Exact integration of the high energy scale in doped Mott insulators

Ting-Pong Choy, Robert G. Leigh, Philip Phillips, and Philip D. Powell

Department of Physics, University of Illinois, 1110 West Green Street, Urbana, Illinois 61801, USA

We expand on our earlier work [R. G. Leigh *et al.*, Phys. Rev. Lett. **99**, 46404 (2007)] in which we constructed the exact low energy theory of a doped Mott insulator by explicitly integrating (rather than projecting) out the degrees of freedom far away from the chemical potential. The exact low energy theory contains degrees of freedom that cannot be obtained from projective schemes. In particular, a charge $\pm 2e$ bosonic field that is not made out of elemental excitations emerges at low energies. Such a field accounts for dynamical spectral weight transfer across the Mott gap. At half-filling, we show that two such excitations emerge which play a crucial role in preserving the Luttinger surface along which the single-particle Green's function vanishes. In addition, the interactions with the bosonic fields defeat the artificial local SU(2) symmetry that is present in the Heisenberg model. We also apply this method to the Anderson-U impurity and show that in addition to the Kondo interaction, bosonic degrees of freedom appear as well. Finally, we show that as a result of the bosonic degree of freedom, the electron at low energies is in a linear superposition of two excitations—one arising from the standard projection into the low energy sector and the other from the binding of a hole and the boson.

I. INTRODUCTION

Low energy theories based on an explicit integration over the degrees of freedom at high energy are the cornerstone[1] for analyzing long wavelength physics of interacting systems. For high-temperature superconductivity in the cuprates, the relevant[2] low energy theory must be constructed for a doped Mott insulator. While no shortage of theories has been proposed,[3–21] none is based on an explicit integration over the degrees of freedom at high energy. The primary difficulties in carrying out such a program appear in the simplest model

$$H_{\text{Hubb}} = -t\sum_{i,j,\sigma} g_{ij}c_{i,\sigma}^\dagger c_{j,\sigma} + U\sum_i c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger c_{i,\downarrow}c_{i,\uparrow} = H_t + H_U \tag{1}$$

applicable to a doped Mott insulator. Here i and j label lattice sites, g_{ij} is equal to one if and only if i and j are nearest neighbors, $c_{i,\sigma}$ annihilates an electron with spin σ on lattice site i, t is the nearest-neighbor hopping matrix element, and U the energy cost when two electrons doubly occupy the same site. The cuprates live in the strongly coupled regime in which the interactions dominate as $t \approx 0.5$ eV and $U \approx 4$ eV. As U is the largest energy scale, it is appropriate to integrate over the fields that generate the physics on the U scale. The operators that correspond to such physics can be written in terms of $\eta_{i,\sigma} = c_{i,\sigma}n_{i,-\sigma}$, noting that $n_{i,\uparrow}n_{i,\downarrow} = \Sigma_\sigma \eta_{i,\sigma}^\dagger \eta_{i,\sigma}/2$. Physically, $\eta_{i,\sigma}$ ($\eta_{i,\sigma}^\dagger$) annihilates (creates) an electron on a doubly (singly) occupied site, hence is associated with the energy scale U. Consequently, the interaction term reduces to a simple quadratic form

$$H_U = U\sum_i n_{i,\uparrow}n_{i,\downarrow} = \frac{U}{2}\sum_{i,\sigma} \eta_{i,\sigma}^\dagger \eta_{i,\sigma}, \tag{2}$$

which would enable an exact integration over the U scale *if* $\eta_{i,\sigma}$ obeyed canonical fermionic or bosonic commutation relations. However, a simple computation gives

$$\{\eta_i, \eta_j^\dagger\} = -\delta_{ij}\frac{1}{2}\sigma^\mu n_\mu = \begin{pmatrix} n_{i,\downarrow} & -c_{i,\downarrow}^\dagger c_{i,\uparrow} \\ -c_{i,\uparrow}^\dagger c_{i,\downarrow} & n_{i,\uparrow} \end{pmatrix}. \tag{3}$$

Hence, standard bosonic or fermionic coherent state representations are of no use in integrating over the fields $\eta_{i,\sigma}$.

The additional problem is spectral weight transfer. When one electron resides on each site (half-filling), a charge gap of order U opens for all U in $d=1$ and provided $U \gg t$ for $d > 1$.[2,22] The band above the gap describes electron motion on singly occupied sites, whereas the band below captures electron motion on empty sites. Such motion is described by $\eta_{i,\sigma}$ and $\xi_{i,\sigma} = c_{i,\sigma}(1 - n_{i,-\sigma})$, respectively. However, unlike the traditional band picture in which electron motion occurs in either the conduction or valence bands, electron spectral weight lives both above and below the Mott gap. This state of affairs is obtained because the electron annihilation operator

$$c_{i,\sigma} = c_{i,\sigma}(1 - n_{i,-\sigma}) + c_{i,\sigma}n_{i,-\sigma} = \xi_{i,\sigma} + \eta_{i,\sigma} \tag{4}$$

can be written as a linear combination of excitations that reside in both bands. That is, unlike the standard band insulator picture, the states lying above the gap are not orthogonal to those below it. As a consequence, adding or removing electrons from a Mott insulator changes the distribution of spectral weight at all energies. In particular, the addition of x holes to a Mott insulator creates at least $2x$ single-particle addition states[23] just above the chemical potential. The deviation from x, as would be the case in a band insulator, is intrinsic to the strong correlations that mediate the Mott insulating state in a half-filled band, thereby distinguishing Mottness from ordering. Each hole reduces the number of

ways of creating a doubly occupied site by one, thereby reducing the spectral weight at high energy. As the x empty sites can be occupied by either spin up or spin down electrons, the $2x$ sum rule is exact[23] in the atomic limit, $U \to \infty$. In the presence of hybridization (with matrix element t), virtual excitations between the lower Hubbard band (LHB) and upper Hubbard band (UHB) increase the loss of spectral weight at high energy thereby leading to a faster than $2x$ growth[23–25] of the low energy spectral weight, a phenomenon confirmed[26–28] widely in the high-temperature copper-oxide superconductors. The hopping-dependent contribution is referred to as dynamical spectral weight transfer.

A true low energy is not exact if it cannot account for *all* low energy degrees of freedom even if they arise from the high energy scale. Hence, the true low energy theory of a doped Mott insulator must preserve the fact that the low-energy spectral weight increases faster than $2x$. In this regard, two approaches are possible: (C1) change the particle statistics so that placing a particle on one site excludes particles of opposite spin or (C2) generate degrees of freedom at low energy so that removal of an electron destroys at least two charge e states. Perturbative methods followed by projection[12–15,17] of the high energy scale as well as slave[19–21] particle techniques all implement (C1). To leading order in t^2/U, the result is the t-J model. The key goal in such approaches is to diagonalize the Hubbard model into sectors with a fixed number of doubly occupied sites. When one performs such a transformation, however, the electron operators must be transformed as well. Although this step is generally ignored,[2–12,14–18] it is crucial because the no double occupancy condition applies only to the transformed fermions not to the bare electrons. In fact, it is the double occupancy in the bare electron basis that generates the dynamical corrections to the $2x$ sum rule. As the relationship between the transformed and bare electrons is nonlinear, it is advantageous to devise a much simpler method in which the mixing to the doubly occupied sectors in the bare electrons is carried by a single degree of freedom. The current method provides a solution to this problem. All the physics associated with the mixing between the UV and IR scales is captured by a charge $2e$ bosonic field. In addition, one might entertain the possibility that slaved-particle methods[19–21] could be tailored to implement an integration of the high energy scale. However, in the slaved operator approach, the interactions involving double occupancy are highly nonlinear as a result of the constraints that remove the unphysical states and hence double occupancy cannot be integrated over explicitly.

