ for most of the cuprates, for example, NDCO, BSCO, and CACLO, if the pseudogap has a universal origin, stripes are not its cause. The precise origin of the pseudogap will be discussed extensively in a later section.

2. Spectral function

Shown in Fig. 8 is the electron spectral function at high temperature, $T=0.25t$ for $n=0.97$, $n=0.90$, $n=0.80$, and $n=0.60$. Several features are evident: (1) the chemical potential moves further into the LHB as the filling decreases, (2) no coherent peaks exist near the chemical potential in the lightly doped regime, $0.9<n<1$, (3) in the dense or weakly interacting regime, sharper features appear, (4) each state in the FBZ has spectral weight both above and below the chemical potential as dictated by Mottness, (5) the Mott gap remains intact but moves to higher energy as the doping increases, and (6) at (π,π), the UHB carries most of the spectral weight regardless of the filling. In the underdoped regime, the characteristic width of each $\mathbf{k}$ state is of order t and even much larger near $(\pi,0)$. Such broad spectral features in the underdoped regime are seen experimentally[60] and

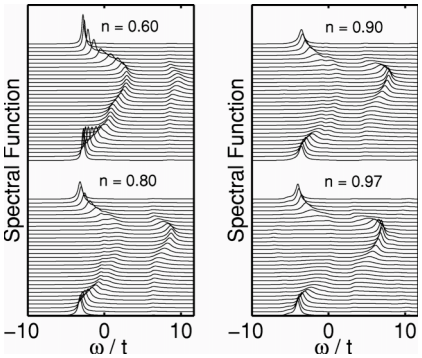

FIG. 8. Doping dependence of the spectral function in the 2D Hubbard model computed using the local cluster approach for $T=0.25t$ and fillings of $n=0.97$, $n=0.90$, 0.80, and $n=0.60$.

arise in this context because $\mathrm{Im}\,\Sigma(\epsilon_F)\neq 0$ as shown in Fig. 7. As a consequence, there is no sharp criterion for unit occupancy of each state in the FBZ. Because the total spectral weight of each $\mathbf{k}$ state is unity, however, and each state lives both below and above the chemical potential, the charge carried by the piece of the state lying below the chemical potential is less than unity. That is, each electronic state is fractionalized. In the heavily overdoped regime, the splitting of the spectral weight above and below the Mott gap, is highly suppressed. As illustrated in Fig. 8, most of the spectral weight resides in the LHB for a filling of $n=0.60$. As a consequence, Mottness vanishes in the overdoped regime. In addition, in the heavily overdoped regime, $\mathrm{Im}\,\Sigma$ acquires the characteristic ω^2 dependence indicative of a Fermi liquid. Hence, with our method we are able to recover the key characteristics to the transition to the traditional Fermi liquid state, namely, (1) a vanishing of spectral weight in the UHB and (2) $\mathrm{Im}\,\Sigma\approx\omega^2$ near the chemical potential.

Does new physics emerge at low temperatures? Figure 9 depicts the spectral function computed at $T=0.07t$. At this relatively low temperature, two new features emerge. First, in the underdoped regime, the spectral weight appears to be suppressed at the chemical potential. Whether this gives rise to a pseudogap will be investigated in the following section. Second, at $n\approx 0.8$ especially in the vicinity of the $(\pi,0)$ point, the band becomes almost dispersionless and seems to split into two subbands. Instead of a strong coherence peak,

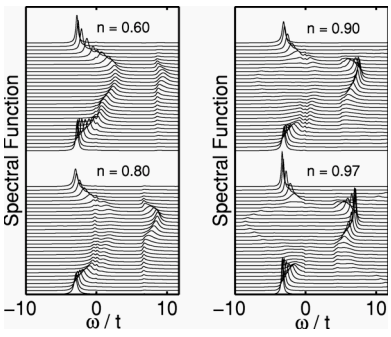

FIG. 9. Doping dependence of the spectral function in the 2D Hubbard model computed using the local cluster approach for $T=0.07t$ and fillings of $n=0.97$, $n=0.90$, 0.80, and $n=0.60$.

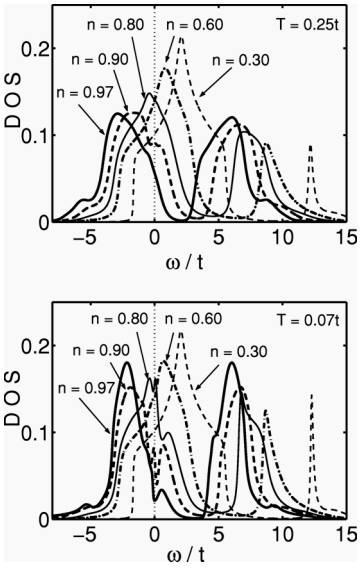

FIG. 10. Density of single-particle states for $T=0.25t$ and $T=0.07t$, $U=8t$ for the fillings shown. No pseudogap exists at high temperature. At low T and low doping levels, a pseudogap emerges at the chemical potential but moves above it at an intermediate doping level. In the overdoped regime, the pseudogap vanishes entirely and a weakly interacting system is recovered.

at the chemical potential we observe a weak maximum adjacent to a region with depleted spectral weight which forms a ''channel'' immediately above the chemical potential. To understand the importance of the low-temperature features, it is expedient to compute the density of states.

3. Pseudogap without preformed pairs or global symmetry breaking

To investigate the possibility of a pseudogap[64] in the lightly doped regime, we integrate the spectral function over momentum to obtain the single-particle density of states (DOS) at high and low temperatures. Displayed in Fig. 10 is the DOS for $T=0.25t$ and $T=0.07t$ for several fillings. As is evident, no local minimum of DOS exists at the chemical potential at high temperature, $T=0.25t$. Features which emerge even at high temperature are the reshuffling of spectral weight from above the charge gap to below as the filling is changed and also a movement of the Mott gap to higher energies. Note that even at $n=0.30$ the Mott gap is still present, though almost all of the spectral weight now resides in the LHB which closely resembles the noninteracting density of states. This is further evidence that we correctly recover Fermi liquid theory as $n\to 0$. What about low temperature? The lower panel of Fig. 10 demonstrates that a pseudogap forms in the DOS for $\delta\approx 0$. The vertical line at 0 indicates that the pseudogap occurs precisely at the chemical potential. Similar qualitative results based on a cluster method have been obtained by Maier *et al.*,[65] except their pseudogap is slightly displaced above E_F. In contrast, in the analysis of Haule *et al.*,[66] the DOS has a negative slope through E_F (as dictated by the proximity to the Mott gap) but never acquires a local minimum at E_F indicative of a true pseudogap. Because the pseudogap exists below some char-

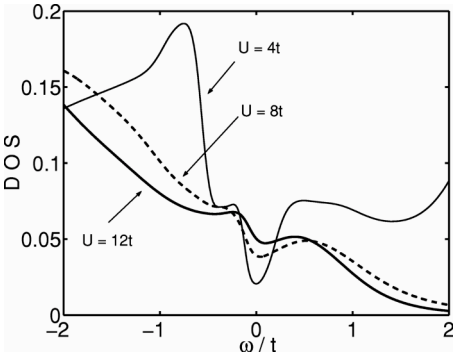

FIG. 11. Density of single-particle states for $T=0.1t$, for $n=0.95$ and three values of the on-site interaction: $U=12t$, $U=8t$, and $U=4t$. The gradual vanishing of the pseudogap as U increases offers direct confirmation that the energy scale for the pseudogap is set by t^2/U.

acteristic temperature and vanishes at higher doping, the result obtained here is nontrivial and highly reminiscent of experimental observations in the cuprates.[67] What is its origin? Incoherence ($\mathrm{Im}\,\Sigma \neq 0$ at E_F) is central, though it is not a sufficient condition[66] [see Fig. 2] for a pseudogap. From Fig. 10, the pseudogap remains intact up to $\delta=0.20$ but simply moves to higher energies as does the Mott gap. This is telling because in $d=\infty$, no[55] pseudogap exists but a Mott gap is present. Absent from $d=\infty$ but present in any lattice of finite connectivity are true short-range correlations. We argue then that the pseudogap is the nearest-neighbor analog of the on-site generated Mott gap. The energy scale for nearest-neighbor interactions scales as t^2/U. Hence, if our hypothesis is correct, we expect the pseudogap to diminish as U increases. The evolution shown in Fig. 11 indeed demonstrates that, at finite temperature, the pseudogap does vanish as U increases. Hence, we can assert with certainty that correlations on neighboring sites do in fact create a depletion in the density of states. The energy scale t^2/U is typically associated with antiferromagnetic spin fluctuations. To explore whether such processes have the right doping dependence to explain the origin of the pseudogap, we display in Fig. 13 the x dependence of J computed from Eq. (28). As is evident, J is only weakly doping dependent in the underdoped regime and hence lacks the strong doping dependence needed to explain the pseudogap. This trend is consistent with that of Jarrell and co-workers[65] who have observed that the pseudogap persists even if antiferromagnetism is killed. Figure 13 also indicates that J_{eff} computed as the energy difference between the nearest-neighbor singlet and triplet states vanishes at $x=0.8$. This is not an accident. Nearest-neighbor spin fluctuaions should desist when no nearest-neighbor sites remain singly occupied. On average, this obtains at a filling of 1/5 or $x=0.8$, precisely the doping value found here. Hence, it is not a surprise that the doping dependence of J is weak in the underdoped regime.

What then is the cause of the physics of the pseudogap? Any two-step process involving the UHB scales as t^2/U. Consider the explicit three-site terms that appear away from half-filling

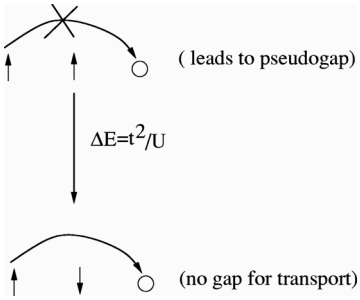

FIG. 12. Local three-site configurations in which spin-blocking leads to a pseudogap. In the upper state, transport directly between the up spin two sites away from the hole and the hole is not possible. Only hole transport via a two-step process is possible. As the amplitude for hole transport is a superposition of all such processes, a pseudogap develops. The only way to overcome the spin-blockade is for the spin neighboring the hole to flip. This process costs an energy t^2/U. Once the spin is flipped, there is no barrier for transport.

$$H_{3\mathrm{B}}=-\frac{J}{4}\sum_{i\delta\neq\delta',\sigma}(\tilde{c}^{\dagger}_{i+\delta,\sigma}\tilde{\eta}_{i,i+\delta'\sigma,-\sigma}-\tilde{c}^{\dagger}_{i+\delta,-\sigma}\tilde{\pi}_{i,i+\delta',\sigma}) \tag{57}$$

in an expansion in t/U, where the tilde represents full projection onto the LHB. It is precisely these terms that thwart the equivalence between the so-called t-J and Hubbard models in the large U limit away from half-filling.[38] This term represents the motion (strictly in the LHB) of a hole in a spin background. We argue that such terms are involved in the pseudogap. The mechanism is as follows. Consider placing a single hole in a Mott insulator. Unlike a site neighboring the hole, a singly occupied site two lattice sites away must temporarily doubly occupy one of its neighbors if it is to move to the hole. For this to be possible, the electrons on neighboring sites must have opposite spins. The matrix element for such a two-step process is t^2/U and described by the three-site terms written above. For sites with the incorrect spin alignment, a local spin fluctuation must obtain for the three-site hopping to occur. The energy barrier for this process is t^2/U. It is from those local three-site configurations in which the spins are incorrectly aligned that the pseudogap arises as illustrated in Fig. 12. Simply invoking spin fluctuations is insufficient to explain the origin of the pseudogap as spin fluctuations alone cannot give rise to transport. However, spin fluctuations can make it possible for an electron two sites away from a hole to transport. Hence, spin fluctuations in the context of three-site hopping can overcome the local spin blockade (or spin gap) that exists in doped Mott insulators. As this effect is entirely local, the pseudogap is the nearest-neighbor analog of the Mott gap: neighboring sites with a parallel arrangement of the spins experience an energy barrier equal to t^2/U for charge transport. Can the doping value at which this process vanishes be estimated? On this account, the pseudogap should be related to the joint probability that a neighboring three-site configuration consisting of a hole and two sites with spin parallel electrons exists. The minimum constraint however is simply that each site has on average one hole as its immediate neighbor,

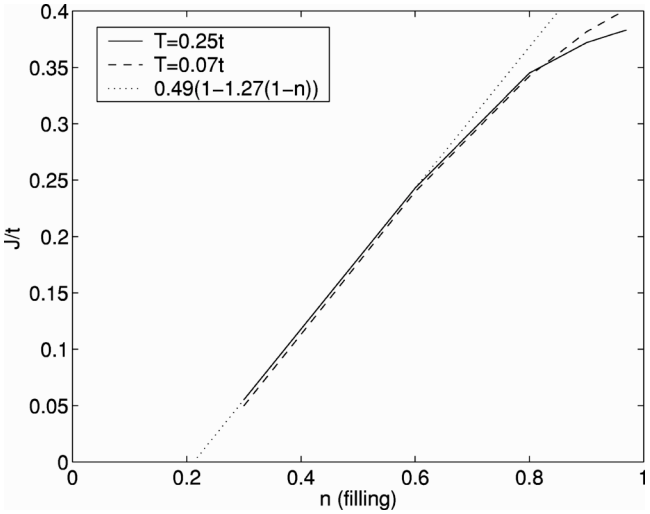

FIG. 13. Effective nearest-neighbor J as a function of filling. J vanishes when four out of five sites are empty. Beyond this concentration, nearest-neighbor spin fluctuations are not possible.

roughly $x=0.25$ for a square lattice. Hence, the pseudogap is of the form $t^2/UP(x)$, where $P(x)$ determines the probability that hole transport involves double occupancy and consequently, is a steadily decreasing function of x vanishing at x_{crit}. From the estimate given above, it is likely that x_{crit} is closer to the end of the superconducting dome than it is to optimal doping.