We present here a detailed description of a method that permits an explicit integration of the degrees of freedom far away from the chemical potential in doped Mott insulators. We show that the degrees of freedom far away from the chemical potential can be integrated out explicitly (that is, without resorting to projection or slave particles) for a doped Mott insulator. The result is that physics emerges at low energies, namely, a charge $2e$ boson, that cannot be thought of as simply related to electronic motion. Our work lays plain that the true low energy theory of the Hubbard model is not the t-J-like model as is commonly believed. The true low energy theory is an example of (C2).

The charge $2e$ boson enters the theory as a Lagrange multiplier field. As such, it does not have dynamics, in much the same way that the σ field in nonlinear σ models does not have dynamics. In that case, dynamics are generated radiatively, by taking into account interactions with the other fields in the model. In the simplest spherical case, the latter fields can be completely integrated out, and a large N expansion organizes the theory. The situation for Mottness is considerably more complicated. In particular, we have not yet elucidated the precise low energy dynamics. Instead, we take the appearance of the $2e$ boson as an indication that the building blocks for the low energy dynamics of strongly correlated electron matter involve degrees of freedom that lack electron quantum numbers. As we show in the present paper, and more fully in companion papers,[29,30] there are indications that the boson should not be thought of as a weakly interacting dynamical field at low energies with, for example, a Fock space of its own, but that instead it should be thought of as a constituent in a strongly coupled theory. For example, from the exact form of the electron creation operator at low energy, we deduce that the boson can mediate charge e excitations by binding a hole. It is the emergence of this state at low energies that serves to preserve the $2x$ sum rule.[23] We believe that there are analogies here between the presence of such composite states and confining dynamics in particle physics. Indeed, the nature of an insulator is of course that electric transport is absent, in analogy to the absence of color transport in QCD.

This work expands considerably our previous paper[30] in which we presented only an outline of the method.

II. LOW ENERGY THEORY

We will be concerned with the limit in which the Hubbard bands are well separated, $U \gg t$. Given that the chemical potential lies in the gap between such well separated bands at half-filling, which band we should associate with high energy is ambiguous at half-filling. Both double occupancy (UHB) and double holes (LHB) are equally costly. Doping removes this ambiguity. Hole doping jumps the chemical potential to the top of the LHB thereby defining double occupancy to be the high energy scale. For electron doping, the chemical potential lies at the bottom of the upper-Hubbard band and it is the physics associated with double holes in the lower-Hubbard band that must be coarse grained. At half-filling, both the UHB and LHB must be integrated out. As each of these limits results in a different theory, we will present each separately. As will be seen, the low energy theories that result from the electron- and hole-doped cases are related, though not by the naive particle-hole transformation.

A. Hole doping

Within the Hilbert space for the Hubbard model, $\otimes_i(\mathcal{F}_\uparrow \otimes \mathcal{F}_\downarrow)$, it is impossible to integrate out the degrees of freedom far away from the chemical potential. The basic idea of our construction is to rewrite the Hubbard model in such a way as to isolate the high energy degrees of freedom so that they can be simply integrated out. To solve this problem, it is
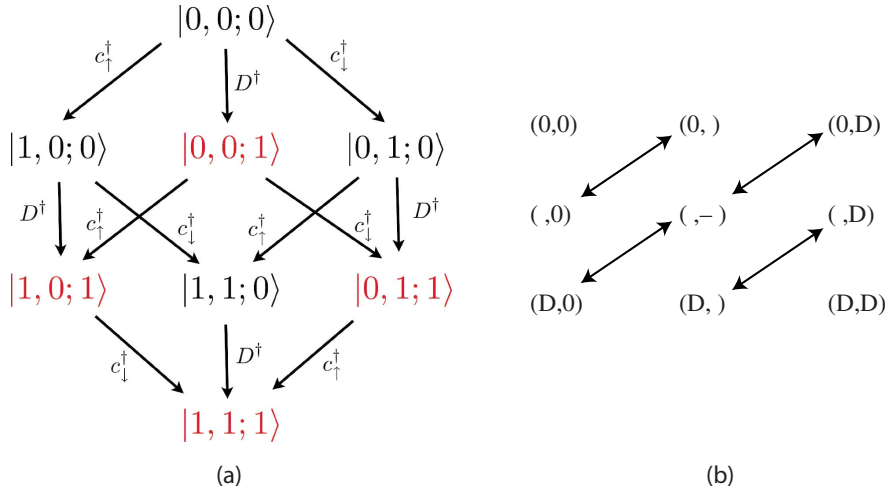
FIG. 1. (Color online) (a) Extended Hilbert space for a single site. (b) Hopping processes between neighboring sites included in the Lagrangian. Double occupation has been replaced by D occupation.

expedient to extend the Hilbert space $\otimes_i (\mathcal{F}_\uparrow \otimes \mathcal{F}_\downarrow \otimes \mathcal{F}_D)$. The key idea is to associate $D^\dagger$ with the creation of double occupation, to be implemented by a constraint. In order to limit the Hilbert space to single occupation in the D sector, we will take D to be fermionic. We refer to D as a fermionic oscillator as it is associated with a two state system. The field D will enter the theory as an elemental field with a large (order U) quadratic term and precise interactions with the electronic degrees of freedom; the low energy (IR) theory is obtained by integrating out D. The interactions of this extended model must be chosen so that the model is precisely equivalent to the Hubbard model; indeed, if instead of integrating out the field D, we merely solve the aforementioned constraint, the model will reduce to the Hubbard model, which we will refer to as the high energy (UV) theory.

The action of the standard electron creation operator, $c_{i,\sigma}^\dagger$, and the fermionic operator, $D^\dagger$, to create the allowed states on a single site is shown in Fig. 1.

There are of course several unphysical states in this Hilbert space. As we will see, such states are removed once the constraint is solved. At present, the expansion of the Hilbert space should be thought of as a tool to enable the integration of the high energy degrees of freedom. To proceed, we formulate a Lagrangian in the extended Hilbert space. The allowed hops involving the D fields and the electron operators which are equivalent to the hops in the Hubbard model are indicated in Fig. 1(b). For example, the hopping process in the upper left-hand corner describes the hopping of a hole in the lower-Hubbard band. The terms in the middle describe the transport between the D field and two electrons in a singlet on neighboring sites. The term in the lower right corner describes a hop in which D_i and $c_j^\dagger$ switch places. There are no further allowable hopping processes. A further requirement of the Lagrangian for the hole-doped theory is that it contains the appropriate dynamical term for motion in the lower-Hubbard band. That is, those sites which contain the occupancy $c_{i,\downarrow}^\dagger c_{i,\uparrow}^\dagger |0\rangle$ must be excluded from hopping processes (such hops are accounted for by the hopping of D). The Euclidean Lagrangian in the extended Hilbert space which describes the hopping processes detailed above can be written as

$$L = \int d^2\theta \left\{ \bar{\theta}\theta \sum_{i,\sigma} (1 - n_{i,-\sigma}) c_{i,\sigma}^\dagger \dot{c}_{i,\sigma} + \sum_i D_i^\dagger \dot{D}_i \right.$$
$$+ U \sum_j D_j^\dagger D_j - t \sum_{i,j,\sigma} g_{ij} [C_{ij,\sigma} c_{i,\sigma}^\dagger c_{j,\sigma} + D_i^\dagger c_{j,\sigma}^\dagger c_{i,\sigma} D_j$$
$$\left. + (D_j^\dagger \theta c_{i,\sigma} V_\sigma c_{j,-\sigma} + \text{h.c.})] + H_{\text{con}} \right\}. \tag{5}$$

Here, g_{ij} selects out nearest neighbors (note that if we wanted to include next-to-nearest neighbor interactions, we need just modify the matrix g_{ij} accordingly), the parameter V_σ has values $V_\uparrow = 1$, $V_\downarrow = -1$, and simply ensures that D couples to the spin singlet and the operator $C_{ij,\sigma}$ is of the form $C_{ij,\sigma} \equiv \bar{\theta}\theta \alpha_{ij,\sigma} \equiv \bar{\theta}\theta(1 - n_{i,-\sigma})(1 - n_{j,-\sigma})$ with number operators $n_{i,\sigma} \equiv c_{i,\sigma}^\dagger c_{i,\sigma}$.