It is common[38,68] to approximate the three-site terms considered here as xKJ, where K is treated as the lattice connectivity. Consequently, the effective nearest-neighbor exchange interaction is doping dependent and given by $J_{\mathrm{eff}}=J(1-xK)$ which vanishes at $x_{\mathrm{crit}}=1/K\approx 0.25$ (Ref. 68) for a square lattice. However, as Fig. 13 demonstrates, J_{eff} vanishes at $x=0.8$ not $x=0.25$. It is likely that the discrepancy found here arises from the fact that the three-body terms in Eq. (57) cannot, in any real sense, be rewritten as an effective spin exchange. In fact, at $x=0.8$, every site has on average four neighboring holes. At this concentration, nearest-neighbor spin fluctuations are not possible; hence, $J_{\mathbf{e}ff}$ should vanish, as seen in Fig. 13.

4. Heat capacity

Evidence for the pseudogap is also found from the heat capacity.[69] We computed the heat capacity numerically from the internal energy

$$C(T)=\frac{1}{N}\frac{dE}{dT}. \tag{58}$$

The energy per site is the sum of the kinetic term and the interaction term

$$\frac{E}{N}=-2t\langle c_i^{\dagger}c_i\alpha\rangle+U\langle n_{i\uparrow}n_{i\downarrow}\rangle, \tag{59}$$

where $t=2dt_0$ ($t=4t_0$ in two dimensions). The double occupancy can be expressed as

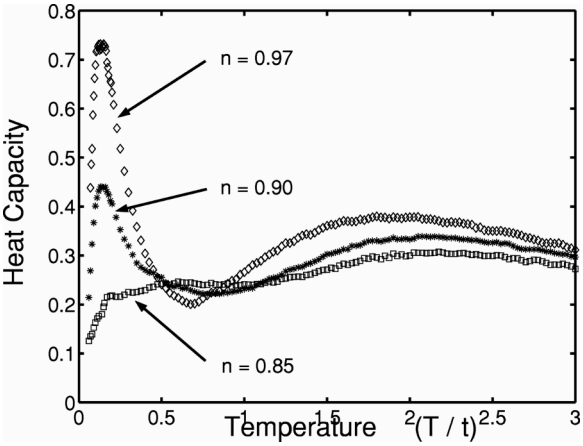

FIG. 14. Heat capacity computed for three fillings computed by numerically differentiating the internal energy obtained from the average of Eqs. (61) and (62). For all three fillings shown, $U=8t$.

$$\langle n_{i\sigma}\bar{n}_{i\sigma}\rangle=\langle c_{i\sigma}^{\dagger}\eta_{i\sigma}\rangle=\frac{n}{2}-\langle\eta_{i\sigma}c_{i\sigma}^{\dagger}\rangle. \tag{60}$$

The correlations can be expressed in terms of the Green function as

$$\frac{E}{N}=\frac{n}{2}U+\int\frac{d\omega}{2\pi}\int\frac{d^2k}{(2\pi)^2}[1-f(\omega)]\Bigg[2t\alpha(k)\frac{-1}{\pi}\times\mathrm{Im}(S_{11}+2S_{12}+S_{22})-U\frac{-1}{\pi}\mathrm{Im}(S_{12}+S_{22})\Bigg]. \tag{61}$$

Equivalently, we can use the equation of motion for the Green function $\langle\langle\eta_i(\tau),c_j^{\dagger}(\tau')\rangle\rangle$ and we obtain an alternate expression

$$\frac{E}{N}=\frac{n}{2}U-\left(1-\frac{n}{2}\right)\mu-\int\frac{d\omega}{2\pi}\int\frac{d^2k}{(2\pi)^2}[1-f(\omega)]\times[\omega-t\alpha(k)]\frac{-1}{\pi}\mathrm{Im}(S_{11}+2S_{12}+S_{22}) \tag{62}$$

for the energy per particle. We found that the difference between Eqs. (61) and (62) was within our numerical errors. Consequently, in our final calculations of the heat capacity shown in Fig. 14, we averaged the two results. In the 1D and 2D Hubbard models at half-filling,[49] two peaks exist in the heat capacity. The high temperature peak corresponds to charge excitations and the low-tempeature peak to spin physics. As is evident, the same separation of energy scales persists even in the doped case. However, for $n<0.9$, we find that the spin peak vanishes and merges with the charge excitation spectrum. This dramatic change represents a possible termination of Mott-dominated physics and the onset of more Fermi liquid behavior. Bonca and Prelovsek[70] observed the identical trend in their exact diagonalization study of a 4×4 system. This agreement lends further credence to our method. Another trend evident from Fig. 14 is that the extrapolated $T=0$ value of the heat capacity in the underdoped

regime, $n>0.9$ is lower than that at $n=0.85$. This represents a loss of spectral weight at low energies as would be the case once a pseudogap opens. Hence, the thermodynamics also corroborate the existence of a pseudogap.

5. Possible time-reversal symmetry breaking

We have shown that in general, a pseudogap exists in a doped Mott insulator without invoking symmetry breaking of any kind. Nonetheless, we entertain the possibility that in analogy with antiferromagnetism and the Mott gap, perhaps some broken symmetry state obtains at lower temperature as a result of the pseudogap found here. Our argument should be construed as a conjecture and hence is entirely speculative. It by no means underlies the calculations we presented here. Loram and colleagues[69] have argued that for $x<x_{\rm crit}$, a glassy phase with an Edwards-Anderson order parameter obtains. While the experimental evidence for a glassy phase extending to $x_{\rm crit}$ is not clear, recent circular dichroism experiments[71] point to time-reversal as the relevant symmetry that is broken in the pseudogap phase. However, this symmetry is broken only along the $(\pi,0)$ and $(0,\pi)$ directions and not along (π,π). Should these results endure, they will provide a benchmark for measuring the validity of the numerous proposals for the pseudogap.[72–76] In the context of the view put forth here, we must determine how purely nearest-neighbor correlations can give rise to a breaking of time-reversal symmetry only along the canonical x and y axes but not along $x=y$. Consider the three-body term in Eq. (57). This term generates correlated motion of a hole among nearest-neighbor sites, that is, local currents. In analogy with the local moments that order antiferromagnetically at $T=0$ as a result of the Mott gap, we propose that the currents may order in the pseudogap phase below T^*. Experimentally,[71] translational symmetry is preserved in the pseudogap phase. Hence, staggered orbital currents are not possible as they automaticaly result in a doubling of the unit cell within a single-band model.[71] Further, experimentally,[77–79] there is no evidence that physics beyond a single-band is relevant to the cuprates. Consequently, any current pattern must preserve translational symmetry within a single-band model. Only one option remains: the currents order below some characteristic temperature T^* along the canonical x and y axes. To ensure that the net current along $x=y$ vanishes, a compensating diagonal current must be present as depicted in Fig. 15. This current pattern can be obtained from the most recently proposed pattern of Simon and Varma[80] by simply integrating out the oxygen sites. Hence, despite claims to the contrary,[80] it is entirely possible to generate a translationally invariant current pattern within a one-band model that is consistent with the experimental observations. In the corrected pattern of Varma and Simon,[80] the oxygen sites do nothing except produce a diagonal current which ensures that the total current in each plaquette vanishes. That the oxygen sites can be integrated out is certainly consistent with the now well accepted work of Zhang and Rice.[81] Nonetheless, our work does not hinge on the current pattern shown in Fig. 15 being the origin of the pseudogap. However, insofar as such a pattern obtains entirely from local nearest-neighbor physics, it is consistent with our finding that the pseudogap is the nearest-neighbor analog of the Mott gap. Should further experiments confirm the presence of t violation, then a more microscopic investigation of the origin of the current pattern shown in Fig. 15 will be warranted. Of course, a current pattern of the type proposed here can only be obtained (if at all) from a Hubbard model if next-nearest-neighbor hopping is included.

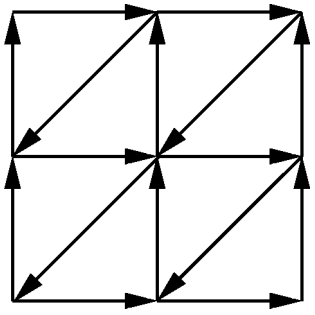

FIG. 15. Current pattern for the motion a hole in a doped Mott insulator that preserves t-reversal symmetry along $x=\pm y$ but violates it along the canonical x and y axes. Each lattice represents a copper site. Hence, translational symmetry is preserved. The diagonal current is chosen such that the net current in each plaquette vanishes.

6. Spectral weight transfer

To quantify the spectral weight transfer evident in Fig. 10, we compute the high and low spectral weight by integrating the DOS from a value of the energy inside the Mott gap which minimizes the DOS to ∞ ($-\infty$ for electron doping) and from μ to that fixed energy, respectively. The results shown in Fig. 16 demonstrate that the initial spectral weight in the UHB which is 1/2 at $n=1$ all moves to low energies as the filling decreases, as is observed experimentally.[29,30] The integrated spectral weight has been normalized per spin. The same is true for electron doping ($n>1$). Further, the curvature of the low-energy spectral weight is positive as a function of doping in agreement with earlier results[38] on the 1D Hubbard model. This signifies that the integrated low-energy spectral weight increases faster than $2x$. The additional low-energy spectral weight above that dictated by state counting

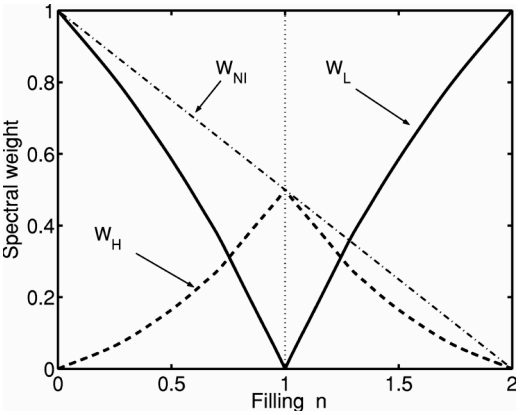

FIG. 16. High (W_H) and low (W_L) spectral weight as a function of filling. $W_{\rm NI}$ is the spectral weight in the noninteracting system.

(see Fig. 1) arises from virtual excitations between the lower and upper Hubbard bands. Such virtual transitions arise from the three-site terms discussed previously. That the number of low-energy degrees of freedom arise from such high-energy processes further attests to the inseparability of the low- and high-energy degrees of freedom in a strongly correlated system. For contrast, the spectral weight transfer for a noninteracting system (W_{NI}) is shown as well. This behavior is expected for a doped band insulator. That Mottness leads to such a drastic deviation from the noninteracting result is a direct consequence of state fractionalization. That is, each state has spectral weight both above and below the chemical potential (see Fig. 16).

7. Hall coefficient

Experimentally, the Hall coefficient in the cuprates is in general positive in the lightly hole-doped regime, scales as $1/x$ in the vicinity of half-filling but falls off faster than $1/x$ for $x \approx 0.1$ and in some instances changes sign[82–84] typically around $x=0.25$. Although the Hall coefficient is in general temperature dependent as emphasized extensively by Anderson,[85] the Takagi *et al.* experiments[83] indicate that the zero-crossing doping level in $La_{2-x}Sr_xCuO_4$ (LSCO) is only weakly temperature dependent. Consequently, we will focus solely on the doping dependence of the Hall coefficient since the existence of the zero crossing is only weakly doping dependent. Nonetheless, because the sign change is not universally observed in all the cuprates, the general conditions under which a sign change of the Hall coefficient should be observed in a doped Mott insulator have not been formulated. In addition, there have been numerous theoretical treatments of the Hall coefficient. For example, perturbative schemes[82,86–88] lead to a sign change of R_H and hence offer a possible explanation for the deviation from $1/x$. However, because perturbation theory is constrained by Luttinger's theorem[89] to yield a Fermi surface occupying half the FBZ at half-filling, such approaches fail to recover the experimentally observed[82] divergence of R_H at half-filling. In strong-coupling calculations, some have obtained a sign change[90–92] while others[93] predict that $R_H<0$ for all hole dopings. In addition, others[94,95] have reached the counterintuitive conclusion that R_H does change sign, but the Fermi surface is closed for all $x>0$. In such studies, it was assumed that the doped Mott insulator is described by doping the diamond-shaped Fermi surface of the weakly-interacting system, an assumption clearly not borne out by experiment.[77,78,96]

On simple grounds, however, the general doping dependence of the Hall coefficient can be easily deduced. Consider a lightly doped Mott insulator in which the Hall coefficient is initially positive. In the heavily overdoped regime where the system is weakly interacting, the Fermi surface must be closed and hence electronlike; thus $R_H<0$. We can deduce the doping level at which the transition from an open to a closed Fermi surface occurs by appealing to the spectral function of a Mott insulator. As Figs. 3 and 4 illustrate, at half-filling every $\mathbf{k}$ state in the FBZ has some spectral weight. Because the chemical potential [see Fig. 7] simply moves down through the LHB upon doping, hole doping simply depletes the spectral weight in the LHB. When half the spectral weight in the LHB is removed, the surface separating occupied from empty states must have zero curvature. As this surface defines the Fermi surface, then its curvature should be related to the sign of the Hall coefficient. Consequently, the critical doping level at which $R_H=0$ is determined by the doping level at which half the spectral weight in the LHB is depleted. Consider the $U=\infty$ limit in which the spectral function is momentum independent. Each state in the LHB carries the weight $1-n/2$. Consequently, the fraction of the spectral weight depleted upon hole doping is $n/2/(1-n/2)$. When this quantity equals 1/2, R_H should vanish. The solution to

$$\frac{n/2}{1-n/2}=\frac{1}{2} \quad \text{(lower limit on the filling at which } R_H=0\text{)} \tag{63}$$

is $n=2/3$ which is a strict lower bound for n_c and is precisely what simulations as well as complicated series expansions in the infinite-U limit[92,94] obtain. Consequently, if a Mott insulator possess a sign change in the Hall coefficient, it must occur for $x<0.333$.

Shown in Fig. 17 is the spectral function in the FBZ evaluated at the chemical potential for $U=8t$ and $T=0.25t$. The upper panel corresponds to $n=0.97$ and the lower to $n=0.3$. As is clear, in the lightly doped regime, the Fermi surface is holelike and the spectral features are broad indicating an absence of well-defined quasiparticles indicative of an incoherent metal as is seen experimentally.[77,78] The source of the incoherence stems from the self-energy shown in Fig. 7 which remains constant at the Fermi level at $n=0.97$. This leads necessarily to a violation of Luttinger's theorem. In fact, the Fermi surface (defined by the maximum in the spectral function) volume at $n=0.97$ is roughly 30% larger than the Luttinger volume. In the overdoped regime, the self-energy has the characteristic ω^2 dependence of a Fermi liquid and hence we recover Luttinger's theorem as the sharp spectral features in Fig. 17 reveal for $n=0.30$. Our results indicate a smooth crossover between the lightly doped regime and overdoped regimes where Luttinger's theorem is reinstated.

However, broad spectral features are not the only contributor to the violation of Luttinger's theorem. Consider the static approximation in which the self-energy in Eq. (5) is explicitly set to zero. The details of this level of theory are derived in Appendix E. At this level of theory, the spectral function for the LHB and UHB's correspond to a series of δ functions. Nonetheless, the bands generated do not describe Fermi liquid quasiparticles because each $\mathbf{k}$ state still has spectral weight both below and above the Fermi level. Relative to the dynamical results, we find that the topology and volume of the Fermi surface do not change as revealed by Fig. 18. The solid line corresponds to $U=8t$, dashed line to $U=1000t$ and dashed-dotted to $U=0$. Clearly shown in Fig. 18 is the evolution from a hole to an electronlike Fermi surface at critical doping levels of 0.791 for $U=8t$ and 0.668 for $U=1000t$. The critical concentrations at which the curvature of the Fermi surface changes sign corresponds to $x_c=0.668$ and 0.791 for $U=1000t$ and $U=8t$, respectively. As

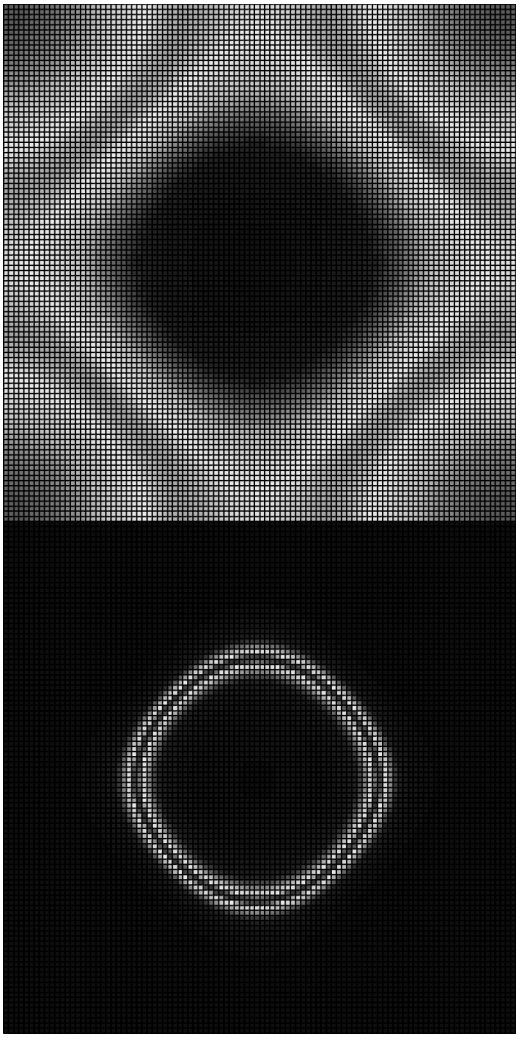

FIG. 17. Spectral function in the first Brillouin zone evaluated at the chemical potential for a filling of $n=0.97$ (top panel) and $n=0.30$ (bottom panel) with $U=8t$ and $T=0.25t$. In the underdoped regime, the spectral function has broad features at the Fermi level. A sharp Fermi surface emerges in the overdoped regime.

anticipated, these values are less than $x_c=2/3$. While for $U=8t$, x_c is remarkably close to the $x_{\mathrm{crit}}=0.19$ of Loram,[69] it is unclear whether the closing of the pseudogap is always accompanied with a sign change of the Hall coefficient. However, a sudden sign change would certainly explain the appearance of the peak in the density of states shown in Fig. 10 once the pseudogap vanishes. Three additional features are apparent. First, at small concentrations, regardless of U, all Fermi surfaces are electronlike and coincide with the noninteracting limit. An analytical proof of this result is given in Appendix E. Second, at intermediate fillings, the Fermi surface (FS) in the interacting system is holelike as opposed to electronlike in the noninteracting system. Finally, the area of the FS for $U=8t$ and $n=1-x=0.97$ is clearly larger than that dictated by Luttinger's theorem, $2\pi^2(1-x)=1.94\pi^2$. From the maximum in the spectral function, we find that the experimental value for the FS area in LSCO (Ref. 79) for $n=0.97$ is $2.06\pi^2$ which represents a nontrivial 8% deviation from the Luttinger result. Such a large deviation cannot

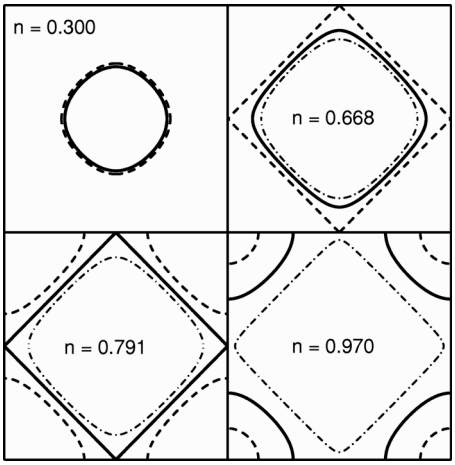

FIG. 18. Fermi surface in the static approximation for four fillings as indicated and three different values of U: (1) solid line, $U=8t$, (2) dashed line, $U=1000t$, and (3) dashed-dotted line, $U=0$, the noninteracting limit.

be attributed to experimental uncertainty (at most 2–3%).[97] Hence, while the FS computed here is clearly larger ($2.5\pi^2$) than the experimental value, both are in qualitative agreement that Luttinger's theorem is violated in the underdoped regime. Of course, improved quantitative agreement with experiment can be obtained by including band parameters such as a next-nearest-neighbor hopping interaction, t' and a doping dependent U. Hence, we see that even though the UHB and LHB's are sharp, the Luttinger volume is not preserved. In Appendix E, we show explicitly that the source of this breakdown stems from the bifurcation of the spectral weight of each $\mathbf{k}$ state into a high- and low-energy part. Consequently, removing a single electron is no longer accomplished simply by removing a single $\mathbf{k}$ state as a result of the breakdown of the band insulator sum rule. Hence, a key consequence of Mottness in two dimensions is a violation of Luttinger's theorem for $n\neq 1$ as additional extensive numerical work attests.[45–48,98] In the heavily overdoped regime [see Fig. 18], the spectral lies predominantly in the LHB and hence one hole=one $\mathbf{k}$ state and Luttinger's theorem is reinstated. Consequently, under hole doping, the hole and electron regimes are fundamentally asymmetrical as emphasized by Hirsch.[99]

To compute the Hall coefficient,

$$R_H=\sigma_{xyz}/\sigma_{xx}^2 \tag{64}$$

we work within Boltzmann transport theory in which[100]

$$\sigma_{xyz}=\frac{e^3\tau^2}{\hbar\Omega c}\sum_{\mathbf{k}} v_x(\mathbf{v}\times\nabla_k)_z v_y\left(-\frac{\partial f}{\partial\epsilon_k}\right), \tag{65}$$

$$\sigma_{xx}=\frac{e^2\tau}{\Omega}\sum_{\mathbf{k}} v_x^2\left(-\frac{\partial f}{\partial\epsilon_k}\right). \tag{66}$$

Here, $1/\tau$ is the scattering rate, Ω the volume, and f the Fermi distribution function. Our use of the Boltzmann equa-

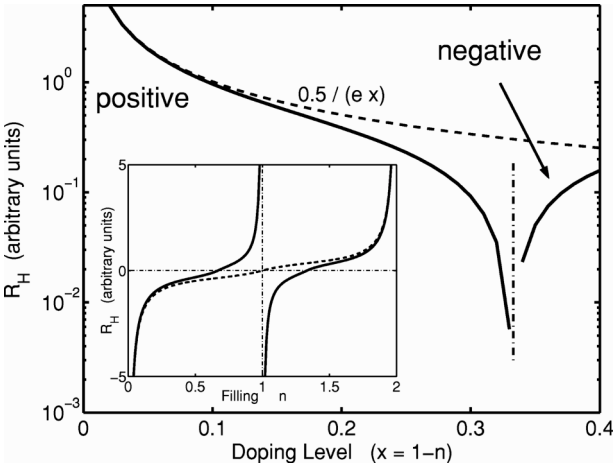

FIG. 19. Hall coefficient as a function of doping using the static approximation for the Hubbard operators with $U=1000t$. The inset shows that there is an antisymmetry between electron and hole doping. In the inset, the solid line corresponds to $U=1000t$ and the dashed line to $U=0$. Both show clearly that the deviation from $1/x$ is induced by the sign change rather than a liberation of charge.

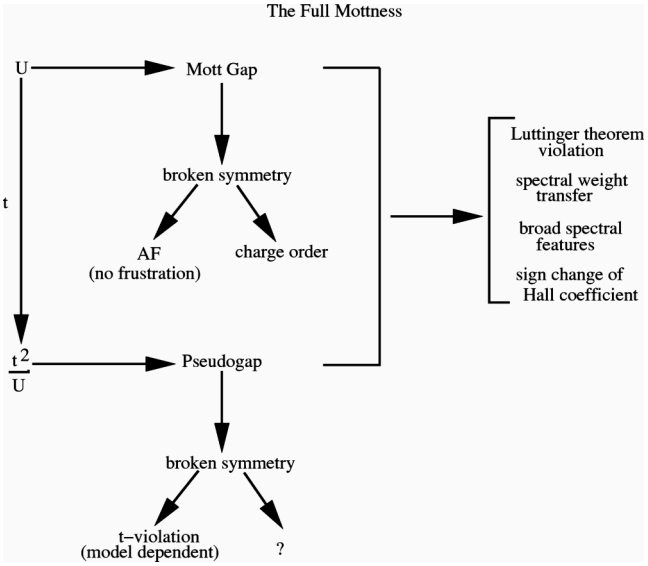

FIG. 20. Heirarchy of energy scales and the corresponding physical processes that obtain in a doped Mott insulator. AF represents antiferromagentic order.

tion should suffice as long as the interactions dominate which is certainly true in the case of interest, $T \ll U$. In using the Boltzmann approach, it is easiest to work in the large U limit because in this case, the spectral function is independent of momentum. Consequently, we used the static energy bands for $U=1000t$ and compute R_H using Eqs. (64)–(66). Figure 19 demonstrates that the sign of the Hall constant is consistent with the curvature of the Fermi surfaces shown in Fig. 18. In addition, the deviation from $1/x$ in the region close to $x_c=1-0.668$ is tied to the impending sign change. As the inset illustrates, R_H diverges at half-filling and changes sign for both electron and hole doping in contrast to weakly interacting scenarios which can yield at most one sign change (dashed line) and no divergence at $n=1$. Of course, the static approximation does not include the pseudogap found earlier. Note that regardless of which model is used for the pseudogap, the $(\pi,0)$ regions of the Fermi surface become gapped. Unless the curvature of the Fermi surface is modified by the removal of the $(\pi,0)$ regions, the pseudogap cannot change the sign of the Hall coefficient nor eliminate the divergence at half-filling. In fact, at the doping level $(x \approx x_c)$ at which the removal of the $(\pi,0)$ regions is most likely to affect the curvature of the FS, the pseudogap vanishes.