For simplicity, we have introduced a complex Grassman constant θ, which we have inserted in order to keep track of statistics; it bears some resemblance to a superspace coordinate. Because D_j is fermionic and $c_{j,\uparrow} c_{j,\downarrow}$ transforms as a boson, a Grassman variable is needed to essentially "fermionize" double occupancy. They are normalized via

$$\int d^2\theta \, \bar{\theta}\theta = 1. \tag{6}$$

The Grassmann variable is an artificial device that will disappear from the UV or IR Lagrangians.

The constraint Hamiltonian H_{con} is taken to be

$$H_{\text{con}} = s\bar{\theta} \sum_j \varphi_j^\dagger (D_j - \theta c_{j,\uparrow} c_{j,\downarrow}) + \text{h.c.}, \tag{7}$$

where φ is a complex charge $2e$ bosonic field which enters the theory as a Lagrange multiplier. The constant s has been inserted to carry the units of energy. At this point, there is some ambiguity in the normalization of φ, but we expect that this will be set dynamically. We will find that if a true infrared limit exists, then s must be of order of the hopping matrix element t. There is a natural parallel between the constraint condition, Eq. (7), and the constraint in the nonlinear sigma model. In fact, the auxiliary field φ will enter the low energy theory in an analogous fashion to σ in the nonlinear sigma

model. In both cases, φ and σ enter as Lagrange multipliers. Both end up playing a crucial role in the phase structure of the true low energy or infrared limit. In this case, φ will serve to create excitations at low energy which will generate the dynamical part of the spectral weight transfer across the Mott gap.

Now, as remarked previously, we have chosen the Lagrangian (5) so that this theory is equivalent to the Hubbard model. To demonstrate this, we first show that once the constraint is solved, we obtain the Hubbard model. Hence, the Lagrangian we have formulated is the Hubbard model written in a nontraditional form—in some sense, we have inserted unity into the Hubbard model path integral in a rather complicated fashion. To this end, we compute the partition function

$$Z = \int [\mathcal{D}c\mathcal{D}c^\dagger \mathcal{D}D\mathcal{D}D^\dagger \mathcal{D}\varphi \mathcal{D}\varphi^\dagger] \exp^{-\int_0^\tau Ldt}, \quad (8)$$

with L given by Eq. (5). We note that φ is a Lagrange multiplier. As shown in the Appendix [Eq. (A5)], in the Euclidean signature, the fluctuations of the real and imaginary parts of φ_i must be integrated along the imaginary axis for φ_i to be regarded as a Lagrangian multiplier. The φ integrations (over the real and imaginary parts) are precisely a representation of (a series of) δ functions of the form

$$\delta\left(\int d\theta D_i - \int d\theta\bar{\theta}c_{i,\uparrow}c_{i,\downarrow}\right). \quad (9)$$

If we wish to recover the Hubbard model, we need only to integrate over D_i, which is straightforward because of the δ functions. The dynamical terms yield

$$\int d^2\theta\bar{\theta}\theta\left[\sum_{i,\sigma}(1-n_{i,-\sigma})c_{i,\sigma}^\dagger \dot{c}_{i,\sigma}\right.$$
$$\left. + \sum_i c_{i,\downarrow}^\dagger c_{i,\uparrow}^\dagger \dot{c}_{i,\uparrow}c_{i,\downarrow} + \sum_i c_{i,\downarrow}^\dagger c_{i,\uparrow}^\dagger c_{i,\uparrow}\dot{c}_{i,\downarrow}\right]$$
$$= \int d^2\theta\bar{\theta}\theta\sum_{i,\sigma}[(1-n_{i,-\sigma})c_{i,\sigma}^\dagger \dot{c}_{i,\sigma} + n_{i,-\sigma}c_{i,\sigma}^\dagger \dot{c}_{i,\sigma}]$$
$$= \int d^2\theta\bar{\theta}\theta\sum_{i,\sigma}c_{i,\sigma}^\dagger \dot{c}_{i,\sigma}. \quad (10)$$

Likewise the term proportional to V_σ yields

$$\int d^2\theta\bar{\theta}\theta\sum_{i,j}g_{ij}[c_{j,\downarrow}^\dagger c_{j,\uparrow}^\dagger(c_{i,\uparrow}c_{j,\downarrow} - c_{i,\downarrow}c_{j,\uparrow})] + \text{h.c.}$$
$$= \int d^2\theta\bar{\theta}\theta\sum_{i,j,\sigma}g_{ij}n_{j,-\sigma}c_{j,\sigma}^\dagger c_{i,\sigma} + \text{h.c.} \quad (11)$$

Finally, the hopping terms that involve two D fields give rise to

$$\int d^2\theta\bar{\theta}\theta\sum_{i,j}g_{ij}[c_{i,\downarrow}^\dagger c_{i,\uparrow}^\dagger(c_{j,\uparrow}^\dagger c_{i,\uparrow} + c_{j,\downarrow}^\dagger c_{i,\downarrow})c_{j,\uparrow}c_{j,\downarrow}]$$
$$= -\int d^2\theta\bar{\theta}\theta\sum_{i,j}g_{ij}n_{j,-\sigma}n_{i,-\sigma}c_{i,\sigma}^\dagger c_{j,\sigma}. \quad (12)$$

Equations (11) and (12) add to the constrained hopping term in the Lagrangian (the term proportional to $C_{ij,\sigma}$) to yield the standard kinetic energy term in the Hubbard model. Finally, the $D^\dagger D$ term generates the on-site repulsion of the Hubbard model. Consequently, by integrating over φ_i followed by an integration over D_i, we recover the Lagrangian,

$$\int d^2\theta\bar{\theta}\theta L_{\text{Hubb}} = \sum_{i,\sigma}c_{i,\sigma}^\dagger \dot{c}_{i,\sigma} + H_{\text{Hubb}}, \quad (13)$$

of the Hubbard model. This constitutes the ultraviolet (UV) limit of our theory. In this limit, it is clear that the Grassman variables amount to an insertion of unity and hence play no role. Furthermore, in this limit the extended Hilbert space contracts, unphysical states such as $|1, 0, 1\rangle$, $|0, 1, 1\rangle$, and $|1, 1, 1\rangle$ are set to zero, and we identify $|1, 1, 0\rangle$ with $|0, 0, 1\rangle$. Note that there is no contradiction between treating D as fermionic and the constraint in Eq. (7). The constraint never governs the commutation relation for D. The value of D is determined by Eq. (7) only when φ is integrated over. This is followed immediately by an integration over D at which point D is eliminated from the theory.