IV. FINAL REMARKS

We have explored here a dynamical method which incorporates the local physics of a doped Mott insulator. The summary of our findings is catalogued in Fig. 20. Physics on the Mott gap scale U as well as the nearest-neighbor interaction scale J play several key roles. The Mott scale, U, sets the energy range for spectral weight transfer and leads to a breakdown of the band insulator sum rule. This ultimately leads to a Fermi surface volume that exceeds that dictated by Luttinger's theorem in the underdoped regime. Reinstatement of the Luttinger volume in the overdoped regime points to a fundamental asymmetry between holes and electrons in a hole-doped Mott insulator. An additional role played by the Mott scale is the generation of a hierarchy of interactions of increasing range. The most important of these is the nearest-neighbor interaction, $J \approx t^2/U$. Spectral weight transfer across the Mott gap points to an inseparability of high- and low-energy scales. Hence, it is unclear in what sense a true low-energy theory can be formulated for a doped Mott insulator. We propose that the antiferromagnet that forms in a Mott insulator is distinct from a spin-density wave antiferromagnet. Finally, we have found that the J is also responsible for the pseudogap. The pseudogap simply reflects the restricted phase space that strongly correlated excitations on neighboring sites encounter. The current pattern shown in Fig. 15 arises from such neighboring correlations and could explain the origin of the direction-dependent t-reversal symmetry breaking observed in the normal states of the cuprates. Three-site correlations, which lead to a doping dependent spin exchange interaction, are crucial to the vanishing of the pseudogap at x_{crit}. Because this state of affairs obtains beyond x_{opt}, our proposal resonates with that of Loram and colleagues.[69] Finally, our work suggests that doping a Mott insulator gives rise to a heirarchy of energy scales all derived from the Mott gap U. Hence, the Mott state found here has the high-energy scale needed to explain the spectral weight transfer from 2 eV to the Fermi energy when superconductivity obtains.[34–36] Whether the emergence of successively lower-energy scales as a function of doping can be formulated within a renormalization group scheme remains an open question in strongly correlated electron physics. Nonetheless, it is along these lines that our current work is directed.

APPENDIX A: EQUATIONS OF MOTION FOR THE TWO-LEVEL OPERATORS

Consider the orthogonalized basis for the two-site problem Φ_n Eqs. (17) and (18). The equations of motion

$$i\frac{\partial}{\partial t}\Phi_n=[\Phi_n,H], \tag{A1}$$

where H is the Hubbard Hamiltonian, can be easily obtained using the equations of motion for the single-site level operators, Eq. (26). Assuming that Φ_n is defined on x and x', a superscript $\bar{\alpha}$ will indicate a sum over all the nearest neighbors of x with the exception of x' divided by the total number of nearest neighbors. Similarly, $\bar{\alpha}'$ will include all the nearest neighbors of x' with the exception of x. Explicitly, we have

$$\begin{aligned}i\frac{\partial}{\partial t}FB_S^\sigma=&\left(2\varepsilon_0-\mu-\frac{\tilde{t}}{2d}\right)FB_S^\sigma-\frac{\tilde{t}}{\sqrt{2}}(c_\sigma^{\bar{\alpha}}+c_\sigma^{\bar{\alpha}'})BB\\&-\frac{\tilde{t}}{\sqrt{2}}(c_\sigma^{\dagger\bar{\alpha}}-c_\sigma^{\dagger\bar{\alpha}'})FF^\sigma-\frac{\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})FF_S\\&+\frac{\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})FF_A-\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})DB_S\\&-\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})DB_A,\end{aligned} \tag{A2}$$

$$\begin{aligned}i\frac{\partial}{\partial t}FB_A^\sigma=&\left(2\varepsilon_0-\mu+\frac{\tilde{t}}{2d}\right)FB_A^\sigma-\frac{\tilde{t}}{\sqrt{2}}(c_\sigma^{\bar{\alpha}}-c_\sigma^{\bar{\alpha}'})BB\\&+\frac{\tilde{t}}{\sqrt{2}}(c_\sigma^{\dagger\bar{\alpha}}-c_\sigma^{\dagger\bar{\alpha}'})FF^\sigma+\frac{\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})FF_S\\&-\frac{\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})FF_A-\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})DB_S\\&-\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})DB_A,\end{aligned} \tag{A3}$$

$$\begin{aligned}i\frac{\partial}{\partial t}FD_S^\sigma=&\left(2\varepsilon_0-3\mu+U+\frac{\tilde{t}}{2d}\right)FD_S^\sigma-\frac{\sigma\tilde{t}}{\sqrt{2}}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})DD\\&-\frac{\sigma\tilde{t}}{\sqrt{2}}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})FF^\sigma+\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}-c_{-\sigma}^{\bar{\alpha}'})FF_S\\&-\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}+c_{-\sigma}^{\bar{\alpha}'})FF_A-\frac{\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}+c_{-\sigma}^{\bar{\alpha}'})DB_S\\&+\frac{\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}-c_{-\sigma}^{\bar{\alpha}'})DB_A,\end{aligned} \tag{A4}$$

$$\begin{aligned}i\frac{\partial}{\partial t}FD_A^\sigma=&\left(2\varepsilon_0-3\mu+U-\frac{\tilde{t}}{2d}\right)FD_A^\sigma-\frac{\sigma\tilde{t}}{\sqrt{2}}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})DD\\&+\frac{\sigma\tilde{t}}{\sqrt{2}}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})FF^\sigma-\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}+c_{-\sigma}^{\bar{\alpha}'})FF_S\\&+\frac{\sigma\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}-c_{-\sigma}^{\bar{\alpha}'})FF_A-\frac{\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}-c_{-\sigma}^{\bar{\alpha}'})DB_S\\&+\frac{\tilde{t}}{2}(c_{-\sigma}^{\bar{\alpha}}+c_{-\sigma}^{\bar{\alpha}'})DB_A,\end{aligned} \tag{A5}$$

$$\begin{aligned}i\frac{\partial}{\partial t}BB=&2\varepsilon_0BB-\frac{\tilde{t}}{\sqrt{2}}\sum_\tau(c_\tau^{\dagger\bar{\alpha}}+c_\tau^{\dagger\bar{\alpha}'})FB_S^\tau\\&-\frac{\tilde{t}}{\sqrt{2}}\sum_\tau(c_\tau^{\dagger\bar{\alpha}}-c_\tau^{\dagger\bar{\alpha}'})FB_A^\tau,\end{aligned} \tag{A6}$$

$$\begin{aligned}i\frac{\partial}{\partial t}DD=&2(\varepsilon_0-2\mu+U)DD-\frac{\tilde{t}}{\sqrt{2}}\sum_\tau\tau(c_{-\tau}^{\bar{\alpha}}+c_{-\tau}^{\bar{\alpha}'})FD_S^\tau\\&-\frac{\tilde{t}}{\sqrt{2}}\sum_\tau\tau(c_{-\tau}^{\bar{\alpha}}-c_{-\tau}^{\bar{\alpha}'})FD_A^\tau,\end{aligned} \tag{A7}$$

$$\begin{aligned}i\frac{\partial}{\partial t}FF^\sigma=&2(\varepsilon_0-\mu)FF^\sigma-\frac{\tilde{t}}{\sqrt{2}}(c_\sigma^{\bar{\alpha}}-c_\sigma^{\bar{\alpha}'})FB_S^\sigma\\&+\frac{\tilde{t}}{\sqrt{2}}(c_\sigma^{\bar{\alpha}}+c_\sigma^{\bar{\alpha}'})FB_A^\sigma-\frac{\sigma\tilde{t}}{\sqrt{2}}(c_{-\sigma}^{\dagger\bar{\alpha}}-c_{-\sigma}^{\dagger\bar{\alpha}'})FD_S^\sigma\\&+\frac{\sigma\tilde{t}}{\sqrt{2}}(c_{-\sigma}^{\dagger\bar{\alpha}}+c_{-\sigma}^{\dagger\bar{\alpha}'})FD_A^\sigma,\end{aligned} \tag{A8}$$

$$\begin{aligned}i\frac{\partial}{\partial t}FF_S=&2(\varepsilon_0-\mu)FF_S-\frac{\tilde{t}}{2}\sum_\tau(c_{-\tau}^{\bar{\alpha}}-c_{-\tau}^{\bar{\alpha}'})FB_S^\tau\\&+\frac{\tilde{t}}{2}\sum_\tau(c_{-\tau}^{\bar{\alpha}}+c_{-\tau}^{\bar{\alpha}'})FB_A^\tau+\frac{\tilde{t}}{2}\sum_\tau\tau(c_\tau^{\dagger\tilde{\alpha}}-c_\tau^{\dagger\tilde{\alpha}'})\\&\times FD_S^\tau-\sum_\tau\tau(c_\tau^{\dagger\tilde{\alpha}}+c_\tau^{\dagger\tilde{\alpha}'})FD_A^\tau,\end{aligned} \tag{A9}$$

$$\begin{aligned}i\frac{\partial}{\partial t}FF_A=&2(\varepsilon_0-\mu)FF_A+\frac{\sigma\tilde{t}}{d}DB_S+\frac{\sigma\tilde{t}}{2}\sum_\tau\tau(c_{-\tau}^{\bar{\alpha}}+c_{-\tau}^{\bar{\alpha}'})\\&\times FB_S^\tau-\frac{\sigma\tilde{t}}{2}\sum_\tau\tau(c_{-\tau}^{\bar{\alpha}}-c_{-\tau}^{\bar{\alpha}'})FB_A^\tau\\&-\frac{\sigma\tilde{t}}{2}\sum_\tau(c_\tau^{\dagger\tilde{\alpha}}+c_\tau^{\dagger\tilde{\alpha}'})FD_S^\tau\\&+\sum_\tau(c_\tau^{\dagger\tilde{\alpha}}-c_\tau^{\dagger\tilde{\alpha}'})FD_A^\tau,\end{aligned} \tag{A10}$$

$$i\frac{\partial}{\partial t}DB_S=(2\varepsilon_0-2\mu+U)DB_S+\frac{\tilde{t}}{d}FF_A-\frac{\tilde{t}}{2}\sum_\tau \tau(c^{\bar{\alpha}}_{-\tau}+c^{\bar{\alpha}'}_{-\tau})FB^\tau_S-\frac{\tilde{t}}{2}\sum_\tau \tau(c^{\bar{\alpha}}_{-\tau}-c^{\bar{\alpha}'}_{-\tau})FB^\tau_A-\frac{\tilde{t}}{2}\sum_\tau (c^{\dagger\bar{\alpha}}_{\tau}+c^{\dagger\bar{\alpha}'}_{\tau})FD^\tau_S-\frac{\tilde{t}}{2}\sum_\tau (c^{\dagger\bar{\alpha}}_{\tau}-c^{\dagger\bar{\alpha}'}_{\tau})FD^\tau_A, \tag{A11}$$

$$i\frac{\partial}{\partial t}DB_A=(2\varepsilon_0-2\mu+U)DB_A-\frac{\tilde{t}}{2}\sum_\tau \tau(c^{\bar{\alpha}}_{-\tau}-c^{\bar{\alpha}'}_{-\tau})FB^\tau_S-\frac{\tilde{t}}{2}\sum_\tau \tau(c^{\bar{\alpha}}_{-\tau}+c^{\bar{\alpha}'}_{-\tau})FB^\tau_A+\frac{\tilde{t}}{2}\sum_\tau (c^{\dagger\bar{\alpha}}_{\tau}-c^{\dagger\bar{\alpha}'}_{\tau})FD^\tau_S+\frac{\tilde{t}}{2}\sum_\tau (c^{\dagger\bar{\alpha}}_{\tau}+c^{\dagger\bar{\alpha}'}_{\tau})FD^\tau_A. \tag{A12}$$

APPENDIX B: SELF-ENERGIES FOR THE TWO-SITE RESOLVENTS

The resolvents associated with the two-site operators Φ_n can be expressed in the form

$$R_{\Phi_n\Phi_m}(\omega)=[\omega-E_{\Phi_n\Phi_m}-\Sigma_{\Phi_n\Phi_m}(\omega)]^{-1}. \tag{B1}$$

For the diagonal components, we use the notation $X_{\Phi_n\Phi_m}\equiv X_{\Phi_n}$, where X is R, E, or Σ. We evaluate the self-energies $\Sigma_{\Phi_n\Phi_m}(\omega)$ within a one-loop approximation. Using the spectral functions ρ_+ and ρ_-, Eq. (48), of the irreducible propagators, we obtain

$$\Sigma_{FB_S}(\omega)=\frac{\tilde{t}^2}{4}\int dx\{2\rho_+(x)[1-f(x)]R_{BB}(\omega-x)-\rho_+(x)f(x)R_{FF_ADB_S}(\omega+x)+\rho_+(x)f(x)R_{DB_S}(\omega+x)+\rho_-(x)f(x)\times R_{DB_A}(\omega+x)+3\rho_-(x)f(x)R_{FF_S}(\omega+x)+\rho_+(x)f(x)R_{FF_A}(\omega+x)\}, \tag{B2}$$

$$\Sigma_{FB_A}(\omega)=\frac{\tilde{t}^2}{4}\int dx\{2\rho_-(x)[1-f(x)]R_{BB}(\omega-x)+\rho_-(x)f(x)R_{FF_ADB_S}(\omega+x)+\rho_-(x)f(x)R_{DB_S}(\omega+x)+\rho_+(x)f(x)\times R_{DB_A}(\omega+x)+3\rho_+(x)f(x)R_{FF_S}(\omega+x)+\rho_-(x)f(x)R_{FF_A}(\omega+x)\}, \tag{B3}$$

$$\Sigma_{FD_S}(\omega)=\frac{\tilde{t}^2}{4}\int dx\{2\rho_+(x)f(x)R_{DD}(\omega+x)+\rho_+(x)[1-f(x)]R_{FF_ADB_S}(\omega-x)+\rho_+(x)\times[1-f(x)]R_{DB_S}(\omega-x)+\rho_-(x)[1-f(x)]\times R_{DB_A}(\omega-x)+3\rho_-(x)[1-f(x)]R_{FF_S}(\omega-x)+\rho_+(x)[1-f(x)]R_{FF_A}(\omega-x)\}, \tag{B4}$$

$$\Sigma_{FD_A}(\omega)=\frac{\tilde{t}^2}{4}\int dx\{2\rho_-(x)f(x)R_{DD}(\omega+x)-\rho_-(x)[1-f(x)]R_{FF_ADB_S}(\omega-x)+\rho_-(x)\times[1-f(x)]R_{DB_S}(\omega-x)+\rho_+(x)[1-f(x)]\times R_{DB_A}(\omega-x)+3\rho_+(x)[1-f(x)]R_{FF_S}(\omega-x)+\rho_-(x)[1-f(x)]R_{FF_A}(\omega-x)\}, \tag{B5}$$

$$\Sigma_{BB}(\omega)=\tilde{t}^2\int dx\{\rho_+(x)f(x)R_{FB_S}(\omega+x)+\rho_-(x)f(x)R_{FB_A}(\omega+x)\}, \tag{B6}$$

$$\Sigma_{DD}(\omega)=\tilde{t}^2\int dx\{\rho_+(x)[1-f(x)]R_{FD_S}(\omega-x)+\rho_-(x)[1-f(x)]R_{FD_A}(\omega-x)\}, \tag{B7}$$

$$\Sigma_{FF_S}(\omega)=\frac{\tilde{t}^2}{2}\int dx\{\rho_-(x)f(x)R_{FD_S}(\omega+x)+\rho_+(x)f(x)R_{FD_A}(\omega+x)+\rho_-(x)[1-f(x)]R_{FB_S}(\omega-x)+\rho_+(x)[1-f(x)]R_{FB_A}(\omega+x)\}, \tag{B8}$$

$$\Sigma_{FF^\sigma}(\omega)=\Sigma_{FF_S}(\omega)=\Sigma_{DB_A}(\omega), \tag{B9}$$

$$\Sigma_{FF_A}(\omega)=\frac{\tilde{t}^2}{2}\int dx\{\rho_+(x)f(x)R_{FD_S}(\omega+x)+\rho_-(x)f(x)R_{FD_A}(\omega+x)+\rho_+(x)[1-f(x)]R_{FB_S}(\omega-x)+\rho_-(x)[1-f(x)]R_{FB_A}(\omega+x)\}, \tag{B10}$$

$$\Sigma_{FF_ADB_S}(\omega)=\frac{\tilde{t}^2}{2}\int dx\{\rho_+(x)f(x)R_{FD_S}(\omega+x)-\rho_-(x)f(x)R_{FD_A}(\omega+x)-\rho_+(x)[1-f(x)]R_{FB_S}(\omega-x)+\rho_-(x)[1-f(x)]R_{FB_A}(\omega+x)\}, \tag{B11}$$

$$\Sigma_{FF_A DB_S}(\omega)=\Sigma_{DB_S FF_A}(\omega), \tag{B12}$$

$$\Sigma_{DB_S}(\omega)=\Sigma_{FF_A}(\omega). \tag{B13}$$

APPENDIX C: SELF-ENERGIES CORRECTIONS DUE TO SPIN FLUCTUATION

We include the effects of spin fluctuation as higher-order corrections to the self-energies of the resolvents. The spectral functions associated with local singlet and triplet states are sharply peaked at well-defined energies separated by the effective antiferromagnetic coupling constant

$$J=\int_{-\infty}^{\infty}\omega(\sigma_{FF_S}-\sigma_{FF_A})d\omega. \tag{C1}$$

The effects of spin fluctuations with the environment can be approximately described by the effective antiferromagnetic interaction

$$\delta H_{eff}=\tfrac{1}{2}J\mathbf{n}\mathbf{n}^{\alpha}, \tag{C2}$$

where $\mathbf{n}=c^{\dagger}\boldsymbol{\sigma}c$ is the spin density, σ_i are the Pauli matrices and $c^{\dagger}=(c_{\uparrow}^{\dagger},c_{\downarrow}^{\dagger})$. Although a proper account of the singlet-triplet mixing can be given by considering vertex corrections, we can use a simpler two-step approach: (1) Working in the basis formed by the eigenstates of the two-site problem (as we, in fact, do) we take care of the spin fluctuations for the cluster without a bath. As we introduce the bath, an additional singlet-triplet mixing occurs which is not captured by NCA (as shown by the very sharp features in the FFA and the FFS resolvents). We try to approximate this mixing by an effective spin-spin interaction. As a result of this mixing, the FFA and FFS states are broadened. (2) Due to the the self-consistency of the approach, this spin-spin interaction is also present between the cluster and the bath. The corrections that we introduce are the effect of this additional effective interaction with the bath on the self-energies of the resolvents. Consequently, some of the equations of motion for the two-site level operators Φ_n (see Appendix A), will be modified. Explicitly we have

$$i\frac{\partial}{\partial t}FB_S^{\sigma}=\frac{J}{2}[\sigma(n_3^{\bar{\alpha}}+n_3^{\bar{\alpha}'})FB_S^{\sigma}+\sigma(n_3^{\bar{\alpha}}-n_3^{\bar{\alpha}'})FB_A^{\sigma}+(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FB_S^{-\sigma}+(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FB_A^{-\sigma}]+\cdots, \tag{C3}$$

$$i\frac{\partial}{\partial t}FB_A^{\sigma}=\frac{J}{2}[\sigma(n_3^{\bar{\alpha}}-n_3^{\bar{\alpha}'})FB_S^{\sigma}+\sigma(n_3^{\bar{\alpha}}+n_3^{\bar{\alpha}'})FB_A^{\sigma}+(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FB_S^{-\sigma}+(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FB_A^{-\sigma}]+\cdots, \tag{C4}$$

$$i\frac{\partial}{\partial t}FD_S^{\sigma}=\frac{J}{2}[\sigma(n_3^{\bar{\alpha}}+n_3^{\bar{\alpha}'})FD_S^{\sigma}+\sigma(n_3^{\bar{\alpha}}-n_3^{\bar{\alpha}'})FD_A^{\sigma}+(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FD_S^{-\sigma}+(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FD_A^{-\sigma}]+\cdots, \tag{C5}$$

$$i\frac{\partial}{\partial t}FD_A^{\sigma}=\frac{J}{2}[\sigma(n_3^{\bar{\alpha}}-n_3^{\bar{\alpha}'})FD_S^{\sigma}+\sigma(n_3^{\bar{\alpha}}+n_3^{\bar{\alpha}'})FD_A^{\sigma}+(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FD_S^{-\sigma}+(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FD_A^{-\sigma}]+\cdots, \tag{C6}$$

$$i\frac{\partial}{\partial t}FF^{\sigma}=J\left[\sigma(n_3^{\bar{\alpha}}+n_3^{\bar{\alpha}'})FF^{\sigma}+\frac{1}{\sqrt{2}}(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FF_S-\frac{\sigma}{\sqrt{2}}(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FF_A\right]+\cdots, \tag{C7}$$

$$i\frac{\partial}{\partial t}FF_S=J\left[\sigma(n_3^{\bar{\alpha}}-n_3^{\bar{\alpha}'})FF_A+\frac{1}{\sqrt{2}}(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FF^{\sigma}+\frac{1}{\sqrt{2}}(n_{-\sigma}^{\bar{\alpha}}+n_{-\sigma}^{\bar{\alpha}'})FF^{-\sigma}\right]+\cdots, \tag{C8}$$

$$i\frac{\partial}{\partial t}FF_A=J\left[\sigma(n_3^{\bar{\alpha}}-n_3^{\bar{\alpha}'})FF_S-\frac{1}{\sqrt{2}}(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FF^{\sigma}+\frac{1}{\sqrt{2}}(n_{-\sigma}^{\bar{\alpha}}-n_{-\sigma}^{\bar{\alpha}'})FF^{-\sigma}\right]+\cdots, \tag{C9}$$

where we used the notation $n_{\sigma}=n_1+i\sigma n_2$. These terms will determine the self-energy corrections

$$\delta\Sigma_{FB_S}(\omega)=\int dx\,\frac{3}{2}J^2\{\varphi_+(x)[1-f(x)]R_{FB_S}(\omega-x)+\varphi_-(x)[1-f(x)]R_{FB_A}(\omega-x)\}, \tag{C10}$$

$$\delta\Sigma_{FB_A}(\omega)=\int dx\,\frac{3}{2}J^2\{\varphi_-(x)[1-f(x)]R_{FB_S}(\omega-x)+\varphi_+(x)[1-f(x)]R_{FB_A}(\omega-x)\}, \tag{C11}$$

$$\delta\Sigma_{FD_S}(\omega)=\int dx\,\frac{3}{2}J^2\{\varphi_+(x)[1-f(x)]R_{FD_S}(\omega-x)+\varphi_-(x)[1-f(x)]R_{FD_A}(\omega-x)\}, \tag{C12}$$

$$\delta\Sigma_{FD_A}(\omega)=\int dx\,\frac{3}{2}J^2\{\varphi_-(x)[1-f(x)]R_{FD_S}(\omega-x)+\varphi_+(x)[1-f(x)]R_{FD_A}(\omega-x)\}, \tag{C13}$$

$$\delta\Sigma_{FF_S}(\omega)=\int dx\ \ 2J^2\{2\varphi_+(x)[1-f(x)]R_{FF_S}(\omega-x)$$
$$+\varphi_-(x)[1-f(x)]R_{FF_A}(\omega-x)\},\qquad\text{(C14)}$$

$$\delta\Sigma_{FF_A}(\omega)=\int dx\ \ 6J^2\varphi_-(x)[1-f(x)]R_{FF_S}(\omega-x),\qquad\text{(C15)}$$

where $\varphi_\pm(\omega)=\varphi(\omega)\pm\varphi'(\omega)$ and

$$\varphi(\omega)=\mathrm{FT}\langle\{n_3^{\bar\alpha},n_3^{\bar\alpha}\}\rangle=\tfrac{1}{2}\mathrm{FT}\langle\{n_{-\sigma}^{\bar\alpha},n_{-\sigma}^{\bar\alpha}\}\rangle,\qquad\text{(C16)}$$

$$\varphi'(\omega)=\mathrm{FT}\langle\{n_3^{\bar\alpha},n_3^{\bar{\alpha'}}\}\rangle=\tfrac{1}{2}\mathrm{FT}\langle\{n_{-\sigma}^{\bar\alpha},n_{-\sigma}^{\bar{\alpha'}}\}\rangle.$$