The advantage of our starting Lagrangian over the traditional writing of the Hubbard model is that we are able to coarse grain the system cleanly for $U \gg t$. The energy scale associated with D is the large on-site energy U. Hence, it makes sense, instead of solving the constraint, to integrate out D. The resultant theory will contain explicitly the bosonic field, φ. As a result of this field, double occupancy will remain, though the energy cost will be shifted from the UV to the infrared (IR). Because the theory is Gaussian, the integration over D_i can be done exactly. This is the ultimate utility of the expansion of the Hilbert space—we have isolated the high energy physics into this Gaussian field. As a result of the dynamical term in the action, integration over D will yield a theory that is frequency dependent. The frequency will enter in the combination $\omega + U$ which will appear in denominators. Since U is the largest energy scale, we expand in powers of ω/U; the leading term yields the proper $\omega = 0$ low energy theory. Since the theory is Gaussian, it suffices to complete the square in the D field. To accomplish this, we define the matrix

$$\mathcal{M}_{ij} = \delta_{ij} - \frac{t}{(\omega + U)}g_{ij}\sum_\sigma c_{j,\sigma}^\dagger c_{i,\sigma} \quad (14)$$

and $b_i = \sum_j b_{ij} = \sum_{j,\sigma}g_{ij}c_{j,\sigma}V_\sigma c_{i,-\sigma}$. At zero frequency the Hamiltonian is

$$H_h^{\text{IR}} = -t\sum_{i,j,\sigma}g_{ij}\alpha_{ij,\sigma}c_{i,\sigma}^\dagger c_{j,\sigma} + H_{\text{int}} - \frac{1}{\beta}\text{Tr}\ln\mathcal{M},$$

where

$$H_{\text{int}} = -\frac{t^2}{U} \sum_{j,k} b_j^\dagger (\mathcal{M}^{-1})_{jk} b_k - \frac{s^2}{U} \sum_{i,j} \varphi_i^\dagger (\mathcal{M}^{-1})_{ij} \varphi_j$$

$$- s \sum_j \varphi_j^\dagger c_{j,\uparrow} c_{j,\downarrow} + \frac{st}{U} \sum_{i,j} \varphi_i^\dagger (\mathcal{M}^{-1})_{ij} b_j + \text{h.c.}, \quad (15)$$

which constitutes the true (IR) limit as long as the energy scale s is not of order U. If $s \sim O(U)$ then we should also integrate out φ_i—this integration is again Gaussian and can be done exactly; one can easily check that this leads precisely back to the UV theory, the Hubbard model. Hence, the only way in which a low energy theory of the Hubbard model exists is if the energy scale for the dynamics that φ mediates is $O(t)$. This observation is significant because it lays plain the principal condition for the existence of an IR limit of the Hubbard model.

To fix s, we note that the fourth term entering our Hamiltonian can mediate spin exchange. As the energy scale for this process is t^2/U, we make the identification $s \simeq t$. Hence, the low energy theory contains a charge $2e$ bosonic field which can either annihilate and/or create doubly occupied sites or nearest-neighbor singlets. That the energy cost for double occupancy in the IR is t^2/U and not U underscores the fact that the UHB and LHB are not orthogonal. The presence of the field φ_i at low energies is the result of the overlap between the high- and low energy scales. Physically, double occupancy occurs at low energies for two distinct reasons. The first is spin exchange which is generated by the term $\varphi_i b_i^\dagger$. The second is motion of a doubly occupied site (a doublon) entirely in the LHB. The latter is not present in projective models but is certainly a low energy process that must be present in the exact low energy theory.

While electron number conservation is broken in the IR, we find [by inspection of Eq. (15)] that a conserved low energy charge does exist[31]

$$Q = \sum_{i,\sigma} c_{i,\sigma}^\dagger c_{i,\sigma} + 2 \sum_i \varphi_i^\dagger \varphi_i. \quad (16)$$

As Eq. (15) makes clear, bosons acquire dynamics only through electron motion. Hence, the low energy theory of a hole-doped Mott insulator is a strongly coupled Bose-Fermi model. On purely phenomenological grounds, Bose-Fermi models have been advanced[32,33] as a starting point for tackling the cuprate problem. In such models, and others,[15] the bosons are viewed as noninteracting and possess a Fock space of their own. However, the current analysis lays plain that while the bosonic degree of freedom exists, it does not extend the Hilbert space of the Hubbard model. That is, the charge $2e$ boson does not have a Fock space of its own. In obtaining the low energy theory, we integrated over the high energy D field which acted in the extended Hilbert space. Consequently, the resultant low energy theory preserves the original Hilbert space of the Hubbard model. As we will show, a distinct possibility is that the boson acts to create composite excitations that have charge e.

Several limits are of interest. First, consider the limit $U = \infty$ (for fixed lattice size). The theory reduces to the restricted hopping term and the third term in Eq. (15). In this limit, the φ integration reduces to a delta function, $\delta(c_{i,\uparrow} c_{i,\downarrow})$,

giving a constraint enforcing the vanishing of double occupancy, the correct result for $U = \infty$.

Second, should $\varphi = 0$, we recover the interactions in the t-J model. To establish this, we note that for $\varphi_i = 0$, we have the restricted hopping term and the first term in Eq. (15). Approximating $\mathcal{M}_{ij}$ by its leading term, δ_{ij}, the second term reduces to

$$\sum_i b_i^\dagger b_i = \sum_{ij\ell\sigma\sigma'} g_{ij} g_{\ell i} c_{i,-\sigma}^\dagger V_\sigma c_{j,\sigma}^\dagger c_{\ell,\sigma'} V_{\sigma'} c_{i,-\sigma'}, \quad (17)$$

which contains the spin-spin interaction $-(\mathbf{S}_i \cdot \mathbf{S}_j - n_i n_j/4)$ as well as the three-site hopping term. Next, we expand the Tr ln term

$$\text{Tr} \ln \mathcal{M}^{-1} = \text{Tr} \ln(\delta_{ij} + A_{ij}) = A_{ii} + \frac{1}{2} A_{ij} A_{ji} + \cdots$$

$$= \frac{t^2}{(\omega + U)^2} \sum_{\sigma,\sigma'} \sum_{i,j} g_{ij} g_{ji} c_{j,\sigma}^\dagger c_{i,\sigma} c_{i,\sigma'}^\dagger c_{j,\sigma'} + \cdots,$$

where

$$A_{ij} = \frac{t}{\omega + U} \sum_\sigma g_{ij} c_{j,\sigma}^\dagger c_{i,\sigma} \quad (18)$$

is nonzero only if i and j are nearest neighbors. When this term appears in the Euclidean Lagrangian, its magnitude is $k_B T t^2/(\omega + U)^2$. Therefore, at low temperature, this term is small compared to H_{IR} and to leading order in t^2/U, the terms in H_{IR} dominate. Hence, the $\varphi = 0$ limit contains the interactions in the t-J model, thereby establishing that the physics contained in φ_i is nonprojective. To make closer contact with the t-J model in which the spin-spin interaction acts only in the singly occupied sector, we note that the theory we have developed here could have been formulated strictly in the projected space by simply substituting $\xi_{i,\sigma}$ for $c_{i,\sigma}$ in the hopping terms containing D_i in our starting Lagrangian, Eq. (5). The only substantive difference would be that the second hopping term (the term quadratic in D_i) in Eq. (5) would enter with the opposite sign. Hence, in the IR limit, $\mathcal{M}_{ij} \rightarrow \delta_{ij} + t/U \xi_{j,\sigma}^\dagger \xi_{i,\sigma}$. This change is dictated by the commutation relations of the $\xi_{i,\sigma}$ operators. The UV limit, the Hubbard model, is obtained as before. Setting $\varphi_i = 0$ in the IR limit leads exactly to the t-J model. Thus, the t-J model[16] in terms of the bare electron operators is not the low-energy limit of the Hubbard model. This is not entirely surprising as Eskes and coworkers[25] have stressed that the operators must be transformed as well in writing the t-J model. Only at $U = \infty$ do the transformed and bare fermion operators agree. This difference signifies that the low energy physics is determined by a finite length scale for encountering double occupancy. Hence the limits $U \rightarrow \infty$ and $L \rightarrow \infty$ (L the size of the system) do not commute as is required for a hard projective model (no double occupancy in the original fermion basis) to be the true low energy theory of the Hubbard model.