Note because $\varphi(\omega)$ and $\varphi'(\omega)$ are expressed in terms of functions defined on $\bar\alpha$ and $\bar{\alpha'}$ which denote the nearest neighbors of the two sites in the cluster, we have avoided double counting the spin fluctuations on the cluster. The functions $\varphi(\omega)$ and $\varphi'(\omega)$ are computed within a noncrossing approximation, taking into account the only dominant contributions (containing products of on-site and nearest-neighbor operators). Introducing the coordination factors $\lambda=(2d-1)/(2d)^2$ and $\lambda'=(2d-2)/(2d)^2$, we obtain

$$\varphi(\omega)\approx\frac{\lambda}{Z}\int dx\Big\{\frac{1}{2}[\sigma_{FB_S}(x)\bar\sigma_{FB_S}(\omega-x)+\bar\sigma_{FB_S}(x)\sigma_{FB_S}(\omega-x)]+\frac{1}{2}[\sigma_{FB_S}(x)\bar\sigma_{FB_A}(\omega-x)+\bar\sigma_{FB_S}(x)\sigma_{FB_A}(\omega-x)]$$
$$+\frac{1}{2}[\sigma_{FB_A}(x)\bar\sigma_{FB_S}(\omega-x)+\bar\sigma_{FB_A}(x)\sigma_{FB_S}(\omega-x)]+\frac{1}{2}[\sigma_{FB_A}(x)\bar\sigma_{FB_A}(\omega-x)+\bar\sigma_{FB_A}(x)\sigma_{FB_A}(\omega-x)]$$
$$+\frac{1}{2}[\sigma_{FD_S}(x)\bar\sigma_{FD_S}(\omega-x)+\bar\sigma_{FB_S}(x)\sigma_{FB_S}(\omega-x)]+\frac{1}{2}[\sigma_{FD_S}(x)\bar\sigma_{FD_A}(\omega-x)+\bar\sigma_{FB_S}(x)\sigma_{FB_A}(\omega-x)]$$
$$+\frac{1}{2}[\sigma_{FD_A}(x)\bar\sigma_{FD_S}(\omega-x)+\bar\sigma_{FB_A}(x)\sigma_{FB_S}(\omega-x)]+\frac{1}{2}[\sigma_{FD_A}(x)\bar\sigma_{FD_A}(\omega-x)+\bar\sigma_{FB_A}(x)\sigma_{FB_A}(\omega-x)]$$
$$+2[\sigma_{FF_S}(x)\bar\sigma_{FF_S}(\omega-x)+\bar\sigma_{FF_S}(x)\sigma_{FF_S}(\omega-x)]+\sigma_{FF_S}(x)\bar\sigma_{FF_A}(\omega-x)+\bar\sigma_{FF_S}(x)\sigma_{FF_A}(\omega-x)$$
$$+\sigma_{FF_A}(x)\bar\sigma_{FF_S}(\omega-x)+\bar\sigma_{FF_A}(x)\sigma_{FF_S}(\omega-x)\Big\},\qquad\text{(C17)}$$

$$\varphi'(\omega)\approx\frac{\lambda'}{Z}\int dx\Big\{\frac{1}{2}[\sigma_{FB_S}(x)\bar\sigma_{FB_S}(\omega-x)+\bar\sigma_{FB_S}(x)\sigma_{FB_S}(\omega-x)]-\frac{1}{2}[\sigma_{FB_S}(x)\bar\sigma_{FB_A}(\omega-x)-\bar\sigma_{FB_S}(x)\sigma_{FB_A}(\omega-x)]$$
$$-\frac{1}{2}[\sigma_{FB_A}(x)\bar\sigma_{FB_S}(\omega-x)-\bar\sigma_{FB_A}(x)\sigma_{FB_S}(\omega-x)]+\frac{1}{2}[\sigma_{FB_A}(x)\bar\sigma_{FB_A}(\omega-x)+\bar\sigma_{FB_A}(x)\sigma_{FB_A}(\omega-x)]$$
$$+\frac{1}{2}[\sigma_{FD_S}(x)\bar\sigma_{FD_S}(\omega-x)+\bar\sigma_{FB_S}(x)\sigma_{FB_S}(\omega-x)]-\frac{1}{2}[\sigma_{FD_S}(x)\bar\sigma_{FD_A}(\omega-x)-\bar\sigma_{FB_S}(x)\sigma_{FB_A}(\omega-x)]$$
$$-\frac{1}{2}[\sigma_{FD_A}(x)\bar\sigma_{FD_S}(\omega-x)-\bar\sigma_{FB_A}(x)\sigma_{FB_S}(\omega-x)]+\frac{1}{2}[\sigma_{FD_A}(x)\bar\sigma_{FD_A}(\omega-x)+\bar\sigma_{FB_A}(x)\sigma_{FB_A}(\omega-x)]$$
$$+2[\sigma_{FF_S}(x)\bar\sigma_{FF_S}(\omega-x)+\bar\sigma_{FF_S}(x)\sigma_{FF_S}(\omega-x)]-\sigma_{FF_S}(x)\bar\sigma_{FF_A}(\omega-x)-\bar\sigma_{FF_S}(x)\sigma_{FF_A}(\omega-x)$$
$$-\sigma_{FF_A}(x)\bar\sigma_{FF_S}(\omega-x)-\bar\sigma_{FF_A}(x)\sigma_{FF_S}(\omega-x)\Big\}.\qquad\text{(C18)}$$

Note that in these equations, it is the Fermi function which appears not the Bose distribution function. In general, the expression for the self-energy depends on which form one uses for the spectral function. We used $\mathrm{FT}\angle\{n_3(t),n_3(t')\}\langle=\varphi(\omega)$. Usually the pectral function is defined by $\mathrm{FT}\langle[n_3(t),n_3(t')]\rangle=\rho(\omega)$. They are related as $\varphi(\omega)=[(e^{\beta\omega}+1)/(e^{\beta\omega}-1)]\rho(\omega)$. The loop contains

$$\int d\omega' \rho(\omega'+\omega) R(-\omega') \frac{e^{\beta(\omega+\omega')}}{e^{\beta(\omega+\omega')}-1} \qquad \text{(C19)}$$

which, in terms of φ, becomes

$$\int d\omega' \varphi(\omega'+\omega) R(-\omega') \frac{e^{\beta(\omega+\omega')}}{e^{\beta(\omega+\omega')}+1}, \qquad \text{(C20)}$$

thereby justifying the Fermi function in the self-energy corrections.

APPENDIX D: LEVEL OPERATOR REPRESENTATION OF δJ AND EXPLICIT EXPRESSIONS FOR Dm_0 AND Dm_1

With the assumption used to derive Eq. (12), we find that the the dynamical correction operator δJ, Eq. (7), can be expressed in terms of two-site level operators,

$$\delta J_\sigma = -\tilde{t} \sum_{m,n} a_{nm} \Phi_n^\dagger \Phi_m . \qquad \text{(D1)}$$

With the notation

$$a_1 = eI_1^{-1}, \quad b_1 = \frac{n}{2} - \tilde{p} I_1^{-1},$$

$$a_2 = eI_2^{-1}, \quad b_2 = \frac{n}{2} + \tilde{p} I_2^{-1}, \qquad \text{(D2)}$$

we obtain for the coefficients a_{nm}:

$$a_{BB\,FB_S^\sigma} = \frac{-1}{\sqrt{2}}(a_1 - b_1), \quad a_{FD_S^{-\sigma}\,DD} = \frac{-\sigma}{\sqrt{2}}(1 - a_2 - b_2),$$

$$a_{BB\,FB_A^\sigma} = \frac{-1}{\sqrt{2}}(a_1 + b_1), \quad a_{FD_A^{-\sigma}\,DD} = \frac{\sigma}{\sqrt{2}}(1 + a_2 - b_2),$$

$$a_{FB_S^{-\sigma}\,FF_S} = \frac{1}{2}(a_1 + b_1), \quad a_{FF_SFD_S^\sigma} = \frac{-\sigma}{2}(1 + a_2 - b_2),$$

$$a_{FB_A^{-\sigma}\,FF_S} = \frac{-1}{2}(a_1 - b_1), \quad a_{FF_SFD_A^\sigma} = \frac{-\sigma}{2}(1 - a_2 - b_2),$$

$$a_{FB_S^\sigma\,FF^\sigma} = \frac{1}{\sqrt{2}}(a_1 + b_1),$$

$$a_{FF^{-\sigma}\,FD_S^{-\sigma}} = \frac{-\sigma}{\sqrt{2}}(1 + a_2 - b_2),$$

$$a_{FB_A^\sigma\,FF^\sigma} = \frac{-1}{\sqrt{2}}(a_1 - b_1),$$

$$a_{FF^{-\sigma}\,FD_A^{-\sigma}} = \frac{-\sigma}{\sqrt{2}}(1 - a_2 - b_2),$$

$$a_{FB_S^{-\sigma}\,FF_A} = \sigma(1 + a_1 - b_1), \quad a_{FF_AFD_S^\sigma} = \tfrac{1}{2}(1 + a_2 + b_2),$$

$$a_{FB_A^{-\sigma}\,FF_A} = \sigma(1 - a_1 - b_1), \quad a_{FF_AFD_A^\sigma} = \tfrac{1}{2}(1 - a_2 + b_2),$$

$$a_{DB_SFD_S^\sigma} = \frac{-1}{2}(2 + a_1 - b_1), \quad a_{FB_S^{-\sigma}\,DB_S} = \frac{\sigma}{2}(1 + a_2 + b_2),$$

$$a_{DB_S\,FD_A^\sigma} = \frac{1}{2}(2 - a_1 - b_1), \quad a_{FB_A^{-\sigma}\,DB_S} = \frac{-\sigma}{2}(1 - a_2 + b_2),$$

$$a_{DB_AFD_S^\sigma} = \frac{1}{2}(a_1 + b_1), \quad a_{FB_S^{-\sigma}\,DB_A} = \frac{\sigma}{2}(1 + a_2 - b_2),$$

$$a_{DB_AFD_A^\sigma} = \frac{1}{2}(a_1 - b_1), \quad a_{FB_A^{-\sigma}\,DB_A} = \frac{-\sigma}{2}(1 - a_2 - b_2). \qquad \text{(D3)}$$

Using these coefficients, we can write explicitly the dynamical corrections Dm_0 and Dm_1. From Eq. (12), we have

$$Dm_0(\omega) = \frac{\tilde{t}^2}{Z} \int dx dx' \frac{1}{\omega - x + x' + i\delta} \{ a^2_{BBFB_S}[\sigma_{FB_S}(x)\bar{\sigma}_{BB}(x') + \bar{\sigma}_{FB_S}(x)\sigma_{BB}(x')] + a^2_{BBFB_A}[\sigma_{FB_A}(x)\bar{\sigma}_{BB}(x')$$

$$+ \bar{\sigma}_{FB_A}(x)\sigma_{BB}(x')] + a^2_{FD_SDD}[\sigma_{DD}(x)\bar{\sigma}_{FD_S}(x') + \bar{\sigma}_{DD}(x)\sigma_{FD_S}(x')] + a^2_{FD_ADD}[\sigma_{DD}(x)\bar{\sigma}_{FD_A}(x')$$

$$+ \bar{\sigma}_{DD}(x)\sigma_{FD_A}(x')] + (a^2_{FB_SFF_S} + a^2_{FB_SFF^\sigma})[\sigma_{FF_S}(x)\bar{\sigma}_{FB_S}(x') + \bar{\sigma}_{FF_S}(x)\sigma_{FB_S}(x')] + (a^2_{FB_AFF_S} + a^2_{FB_AFF^\sigma})$$
$$\times [\sigma_{FF_S}(x)\bar{\sigma}_{FB_A}(x') + \bar{\sigma}_{FF_S}(x)\sigma_{FB_A}(x')] + (a^2_{FF_SFD_S} + a^2_{FF^\sigma FD_S})[\sigma_{FD_S}(x)\bar{\sigma}_{FF_S}(x') + \bar{\sigma}_{FD_S}(x)\sigma_{FF_S}(x')]$$
$$+ (a^2_{FF_SFD_A} + a^2_{FF^\sigma FD_A})[\sigma_{FD_A}(x)\bar{\sigma}_{FF_S}(x') + \bar{\sigma}_{FD_A}(x)\sigma_{FF_S}(x')] + a^2_{FB_SFF_A}[\sigma_{FF_A}(x)\bar{\sigma}_{FB_S}(x') + \bar{\sigma}_{FF_A}(x)\sigma_{FB_S}(x')]$$
$$+ a^2_{FB_AFF_A}[\sigma_{FF_A}(x)\bar{\sigma}_{FB_A}(x') + \bar{\sigma}_{FF_A}(x)\sigma_{FB_A}(x')] + a^2_{FF_AFD_S}[\sigma_{FD_S}(x)\bar{\sigma}_{FF_A}(x') + \bar{\sigma}_{FD_S}(x)\sigma_{FF_A}(x')]$$

$$+a^2_{FF_AFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{FF_A}(x')+\overline{\sigma}_{FD_A}(x)\sigma_{FF_A}(x')]+2a_{FB_SDB_S}a_{FB_SFF_A}[\sigma_{DB_SFF_A}(x)\overline{\sigma}_{FB_S}(x')$$
$$+\overline{\sigma}_{DB_SFF_A}(x)\sigma_{FB_S}(x')]+2a_{FB_ADB_S}a_{FB_AFF_A}[\sigma_{DB_SFF_A}(x)\overline{\sigma}_{FB_A}(x')+\overline{\sigma}_{DB_SFF_A}(x)\sigma_{FB_A}(x')]$$
$$+2a_{DB_SFD_S}a_{FF_AFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{DB_SFF_A}(x')+\overline{\sigma}_{FD_S}(x)\sigma_{DB_SFF_A}(x')]+2a_{DB_SFD_A}a_{FF_AFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{DB_SFF_A}(x')$$
$$+\overline{\sigma}_{FD_A}(x)\sigma_{DB_SFF_A}(x')]+a^2_{FB_SDB_S}[\sigma_{DB_S}(x)\overline{\sigma}_{FB_S}(x')+\overline{\sigma}_{DB_S}(x)\sigma_{FB_S}(x')]+a^2_{FB_ADB_S}[\sigma_{DB_S}(x)\overline{\sigma}_{FB_A}(x')$$
$$+\overline{\sigma}_{DB_S}(x)\sigma_{FB_A}(x')]+a^2_{DB_SFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{DB_S}(x')+\overline{\sigma}_{FD_S}(x)\sigma_{DB_S}(x')]+a^2_{DB_SFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{DB_S}(x')$$
$$+\overline{\sigma}_{FD_A}(x)\sigma_{DB_S}(x')]+a^2_{FB_SDB_A}[\sigma_{DB_A}(x)\overline{\sigma}_{FB_S}(x')+\overline{\sigma}_{DB_A}(x)\sigma_{FB_S}(x')]+a^2_{FB_ADB_A}[\sigma_{DB_A}(x)\overline{\sigma}_{FB_A}(x')$$
$$+\overline{\sigma}_{DB_A}(x)\sigma_{FB_A}(x')]+a^2_{DB_AFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{DB_A}(x')+\overline{\sigma}_{FD_S}(x)\sigma_{DB_A}(x')]+a^2_{DB_AFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{DB_A}(x')$$
$$+\overline{\sigma}_{FD_A}(x)\sigma_{DB_A}(x')]\}. \tag{D4}$$