No such problem besets our low energy theory. We can recover the original Hubbard model from our low energy theory by simply integrating over φ_i. Although this is not a sensible thing to do from a low energy perspective, it can be done exactly. To see how this happens, we rewrite Eq. (15)

including the frequency dependence,

$$
L_{\text{IR}} = \sum_{i,\sigma}(1-n_{i\bar{\sigma}})c_{i,\sigma}^{\dagger}\dot{c}_{i,\sigma} - t\sum_{i,\sigma}(1-n_{i\bar{\sigma}})c_{i,\sigma}^{\dagger}c_{j,\sigma}(1-n_{j\bar{\sigma}}) - \sum_{ij}\left[s\varphi_i^{\dagger} - t\sum_{\langle i,k\rangle}b_{ki}^{\dagger} + (\omega+\hat{U})c_{k\downarrow}^{\dagger}c_{k\uparrow}^{\dagger}\mathcal{M}_{ki}\right]\frac{(\mathcal{M}^{-1})_{ij}}{\omega+\hat{U}}
$$

$$
\times\left[s\varphi_j - t\sum_{\langle j,n\rangle}b_{nj} + (\omega+\hat{U})\mathcal{M}_{jn}c_{n\uparrow}c_{n\downarrow}\right] - tc_{i,\downarrow}^{\dagger}c_{i,\uparrow}^{\dagger}b_i - tb_i^{\dagger}c_{i,\uparrow}c_{i,\downarrow} + (\omega+\hat{U})c_{i,\downarrow}^{\dagger}c_{i,\uparrow}^{\dagger}\mathcal{M}_{ij}c_{j,\uparrow}c_{j,\downarrow} - \frac{1}{\beta}\operatorname{Tr}\ln\mathcal{M}.
$$

To simplify the expression, note that

$$
-tb_i^{\dagger}c_{i,\uparrow}c_{i,\downarrow} = -t\sum_j(c_{i,\downarrow}^{\dagger}c_{j,\uparrow}^{\dagger} - c_{i,\uparrow}^{\dagger}c_{j,\downarrow}^{\dagger})c_{i,\uparrow}c_{i,\downarrow} = -t\sum_{\sigma}c_{j,\sigma}^{\dagger}c_{i,\sigma}n_{i,-\sigma}
\tag{19}
$$

and

$$
-\sum_{\langle i,j\rangle\sigma}c_{i,\downarrow}^{\dagger}c_{i,\uparrow}^{\dagger}c_{j,\sigma}^{\dagger}c_{i,\sigma}c_{j,\uparrow}c_{j,\downarrow} = \sum_{\langle i,j\rangle,\sigma}n_{i,-\sigma}c_{i,\sigma}^{\dagger}c_{j,\sigma}n_{j,-\sigma},
\tag{20}
$$

which comes from $c_{i,\downarrow}^{\dagger}c_{i,\uparrow}^{\dagger}\mathcal{M}_{ij}c_{j,\uparrow}c_{j,\downarrow}$. The Lagrangian becomes

$$
L = \sum_{i,\sigma}c_{i,\sigma}^{\dagger}\dot{c}_{i,\sigma} + H_{\text{Hubb}} - \sum_{ij}\left[s\varphi_i^{\dagger} - t\sum_{\langle i,k\rangle}b_{ki}^{\dagger} + (\omega+\hat{U})c_{k\downarrow}^{\dagger}c_{k\uparrow}^{\dagger}\mathcal{M}_{ki}\right]\frac{(\mathcal{M}^{-1})_{ij}}{\omega+U}\left[s\varphi_j - t\sum_{\langle j,n\rangle}b_{nj} + (\omega+\hat{U})\mathcal{M}_{jn}c_{n\uparrow}c_{n\downarrow}\right] - \frac{1}{\beta}\operatorname{Tr}\ln\mathcal{M},
\tag{21}
$$

which yields the Hubbard model upon integration over φ_i (see the Appendix for details).

We conclude from this analysis that our low energy theory permits immediate correspondence with the original Hubbard model; that is, we have not lost any information regarding the high energy scale, unlike projective methods. All information regarding the high energy scale is encoded into the emergent charge $2e$ bosonic excitation and its interactions. The IR physics will be determined by examining the low energy dynamics of the electrons and/or holes and φ.

B. What this theory is not

Rather than decoupling the on-site repulsion term, we derived our low energy theory by exponentiating a δ-functional constraint on the heavy field, D. Nonetheless, one might contemplate that our theory could be obtained by more traditional schemes, for example, by some sort of Hubbard Stratonovich (HS) decoupling scheme. Since the interaction in the Hubbard model is entirely local, any decoupling by means of introducing an auxiliary field would yield only local interactions. The auxiliary field φ_i in Eq. (15) clearly generates nonlocal interactions as well as on-site interactions. Hence, Eq. (15) cannot be obtained from a HS transformation. However, the nonlocal terms are scaled by t/U. Hence, it might still be maintained that the local terms dominate and, in fact, that they could be obtained by some sort of HS transformation. Consider the identity,

$$
e^{-UX^{\dagger}X} = \int d^2\sigma\, e^{-(\lambda\sigma^{\dagger}\sigma+\lambda_1\sigma X^{\dagger}+\lambda_2 X\sigma^{\dagger})},
\tag{22}
$$

which is true as long as $\lambda\in\mathbb{R}^+$ and $\lambda_1\lambda_2=-U\lambda$. The standard HS transformation assumes that $\lambda_1\lambda_2\in\mathbb{R}^+$ and $\lambda_2=\lambda_1^*$. This

necessarily leads only to the $-U$ Hubbard model with $X=c_{\uparrow}c_{\downarrow}$. However, the constraint that the exponent on the right-hand side of Eq. (22) be real can be relaxed[34] in which case it can be applied to the $+U$ Hubbard model as well. Nonetheless, this procedure will not yield the nonlocal terms in our low energy theory. Furthermore, it does not permit a clean identification of the field associated with the high energy degrees of freedom. That is, there is no D field in any version of Eq. (22). Consequently, this procedure is a non-starter for the construction of a proper low energy theory.

C. Electron doping

For electron doping, the chemical potential jumps to the bottom of the UHB. Consequently, the degrees of freedom that lie far away from the chemical potential no longer correspond to double occupancy but rather double holes. To coarse grain these degrees of freedom, we extend the Hilbert space in a similar way as the hole-doped theory, defining a field $\tilde{D}$ which will be constrained to describe the creation of double holes. Mathematically, all results in hole doping obtained in the previous section can be transformed to electron doping via a generalized particle-hole transformation (GPHT), namely,

$$
c_{i,\sigma} \rightarrow e^{i\mathbf{Q}\cdot\mathbf{R}_i}c_{i,\sigma}^{\dagger},
$$

$$
D_i \rightarrow \tilde{D}_i,
$$

$$
\varphi_i \rightarrow \tilde{\varphi}_i,
\tag{23}
$$

where $\mathbf{Q}=(\pi,\pi)$ and $\tilde{D}_i$ is a fermion operator associated with double holes. The bosonic field, $\tilde{\varphi}$, is the Lagrangian

multiplier defined by the constraint, $\tilde{H}_{con} = \tilde{s}\bar{\theta}\theta\Sigma_i\tilde{\varphi}_i(\tilde{D}_i - \theta c^\dagger_{i,\uparrow}c^\dagger_{i,\downarrow})$+h.c. According to Eq. (5), an appropriate Lagrangian for the extended theory at electron doped can be constructed,

$$
L = \int d^2\theta \left\{ \bar{\theta}\theta\sum_{i,\sigma} n_{i,-\sigma}c_{i,\sigma}\dot{c}^\dagger_{i,\sigma} + \sum_i \tilde{D}^\dagger_i\dot{\tilde{D}}_i + U\sum_j \tilde{D}^\dagger_j\tilde{D}_j \right.
$$
$$
- t\sum_{i,j,\sigma} g_{ij}[c^\dagger_{i,\sigma}n_{i,-\sigma}c_{j,\sigma}n_{j,-\sigma} + \tilde{D}^\dagger_j c^\dagger_{j,\sigma}c_{i,\sigma}\tilde{D}_i
$$
$$
\left. + (\bar{\theta}c_{i,\sigma}V_\sigma c_{j,-\sigma}\tilde{D}_i + \text{h.c.})] + \tilde{H}_{con} \right\},
\tag{24}
$$

which preserves the distinct hops in the Hubbard model. Two differences to note are that (1) because the chemical potential resides in the UHB, the electron hopping term now involves sites that are at least singly occupied and (2) the order of the $\tilde{D}_i$ and c_i operators is important. If we integrate over $\tilde{\varphi}_i$ and then $\tilde{D}_i$, all the unphysical states are removed and we obtain as before precisely L_{Hubb}. Hence, both theories yield the Hubbard model in their UV limits. They differ, however, in the IR as can be seen by performing the integration over $\tilde{D}_i$. The corresponding integral is again Gaussian and yields

$$
H^{\text{IR}}_e = -t\sum_{i,j,\sigma} g_{ij}c^\dagger_{i,\sigma}n_{i,-\sigma}c_{j,\sigma}n_{j,-\sigma} + \tilde{H}_{\text{int}} - \frac{1}{\beta}\text{Tr}\ln\tilde{M},
$$

where

$$
\tilde{H}_{\text{int}} = -\frac{t^2}{U}\sum_{j,k} b_j(\tilde{M}^{-1})_{jk}b^\dagger_k - \frac{\tilde{s}^2}{U}\sum_{i,j} \tilde{\varphi}^\dagger_i(\tilde{M}^{-1})_{ij}\tilde{\varphi}_j
$$
$$
- \tilde{s}\sum_j \tilde{\varphi}^\dagger_j c^\dagger_{j,\uparrow}c^\dagger_{j,\downarrow} + \frac{\tilde{s}t}{U}\sum_{i,j} \tilde{\varphi}^\dagger_i(\tilde{M}^{-1})_{ij}b^\dagger_j + \text{h.c.}
$$

as the IR limit of the electron-doped theory. In Eq. (25), the matrix $\tilde{M}_{ij}=\mathcal{M}_{ji}$ is defined via the GPHT on the $\mathcal{M}$ matrix in Eq. (14). As $\tilde{\varphi}_i$ is a complex field, the GPHT interchanges the creation operators of opposite charge. We again make the identification $\tilde{s}\sim t$ because the last term can also mediate spin exchange.

When the boson vanishes, we do recover the exact particle-hole symmetric analog of the hole-doped theory. Because the field $\tilde{\varphi}$ now couples to double holes, the relevant creation operator has charge $-2e$ and the conserved charge is $\tilde{Q}=\Sigma_{i,\sigma}c^\dagger_{i,\sigma}c_{i,\sigma}-2\varphi^\dagger_i\varphi_i$. This sign change in the conserved charge will manifest itself as a sign change in the chemical potential as long as $\langle\tilde{\varphi}_i\tilde{\varphi}_i\rangle\neq 0$. Likewise, the correct $U\to\infty$ limit is obtained as before.

D. Half-filled models

At half-filling when the chemical potential lies within the Mott gap, both double hole and double occupancy lie far away from the chemical potential. As two different degrees of freedom per site need to be coarse grained, we introduce two fields D and $\tilde{D}$ which when constrained correspond to double occupancy and double holes, respectively, and extend

the Hilbert space to $\otimes_i(\mathcal{F}_\uparrow\otimes\mathcal{F}_\downarrow\otimes\mathcal{F}_D\otimes\mathcal{F}_{\tilde{D}})$. The corresponding low energy theory will be obtained by integrating over both D and $\tilde{D}$ rather than by solving the constraint. As a result, integrating over D and $\tilde{D}$ does not yield identical results as would be the case if double occupancy or double holes were integrated over.

1. Anderson impurity

To illustrate the process of coarse graining two high energy fields, we begin with a simpler model, the Anderson[35] impurity model.

$$
H = E_f(n_\uparrow + n_\downarrow) + Un_\downarrow n_\uparrow + \sum_{k,\sigma} n_{k,\sigma}\varepsilon_k + \sum_k (v_k c^\dagger_{k,\sigma}a_\sigma + \text{h.c.}),
\tag{25}
$$

where a_σ destroys an impurity electron and $c_{k,\sigma}$ destroys a continuum electron. By setting $E_f=-U/2$, it costs an energy $U/2$ to create a double hole or a doubly occupied state on the impurity site which is analogous to the half-filled Hubbard model. In the following, we would like to shown that, by introducing two heavy fields D and $\tilde{D}$ which correspond to the doubly occupied or double hole states on the impurity site, respectively, the Kondo model with additional coupling to bosonic fields can be derived perturbatively if $v_k\ll U$.

The appropriate extended Hamiltonian is

$$
H = \int d^2\theta \left[E_f\sum_\sigma n_\sigma + \frac{1}{2}UD^\dagger D - \frac{1}{2}U\tilde{D}\tilde{D}^\dagger + \bar{\theta}\theta\sum_{,\sigma} n_{k,\sigma}\varepsilon_k \right.
$$
$$
+ \sum_{k,\sigma} (\bar{\theta}c^\dagger_{k,-\sigma}V_\sigma a^\dagger_\sigma D + \bar{D}^\dagger\theta c^\dagger_{k,-\sigma}V_\sigma a^\dagger_\sigma + \text{h.c.})
$$
$$
\left. + s\bar{\theta}\varphi^\dagger(D - \theta a_\uparrow a_\downarrow) + \text{h.c.} + \tilde{s}\bar{\theta}\tilde{\varphi}^\dagger(\tilde{D} - \theta a^\dagger_\uparrow a^\dagger_\downarrow) + \text{h.c.} \right].
\tag{26}
$$

In the current model, two bosonic field φ and $\tilde{\varphi}$ are introduced which correspond to the two constraints on D and $\tilde{D}$ fields, respectively. If we first integrate out φ and $\tilde{\varphi}$ which result in delta functions corresponding to the removal of the unphysical states and then integrate out D and $\tilde{D}$, Eq. (25) is obtained precisely. This constitutes the UV limit of this theory. Similar to the previous discussions, we can derive the IR limit of the Anderson impurity model by first integrating out both D and $\tilde{D}$ fields, which amounts to substituting

$$
D^\dagger = -\frac{2}{U}\bar{\theta}\left(s\varphi^\dagger + \sum_{k,\sigma} v_k c^\dagger_{k,\sigma}V_\sigma a^\dagger_\sigma\right),
\tag{27}
$$

$$
\tilde{D} = -\frac{2}{U}\left(\tilde{s}\varphi + \sum_{k,\sigma} v_k c^\dagger_{k,\sigma}V_\sigma a^\dagger_\sigma\right)\theta
\tag{28}
$$

into H. We finally obtain

$$H_{IR} = H_{\text{Kondo}} - \frac{2}{U}(s^2\varphi^\dagger\varphi + \tilde{s}^2\tilde{\varphi}^\dagger\tilde{\varphi}) - (s\varphi^\dagger + \tilde{s}\tilde{\varphi}^\dagger)a_\uparrow a_\downarrow$$

$$- \frac{2}{U}(s\varphi^\dagger + \tilde{s}\tilde{\varphi})\sum_{\mathbf{k},\sigma} v_{\mathbf{k}} a_\sigma V_\sigma c_{\mathbf{k},\sigma} + \text{h.c.}, \qquad (29)$$

where H_{Kondo} is the Kondo Hamiltonian

$$H_{\text{Kondo}} = \sum_{\mathbf{k},\sigma} n_{\mathbf{k},\sigma}\varepsilon_{\mathbf{k}} - \frac{U}{2}(n_\uparrow + n_\downarrow)$$

$$+ \sum_{\mathbf{k},\mathbf{k}'} \frac{4v_{\mathbf{k}}v_{\mathbf{k}'}}{U}\left[(c_{\mathbf{k}}^\dagger\vec{\sigma}c_{\mathbf{k}'})\cdot\mathbf{S}_{imp} - \frac{1}{2}c_{\mathbf{k}}^\dagger c_{\mathbf{k}'}(n_\uparrow + n_\downarrow)\right]. \qquad (30)$$