Similarly, for Dm_1 we obtain

$$Dm_1(\omega)=\frac{\tilde{t}^2}{Z}\int dxdx'\frac{1}{\omega-x+x'+i\delta}\{a^2_{BBFB_S}[\sigma_{FB_S}(x)\overline{\sigma}_{BB}(x')+\overline{\sigma}_{FB_S}(x)\sigma_{BB}(x')]-a^2_{BBFB_A}[\sigma_{FB_A}(x)\overline{\sigma}_{BB}(x')$$
$$+\overline{\sigma}_{FB_A}(x)\sigma_{BB}(x')]+a^2_{FD_SDD}[\sigma_{DD}(x)\overline{\sigma}_{FD_S}(x')+\overline{\sigma}_{DD}(x)\sigma_{FD_S}(x')]-a^2_{FD_ADD}[\sigma_{DD}(x)\overline{\sigma}_{FD_A}(x')$$
$$+\overline{\sigma}_{DD}(x)\sigma_{FD_A}(x')]-(a^2_{FB_SFF_S}+a^2_{FB_SFF^\sigma})[\sigma_{FF_S}(x)\overline{\sigma}_{FB_S}(x')+\overline{\sigma}_{FF_S}(x)\sigma_{FB_S}(x')]+(a^2_{FB_AFF_S}+a^2_{FB_AFF^\sigma})$$
$$\times[\sigma_{FF_S}(x)\overline{\sigma}_{FB_A}(x')+\overline{\sigma}_{FF_S}(x)\sigma_{FB_A}(x')]-(a^2_{FF_SFD_S}+a^2_{FF^\sigma\ FD_S})[\sigma_{FD_S}(x)\overline{\sigma}_{FF_S}(x')+\overline{\sigma}_{FD_S}(x)\sigma_{FF_S}(x')]$$
$$+(a^2_{FF_SFD_A}+a^2_{FF^\sigma\ FD_A})[\sigma_{FD_A}(x)\overline{\sigma}_{FF_S}(x')+\overline{\sigma}_{FD_A}(x)\sigma_{FF_S}(x')]+a^2_{FB_SFF_A}[\sigma_{FF_A}(x)\overline{\sigma}_{FB_S}(x')$$
$$+\overline{\sigma}_{FF_A}(x)\sigma_{FB_S}(x')]-a^2_{FB_AFF_A}[\sigma_{FF_A}(x)\overline{\sigma}_{FB_A}(x')+\overline{\sigma}_{FF_A}(x)\sigma_{FB_A}(x')]+a^2_{FF_AFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{FF_A}(x')$$
$$+\overline{\sigma}_{FD_S}(x)\sigma_{FF_A}(x')]-a^2_{FF_AFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{FF_A}(x')+\overline{\sigma}_{FD_A}(x)\sigma_{FF_A}(x')]$$
$$+2a_{FB_SDB_S}a_{FB_SFF_A}[\sigma_{DB_SFF_A}(x)\overline{\sigma}_{FB_S}(x')+\overline{\sigma}_{DB_SFF_A}(x)\sigma_{FB_S}(x')]-2a_{FB_ADB_S}a_{FB_AFF_A}[\sigma_{DB_SFF_A}(x)\overline{\sigma}_{FB_A}(x')$$
$$+\overline{\sigma}_{DB_SFF_A}(x)\sigma_{FB_A}(x')]+2a_{DB_SFD_S}a_{FF_AFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{DB_SFF_A}(x')+\overline{\sigma}_{FD_S}(x)\sigma_{DB_SFF_A}(x')]$$
$$-2a_{DB_SFD_A}a_{FF_AFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{DB_SFF_A}(x')+\overline{\sigma}_{FD_A}(x)\sigma_{DB_SFF_A}(x')]+a^2_{FB_SDB_S}[\sigma_{DB_S}(x)\overline{\sigma}_{FB_S}(x')$$
$$+\overline{\sigma}_{DB_S}(x)\sigma_{FB_S}(x')]-a^2_{FB_ADB_S}[\sigma_{DB_S}(x)\overline{\sigma}_{FB_A}(x')+\overline{\sigma}_{DB_S}(x)\sigma_{FB_A}(x')]+a^2_{DB_SFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{DB_S}(x')$$
$$+\overline{\sigma}_{FD_S}(x)\sigma_{DB_S}(x')]-a^2_{DB_SFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{DB_S}(x')+\overline{\sigma}_{FD_A}(x)\sigma_{DB_S}(x')]-a^2_{FB_SDB_A}[\sigma_{DB_A}(x)\overline{\sigma}_{FB_S}(x')$$
$$+\overline{\sigma}_{DB_A}(x)\sigma_{FB_S}(x')]+a^2_{FB_ADB_A}[\sigma_{DB_A}(x)\overline{\sigma}_{FB_A}(x')+\overline{\sigma}_{DB_A}(x)\sigma_{FB_A}(x')]-a^2_{DB_AFD_S}[\sigma_{FD_S}(x)\overline{\sigma}_{DB_A}(x')$$
$$+\overline{\sigma}_{FD_S}(x)\sigma_{DB_A}(x')]+a^2_{DB_AFD_A}[\sigma_{FD_A}(x)\overline{\sigma}_{DB_A}(x')+\overline{\sigma}_{FD_A}(x)\sigma_{DB_A}(x')]\}. \tag{D5}$$

APPENDIX E: STATIC APPROXIMATION

We show explicitly in this appendix that the Hubbard operator technique in the static approximation[101] correctly recovers the noninteracting or Fermi liquid limit when $U=0$ regardless of the filling and as $n\to 0$ for any U. Consequently, the violation of Luttinger's theorem found here is not an artifact of the method. Consider the Hubbard operator basis, $\psi_1=\xi_{i\sigma}$ and $\psi_2=\eta_{i\sigma}$ and the associated Green functions $S_{\alpha\beta}=\langle\langle\psi_\alpha;\psi_\beta\rangle\rangle$. Within the static approximation,[101] the expression for the retarded Green function in Fourier space becomes

$$S_{\alpha\beta}(\mathbf{k},\omega)=\sum_{j=1}^{2}\frac{\sigma^{(j)}_{\alpha\beta}(\mathbf{k})}{\omega-\epsilon_j(\mathbf{k})+i\delta}. \tag{E1}$$

The dispersion relations for the two bands are given by $\epsilon_{1,2}(\mathbf{k}) = R(\mathbf{k}) \pm Q(\mathbf{k})$ with

$$R(\mathbf{k}) = \frac{1}{2}U - \mu - \frac{1}{2I_1I_2}[m(\mathbf{k}) + 8t\alpha(\mathbf{k})I_1I_2], \quad \text{(E2)}$$

$$Q(\mathbf{k}) = \frac{1}{2}\sqrt{g^2(\mathbf{k}) + \frac{4m^2(\mathbf{k})}{I_1I_2}}, \quad \text{(E3)}$$

where $I_1 = 1 - n/2$ and $I_2 = n/2$ and $\alpha(\mathbf{k}) = \frac{1}{2}[\cos(k_x) + \cos(k_y)]$ and we used the notation $m(\mathbf{k}) = 4t[e + \alpha(\mathbf{k})(p - I_2)]$ and $g(\mathbf{k}) = -U + (1-n)/[I_1I_2m(\mathbf{k})]$ where $e = \langle \xi^\dagger_{i\sigma}\xi^\alpha_{i\sigma}\rangle - \langle \eta^\alpha_{i\sigma}\eta_{i\sigma}\rangle$ and $p = \langle n_{i\sigma}n^\alpha_{i\sigma}\rangle + \langle S_iS_i^{\dagger\alpha}\rangle - \langle b_ib_i^{\dagger\alpha}\rangle$ with $S_i = c^\dagger_{i\downarrow}c_{i\uparrow}$, $b_i = c_{i\uparrow}c_{i\downarrow}$ and α indicates a sum over nearest neighbors of site i. The explicit expressions for the spectral functions $\sigma^{(j)}_{\alpha\beta}(\mathbf{k})$ are given by

$$\sigma^{(1)}_{\xi\xi}(\mathbf{k}) = \frac{I_1}{2}\left[1 + \frac{g(\mathbf{k})}{2Q(\mathbf{k})}\right], \quad \sigma^{(2)}_{\xi\xi}(\mathbf{k}) = \frac{I_1}{2}\left[1 - \frac{g(\mathbf{k})}{2Q(\mathbf{k})}\right],$$

$$\sigma^{(1)}_{\xi\eta}(\mathbf{k}) = \frac{m(\mathbf{k})}{2Q(\mathbf{k})}, \quad \sigma^{(2)}_{\xi\eta}(\mathbf{k}) = -\frac{m(\mathbf{k})}{2Q(\mathbf{k})},$$

$$\sigma^{(1)}_{\eta\eta}(\mathbf{k}) = \frac{I_2}{2}\left[1 - \frac{g(\mathbf{k})}{2Q(\mathbf{k})}\right], \quad \sigma^{(2)}_{\eta\eta}(\mathbf{k}) = \frac{I_2}{2}\left[1 + \frac{g(\mathbf{k})}{2Q(\mathbf{k})}\right]. \quad \text{(E4)}$$

Note the fact that the spectral functions are $\mathbf{k}$ dependent and they also depend on the doping level, through I_1 and I_2, and temperature, due to the self-consistent parameters e and p. At half-filling, the spectral function for the lower Hubbard band determines the amplitudes in $|\mathrm{MI}\rangle$ through $\sigma^1 = (u_k + v_k)^2/2$. In the strong-coupling limit, $U \gg t$, the only dependence that remains is on filling. In this limit $\sigma^{(1)}_{\xi\xi} = I_1$, $\sigma^{(2)}_{\eta\eta} = I_2$ and all the other functions vanish. All of our calculations of the Hall coefficient were performed in this limit.

Consider now the two weak-coupling limits in which Fermi-liquid theory should hold: Case (a) $n \to 0$ and Case (b) $U \to 0$. In the first case, $g \to -U + 2m/n$, $Q \to |g|/2$, the correlations in p become independent and hence $p \propto n^2$, $e \propto n$ implying that $m(\mathbf{k}) \propto n$. Consequently, the dispersion for the lower Hubbard band reduces exactly to that of the noninteracting limit, $\epsilon_1(\mathbf{k}) = -\mu - 4t\alpha(\mathbf{k}) = \epsilon_0(\mathbf{k})$. Moreover, all the spectral weight resides in this band because as $n \to 0$, $g/2Q = 1$ and $m(\mathbf{k}) \to 0$, implying that $\sigma^{(1)} = 1$ and $\sigma^{(2)} = 0$. That the static approximation reduces to the correct noninteracting limit is not unexpected as Figs. 1 and 2 illustrate that the Fermi surface is independent of U as $n \to 0$. In the $U \to 0$ limit, $g(\mathbf{k}) \to (1-n)m(\mathbf{k})/I_1I_2$ and as a consequence, $Q = |m(\mathbf{k})|/2I_1I_2$. As a result, the band dispersion relations are $\epsilon_{1,2} = \epsilon_0(\mathbf{k}) - [m(\mathbf{k}) \mp |m(\mathbf{k})|]/2I_1I_2$ with spectral weights $\sigma^{(1,2)} = 1/2 \pm m(\mathbf{k})/[2|m(\mathbf{k})|]$ which are either unity or zero. Consequently, although two bands still exist, only the free particle dispersion carries unit spectral weight because the $|m(\mathbf{k})|$ terms enter with opposite signs. Hence, the static approximation correctly reproduces the noninteracting limit when $U \to 0$.[102] As a consequence, the violation of Luttinger's theorem seen here is not an artifact of the approximation scheme but stems fundamentally from the splitting of the spectral weight over two bands although no symmetries are broken, the hallmark of Mottness.