Here, $\mathbf{S}_{imp} = a_\sigma^\dagger\vec{\sigma}_{\sigma\sigma'}a_\sigma'$ is the spin operator of the impurity site. Thus the IR limit consists of the Kondo model in addition to the coupling between the electron and the bosonic degree of freedom.

2. Half-filled Hubbard model

Next, we perform the same procedure for the Hubbard model at half-filling by introducing two fermionic fields D and $\tilde{D}$ associated with the double occupancy and double holes, respectively. We consider the generalized Lagrangian

$$L = \int d^2\theta\left[\sum_i (D_i^\dagger\dot{D}_i + \dot{\tilde{D}}_i^\dagger\tilde{D}_i) - \frac{t}{2}\sum_{i,j,\sigma} g_{ij}(D_j^\dagger\theta c_{i,\sigma}V_\sigma c_{j,-\sigma}\right.$$

$$\left. + \bar{\theta}c_{i,\sigma}V_\sigma c_{j,-\sigma}\tilde{D}_j + \text{h.c.}) + \frac{U}{2}\sum_j (D_j^\dagger D_j - \tilde{D}_j\tilde{D}_j^\dagger) + H_{\text{con}}\right], \qquad (31)$$

with the constraint terms given by

$$H_{\text{con}} = s\bar{\theta}\sum_i \varphi_i^\dagger(D_i - \theta c_{i,\uparrow}c_{i,\downarrow}) + \text{h.c.}$$

$$+ \tilde{s}\bar{\theta}\sum_i \tilde{\varphi}_i^\dagger(\tilde{D}_i - \theta c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger) + \text{h.c.} \qquad (32)$$

Here, φ_i and $\tilde{\varphi}_i$ are the two bosonic fields with charge $2e$ and $-2e$, respectively. Similar to the previous result, if we first integrate out both the bosonic fields φ_i and $\tilde{\varphi}_i$ and then D_i and $\tilde{D}_i$, the Hubbard model is obtained and the generalized theory [Eq. (31)] yields the correct UV limit. However, a different IR limit is obtained if we first integrate out D_i and $\tilde{D}_i$,

$$H_{IR}^{\text{hf}} = -\frac{t^2}{2U}\sum_i (b_i^\dagger b_i + b_i b_i^\dagger) - \sum_i \left(\frac{2s^2}{U}\varphi_i^\dagger\varphi_i + \frac{2\tilde{s}^2}{U}\tilde{\varphi}_i^\dagger\tilde{\varphi}_i\right)$$

$$+ \frac{t}{U}\sum_i (s\varphi_i^\dagger + \tilde{s}\tilde{\varphi}_i)b_i + \text{h.c.}$$

$$- \sum_i (s\varphi_i^\dagger - \tilde{s}\tilde{\varphi}_i)c_{i,\uparrow}c_{i,\downarrow} + \text{h.c.} \qquad (33)$$

This Hamiltonian is invariant under the transformation $c_{i,\sigma} \rightarrow \exp(i\mathbf{Q}\cdot\mathbf{R}_i)c_{i,\sigma}^\dagger$, $\varphi_i \leftrightarrow \tilde{\varphi}_i$, and $s \leftrightarrow \tilde{s}$. This invariant reflects

the symmetry between the double occupancy and the double hole in the system at half-filling. In contrast to the doped case as in Eqs. (15) and (25), no $\mathcal{M}$ matrices appear in the IR theory at half-filling. Consequently, we arrive at a closed form for the low energy theory at half-filling. The $b^\dagger b + bb^\dagger$ terms include a spin-spin interaction as well as a three-site hopping term. However, at half-filling, the three-site hopping term vanishes. As a result, charge dynamics appear solely from motion of the charge $2e$ boson. This state of affairs is obtained because at half-filling, charge dynamics persists only on the energy scale U. Since it is the boson that encodes the high energy scale, it stands to reason that only the boson term mediates charge transport. As we show in Appendix A, it is the φ_i terms that break the spurious local SU(2) symmetry[36] of the Heisenberg model and reinstate the global SU(2) symmetry of the Hubbard model. The local SU(2) symmetry in the Heisenberg model arose entirely from projection. This local SU(2) symmetry was noticed quite some time earlier[36] but its origin was never clarified. In fact, it is straightforward to check that in the lower-Hubbard band at half-filling, all perturbative terms in t/U only mediate spin physics and hence preserve the local SU(2) symmetry of the Heisenberg model. Our derivation (see Appendix B) lays plain that as a result of the boson terms, this symmetry is absent from the true low energy theory of the Hubbard model. Note that, in the Anderson impurity models, the bosonic terms play a similar role to destroy the local SU(2) symmetry that appears as a result of projection.

Second, the true low energy model preserves the sum rules associated with the original model. An essential property[37,38] of the half-filled Hubbard model when $U \gg t$ is the presence of a surface in momentum space (the Luttinger surface) where the single-particle Green's function vanishes. Since the Mott state at half-filling has a gap, the nontrivial implication of the zero surface is that the real part of the Green's function,

$$R_\sigma(\mathbf{k},0) = -\int_{-\infty}^{-\Delta_-} d\epsilon'\frac{A_\sigma(\mathbf{k},\epsilon')}{\epsilon'} - \int_{\Delta_+}^\infty d\epsilon'\frac{A_\sigma(\mathbf{k},\epsilon')}{\epsilon'}, \qquad (34)$$

vanishes. Here $A_\sigma(k,\epsilon)$ is the single-particle spectral function which we are assuming to have a gap of width 2Δ symmetrically located about the chemical potential at $\epsilon = 0$. Because $A(\mathbf{k},\epsilon) > 0$ away from the gap, and ϵ changes sign above and below the gap, Eq. (34) can pass through zero. For this state of affairs to be obtained, the pieces of the integral below and above the gap must be retained. Projected models which throw away the UHB fail to recover the zero surface. What Eq. (33) makes clear is that all the information regarding the surface of zeros is now encoded into the bosonic fields $\varphi_i^\dagger$ and $\bar{\varphi}_i$. On the Luttinger surface, the self-energy diverges, representing a breakdown of perturbation theory. As the bosonic fields cannot be obtained from perturbation theory, we conclude that it is the emergence of the bosonic field that accounts for the breakdown of Luttinger's theorem[38] and ultimately the Mott insulating state.