[1] N.F. Mott, *Metal-Insulator Transitions* (Taylor & Francis, London, 1974).
[2] P.W. Anderson, in *Lectures on the Many Body Problem*, edited by E.R. Caianiello (Academic, New York, 1964), Vol. 2, p. 127.
[3] S. Doniach, Phys. Rev. B **24**, 5063 (1981); E. Simanek, *ibid.* **22**, 459 (1979).
[4] P. Phillips, *Advanced Solid State Physics* (Westview Press, Colorado, 2003), Chap. 13.
[5] M. Greiner, O. Mandel, T. Esslinger, T.W. Hänsch, and I. Bloch, Nature (London) **415**, 39 (2002).
[6] R. Laughlin, Adv. Phys. **47**, 943 (1998).
[7] S.C. Zhang, Science **275**, 1089 (1997).
[8] A.P. Kampf and J.R. Schrieffer, Phys. Rev. B **42**, 7967 (1990).
[9] D. Pines, Physica C **282-287**, 273 (1997).
[10] J. Zaanen and O. Gunnarsson, Phys. Rev. B **40**, 7391 (1989).
[11] V.J. Emery and S.A. Kivelson, Physica C **209**, 597 (1993); S.A. Kivelson and V.J. Emery, in Proceedings of the Los Alamos Symposium, 1993, edited by K.S. Bedell, Z. Wang, D.E. Meltzer, A.V. Balatsky, and E. Abrahams (Addison-Wesley, New York, 1994), p. 619.
[12] V.J. Emery, S.A. Kivelson, and J.M. Tranquada, Proc. Natl. Acad. Sci. U.S.A. **96**, 8814 (1999).
[13] A.H. Castro Neto, Phys. Rev. B **64**, 104509 (2001).
[14] I. Martin, G. Ortiz, A.V. Balatsky, and A.R. Bishop, Europhys. Lett. **56**, 849 (2001).
[15] A.H. Castro Neto, Phys. Rev. B **64**, 104509 (2001).
[16] A. Polkovnikov, S. Sachdev, M. Vojta, and E. Demler, Int. J. Mod. Phys. B **16**, 3156 (2002).
[17] J. Zaanen and A.M. Oles, Ann. Phys. (Leipzig) **5**, 224 (1996).
[18] S.R. White and D.J. Scalapino, Phys. Rev. Lett. **81**, 3227 (1998).
[19] D.-H. Lee, S.A. Kivelson, Phys. Rev. B **67**, 024506 (2003).
[20] S. Sachdev, Ann. Phys. (N.Y.) **303**, 226 (2003).
[21] P.W. Anderson, *The Theory of Superconductivity in the High-T_c Cuprates* (Princeton University Press, Princeton, 1997).
[22] B. Baskaran, Z. Zou, and P.W. Anderson, Solid State Commun. **63**, 973 (1987).
[23] S.A. Kivelson, D.S. Rokhsar, and J.P. Sethna, Phys. Rev. B **35**, 8865 (1987).
[24] N. Read and B. Chakrabortyy, Phys. Rev. B **40**, 7133 (1989); N. Read and S. Sachdev, Phys. Rev. Lett. **66**, 1773 (1991).
[25] C. Mudry and E. Fradkin, Phys. Rev. B **49**, 5200 (1992).
[26] T. Senthil and M.P.A. Fisher, Phys. Rev. B **62**, 7850 (2000).
[27] R. Moessner and S.L. Sondhi, Phys. Rev. Lett. **86**, 1881 (2001).
[28] X.-G. Wen, Phys. Rev. B **44**, 2664 (1991); **65**, 165113 (2002).
[29] S.L. Cooper *et al.*, Phys. Rev. B **41**, 11 605 (1990).
[30] S. Uchida *et al.*, Phys. Rev. B **43**, 7942 (1991).
[31] C.T. Chen *et al.*, Phys. Rev. Lett. **66**, 104 (1991).
[32] N.P. Armitage, F. Ronning, D.H. Lu, C. Kim, A. Damascelli, K.M. Shen, D.L. Feng, H. Eisaki, Z.-X. Shen, P.K. Mang, N.

Kaneko, M. Greven, Y. Onose, Y. Taguchi, and Y. Tokura, Phys. Rev. Lett. **88**, 257001 (2002).
[33] Y. Kohsaka, T. Sasagawa, F. Ronning, T. Yoshida, C. Kim, T. Hanaguri, M. Azuma, M. Takano, Z.-X. Shen, and H. Takagi, J. Phys. Soc. Jpn. **72**, 1018 (2003).
[34] M. Rübhausen, A. Gozar, M.V. Klein, P. Guptasarma, D.G. Hinks, Phys. Rev. B **63**, 224514 (2001).
[35] H.J.A. Molegraaf, C. Presura, D. van der Marel, P.H. Kes, M. Li, Science **295**, 2239 (2002).
[36] A.F. Santander-Syro, R.P. Lobo, N. Bontemps, Z. Konstantinovic, Z.Z. Li, and H. Raffy, Europhys. Lett. **62**, 568 (2003).
[37] M.B.J. Meinders, H. Eskes, and G.A. Sawatzky, Phys. Rev. B **48**, 3916 (1993).
[38] H. Eskes and R. Eder, Phys. Rev. B **54**, 14 226 (1996); H. Eskes, A.M. Oleś, M.B.J. Meinders, and W. Stephan, *ibid.* **50**, 17 980 (1994).
[39] H. Matsumoto and F. Mancini, Phys. Rev. B **55**, 2095 (1997).
[40] G. Kotliar, S.Y. Savrasov, and G. Palsson, Phys. Rev. Lett. **87**, 186401 (2001).
[41] J. Hubbard, Proc. R. Soc. London, Ser. A **276**, 238 (1963).
[42] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
[43] M. Jarrell and T. Pruschke, Phys. Rev. B **49**, 1458 (1993).
[44] E. Dagotto, A. Moreo, F. Ortolani, J. Riera, and D.J. Scalapino, Phys. Rev. Lett. **67**, 1918 (1991).
[45] W.O. Putikka, M.U. Luchini, and R.R.P. Singh, Phys. Rev. Lett. **81**, 2966 (1998).
[46] A. Himeda and M. Ogata, Phys. Rev. Lett. **85**, 4345 (2000).
[47] Th.A. Maier, Th. Pruschke, and M. Jarrell, Phys. Rev. B **66**, 075102 (2002).
[48] C. Gröber, R. Eder, and W. Hanke, Phys. Rev. B **62**, 4336 (2000).
[49] T. Stanescu and P. Phillips, Phys. Rev. B **64**, 235117 (2001).
[50] L.M. Roth, Phys. Rev. **184**, 451 (1969).
[51] Y. Kuramoto, Z. Phys. B: Condens. Matter **53**, 271 (1983); N. Grewe, *ibid.* **53**, 319 (1983).
[52] Th. Pruschke, Z. Phys. B: Condens. Matter **81**, 319 (1990).
[53] F. Mancini, Phys. Lett. A **249**, 231 (1998); for a review see A. Avella, F. Mancini, D. Villani, L. Siurakshina, and V.Yu. Yusahnkhai, Int. J. Mod. Phys. B **12**, 81 (1998).
[54] DF. Woynarovich, J. Phys. C **15**, 85 (1982); **15**, 97 (1982).
[55] D.S. Fisher, G. Kotliar, and G. Moeller, Phys. Rev. B **52**, 17 112 (1995).
[56] W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 4302 (1970).
[57] G. Rietveld, M. Glastra, and D. van der Marel, Physica C **241**, 257 (1995); A. Ino, Phys. Rev. Lett. **79**, 2101 (1997).
[58] N. Harima *et al.*, Phys. Rev. B **64**, 220507 (2001).
[59] N. Harima, A. Fujimori, T. Sugaya, and I. Terasaki, Phys. Rev. B **67**, 172501 (2003).
[60] M.Z. Hassan *et al.*, Science **288**, 1811 (2000).
[61] G. Ritveld, S.J. Collocot, and D. van der Marel, Physica C **241**, 273 (1995).
[62] M.A. van Veenendaal *et al.*, Phys. Rev. B **47**, 446 (1993).
[63] A.L. Chernyshev, S.R. White, A.H. Castro Neto, Phys. Rev. B **65**, 214527 (2002); E.W. Carlson, V.J. Emery, S.A. Kivelson, and D. Orgad, in *The Physics of Conventional and Unconventional Superconductors*, edited by K.H. Bennemann and J.B. Ketterson (Springer-Verlag, Berlin, 2003).
[64] A brief report of the pseudogap result can be found in T.D. Stanescu and P. Phillips, Phys. Rev. Lett. **91**, 017002 (2003).
[65] Th.A. Maier, M. Jarrell, A. Macridin, and C. Slezak, Phys. Rev. Lett. **92**, 027005 (2004).
[66] K. Haule, A. Rosch, and P. Wölfle, Phys. Rev. Lett. **89**, 236402 (2002).
[67] T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).
[68] P.W. Anderson, cond-mat/0108522 (unpublished).
[69] J.L. Tallon, J.W. Loram, J.R. Cooper, C. Panagopoulos, and C. Bernhard, Phys. Rev. B **68**, 180501(R) (2003); J.L. Tallon, J.w. Loram, and C. Panagopoulos, J. Low Temp. Phys. **131**, 387 (2003).
[70] J. Bonca and P. Prelovsek, Phys. Rev. B **67**, 085103 (2003).
[71] A. Kaminski, H.M. Fretwell, J.C. Campuzano, Z. Li, H. Raffy, W.G. Cullen, H. You, C.G. Olson, C.M. Varma, and H. Höchet, Nature (London) **416**, 610 (2002).
[72] V.J. Emery and S.A. Kivelson, Nature (London) **374**, 434 (1995).
[73] M. Randeria and N. Trivedi, J. Phys. Chem. Solids **59**, 1754 (1998).
[74] C.M. Varma, Phys. Rev. Lett. **83**, 3538 (1999).
[75] S.C. Zhang, Science **275**, 1089 (1997).
[76] S. Chakravarty *et al.*, Phys. Rev. B **63**, 094503 (2001); D.A. Ivanov, P.A. Lee, and X.-G. Wen, Phys. Rev. Lett. **84**, 3958 (2000).
[77] J. Mesot, M. Randeria, M.R. Norman, A. Kaminski, H.M. Fretwell, J.C. Campuzano, H. Ding, T. Takeuchi, T. Sato, T. Yokoya, T. Takahashi, I. Chong, T. Terashima, M. Takano, T. Mochiku, and K. Kadowaki, Phys. Rev. B **63**, 224516 (2001).
[78] P.V. Bogdanov, A. Lanzara, X.J. Zhou, S.A. Kellar, D.L. Feng, E.D. Lu, H. Eisaki, J.-I. Shimoyama, K. Kishio, Z. Hussain, and Z.-X. Shen, Phys. Rev. B **64**, 180505 (2001).
[79] T. Yoshida, X.J. Zhou, T. Sasagawa, W.L. Yang, P.V. Bogdanov, A. Lanzara, Z. Hussain, T. Mizokawa, A. Fujimori, H. Eisaki, Z.-X. Shen, T. Kakeshita, and S. Uchida, Phys. Rev. Lett. **91**, 027001 (2003).
[80] M.E. Simon and C. Varma, Phys. Rev. Lett. **89**, 247003 (2002).
[81] F.C. Zhang and T.M. Rice, Phys. Rev. B **37**, 3759 (1988).
[82] N.P. Ong, *Physical Properties of High-Temperature Superconductors II*, edited by D.M. Ginsberg (World Scientific, Singapore, 1990).
[83] H. Takagi, T. Ido, S. Ishibashi, M. Uota, S. Uchida, and Y. Tokura, Phys. Rev. B **40**, 2254 (1989).
[84] K. Tamasaku, T. Ito, H. Takagi, and S. Uchida, Phys. Rev. Lett. **72**, 3088 (1994).
[85] P. W. Anderson, in *Electronic Properties and Mechanisms of High* T_c *Superconductors, Proceedings of the International Workshop*, edited by T. Oguchi, K. Kadowaki, and T. Sasaki (North-Holland, Amsterdam, 1992), pp. 92–181.
[86] J.H. Kim, K. Levin, and A. Auerbach, Phys. Rev. B **39**, 11 633 (1989).
[87] S.A. Trugman, Phys. Rev. Lett. **65**, 500 (1990).
[88] C.A.R. Sa de Melo, Z. Wang, and S. Doniach, Phys. Rev. Lett. **68**, 2078 (1992).
[89] J.M. Luttinger, Phys. Rev. **119**, 1153 (1960). In proving Luttinger's theorem, one must count the poles and zeros of the single-particle Green function.
[90] H.E. Castillo, and C.A. Balseiro, Phys. Rev. Lett. **68**, 121 (1989).
[91] A.G. Rojo, G. Kotliar, and G.S. Canright, Phys. Rev. B **47**, 9140 (1993).
[92] P. Prelovsek, M. Long, T. Markez, and K. Zotos, Phys. Rev. Lett. **83**, 2785 (1999).
[93] F.F. Assaad, and M. Imada, Phys. Rev. Lett. **74**, 3868 (1995).

[94] B.S. Shastry, B.I. Shraiman, and R.R.P. Singh, Phys. Rev. Lett. **70**, 2004 (1993).
[95] G. Baumgärtel, J. Schmalian, and K.-H. Bennemann, Phys. Rev. B **48**, 3983 (1993).
[96] T. Yoshida, X.J. Zhou, M. Nakamura, S.A. Kellar, P.V. Bogdanov, E.D. Lu, A. Lanzara, Z. Hussain, A. Ino, T. Mizokawa, A. Fujimori, H. Eisaki, C. Kim, Z.-X. Shen, T. Kakeshita, and S. Uchida, Phys. Rev. B **63**, 220501 (2001).
[97] T. Yoshida (private communication).
[98] A. Avella, F. Mancini, and D. Villani, Solid State Commun. **108**, 723 (1998).
[99] J.E. Hirsch, Phys. Rev. B **65**, 184502 (2002).
[100] J.M. Ziman, *Electrons and Phonons* (Oxford University Press, London, 1960), pp. 502 & 503; T.P. Beaulac, F.J. Pinski, and P.B. Allen, Phys. Rev. B **23**, 3617 (1981); W.W. Schulz, P.B. Allen, and N. Trivedi, *ibid.* **45**, 10 886 (1992).
[101] A. Avella, M. Mancini, D. Villani, L. Siurakshina, and V. Yu Yushankhai, Int. J. Mod. Phys. B **12**, 81 (1998).
[102] S. Moukouri and M. Jarrell, Phys. Rev. Lett. **87**, 167010 (2001).