E. Electron operator

In each of the low energy theories, the operator which creates a single electron represents a composite excitation. To determine its form, we add to each of the starting Lagrangians a source term that generates the canonical electron operator when the constraint is solved. The appropriate transformation that yields the canonical electron operator in the UV is

$$L \to L + \sum_{i,\sigma} J_{i,\sigma} [\bar{\theta}\theta(1 - n_{i,-\sigma})c_{i,\sigma}^{\dagger} - V_{\sigma}D_i^{\dagger}\theta c_{i,-\sigma}] + \text{h.c.}$$

However, in the IR in which we only integrate over the heavy degree of freedom, D_i, the electron creation operator

$$c_{i,\sigma}^{\dagger} \to (1 - n_{i,-\sigma})c_{i,\sigma}^{\dagger} - V_{\sigma}\frac{t}{U}b_j^{\dagger}\mathcal{M}_{ji}^{-1}c_{i,-\sigma} + V_{\sigma}\frac{s}{U}\varphi_j^{\dagger}\mathcal{M}_{ji}^{-1}c_{i,-\sigma}.$$

$$(35)$$

For electron doping, we apply the generalized particle-hole transformation to obtain

$$c_{i,\sigma}^{\dagger} \to n_{i,-\sigma}c_{i,\sigma}^{\dagger} - V_{\sigma}\frac{t}{U}c_{i,-\sigma}\mathcal{M}_{ij}^{-1}b_j^{\dagger} + V_{\sigma}\frac{s}{U}c_{i,-\sigma}\mathcal{M}_{ij}^{-1}\tilde{\varphi}_j,$$

$$(36)$$

as the generator of electron excitations in the IR. For either doping, the electron operator contains the standard term for motion in the LHB, $(1 - n_{i,-\sigma})c_{i,\sigma}^{\dagger}$ ($n_{i,-\sigma}c_{i,\sigma}$ in the UHB for electron doping) with a renormalization from spin fluctuations (second term) and a charge e excitation, $\varphi_j^{\dagger}\mathcal{M}_{ji}^{-1}c_{i,-\sigma}$. Consequently, we predict that an electron at low energies is in a superposition of the standard LHB state (modified with spin fluctuations) and a composite charge e state described by $c_{i,-\sigma}\mathcal{M}_{ij}^{-1}\varphi_j^{\dagger}$. The projected state [the first term in Eq. (35)] preserves the static part of the spectral weight transfer (that is, the $2x$ sum rule) whereas the additional charge e state generates the dynamical (hopping dependent) part of the spectral weight transfer across the Mott gap. As shown in a companion paper,[29] there are also experimental ramifications for the composite structure of the electron.

At half-filling, a similar trick can be applied to generate the electron operator. In this case,

$$L \to L + \sum_{i,\sigma} J_{i,\sigma}(V_{\sigma}D_i^{\dagger}c_{i,-\sigma}\theta + V_{\sigma}\bar{\theta}c_{i,-\sigma}\tilde{D}_i) + \text{h.c.}$$

is the correct transformation to generate the canonical electron operator in the UV. If we now integrate the partition function over D_i and $\tilde{D}_i$, we find that the electron creation operator at half-filling

$$c_{i,\sigma}^{\dagger} \to -V_{\sigma}\frac{t}{U}(c_{i,-\sigma}b_i^{\dagger} + b_i^{\dagger}c_{i,-\sigma}) + V_{\sigma}\frac{2}{U}(s\varphi_i^{\dagger} + \tilde{s}\tilde{\varphi}_i)c_{i,-\sigma}$$

has two important differences with its counterpart for $n \neq 1$. First, it lacks the standard LHB and UHB components as the chemical potential lies between both bands at half-filling. Second, the propagator $\mathcal{M}$ is absent. Nonetheless, the electron at half-filling still has two components both above and below the chemical potential. The simplification that $c_{i,-\sigma}\varphi_i^{\dagger}$

(that is, the $\mathcal{M}$ matrix is absent) constitutes the charge e excitation may make subsequent calculations of the strength of the binding between the boson and a hole at least tractable within the framework of the Bethe-Saltpeter equations.

III. FINAL REMARKS

We have shown that a true low energy theory of a doped Mott insulator possesses degrees of freedom which do not have the quantum numbers of the electron. The degree of freedom is a local nonretarded charge $2e$ boson and hence stands in stark contrast to the charge e boson in the slave boson[19–21] approach in which a direct integration of the high energy scale is not possible. Fundamental to theory here is that the boson does not act in its own Fock space, in contrast to other Fermi-Bose models.[32,33] That is, there are no free charge $2e$ boson states just as there are no free quark states in confining theories. Rather, the charge $2e$ boson mediates electronic states by forming composite excitations. As such the charge $2e$ degree of freedom is detectable[29,30] through the substructure it provides in the electron excitation spectrum. In addition, the boson is not minimally coupled to the electromagnetic gauge field. It acquires dynamics and hence a gauge coupling through high order terms in the $\mathcal{M}$ matrix, essentially t^3/U^2. At half-filling, the bosonic mode preserves the Luttinger surface on which the self-energy diverges or equivalently, the single-particle Green's function vanishes. Since the boson represents a nonperturbative effect, it is not surprising that the Luttinger surface cannot occur without it. In a future publication[29] we explore the role of the boson in mediating the normal state properties of the cuprates as well as the possibility that the Mott state is ultimately characterized by charge neutral bound states mediated by the hidden charge $\pm 2e$ boson.

ACKNOWLEDGMENTS

APPENDIX A: LAGRANGE MULTIPLIERS

Here we offer some details about the mechanics of Lagrange multipliers. Though this is standard stuff, we review it to avoid any confusion in our derivation. To illustrate the method, we will begin with the familiar example of the nonlinear σ model, that is, a bosonic theory with spherical target manifold. In Lorentzian signature, we introduce the spherical constraint by writing the corresponding functional δ function as an integral of a complex exponential, with Lagrange multiplier σ,

$$Z_L[J] = \int [d\phi^a d\sigma]e^{i\int d^d x(1/2\phi^a\partial^2\phi^a)}e^{-i/2\int d^d x\sigma(\phi^a\phi^a - N/g)},$$

$$(A1)$$

which after Wick rotation becomes

$$Z_E[J] = \int [d\phi^a d\sigma] e^{-\int d^d x [1/2\phi^a(-\partial^2 + \sigma)\phi^a - N/2g\sigma]} \quad \text{(A2)}$$

$$= \int [d\sigma] e^{-NS_{eff}(\sigma)}, \quad \text{(A3)}$$

with

$$S_{eff}(\sigma) = \frac{1}{2}\text{Tr}\ln(-\partial^2 + \sigma) - \frac{1}{2g}\int_x \sigma. \quad \text{(A4)}$$

In Eq. (A3), we have performed the ϕ^a functional integrations. To proceed with the analysis of the model, we investigate Eq. (A4) by expanding σ around its vev $\langle\sigma\rangle$. It is crucial though to appreciate that if we go back to Eq. (A2), we see that in order for σ to be a Lagrange multiplier field (in the Euclidean formulation), the fluctuations in σ should be taken along the imaginary axis in field space. That is, we write

$$\sigma(x) = \langle\sigma\rangle + \frac{i}{\sqrt{N}}\lambda(x). \quad \text{(A5)}$$

For uniform $\langle\sigma\rangle$, we then obtain

$$\text{Tr}\ln(-\partial^2 + \sigma) = -\text{Tr}\ln\Delta_0 + \frac{i}{\sqrt{N}}\Delta_0(0)\int_x \lambda(x) + \frac{1}{2N}$$

$$\times \int_{x,y} \Delta_0(x,y)\lambda(y)\Delta_0(y,x)\lambda(x) + \cdots, \quad \text{(A6)}$$

where $\Delta_0^{-1} = -\partial^2 + \langle\sigma\rangle$. We thus find

$$S_{eff} = -\frac{\langle\sigma\rangle}{2g}V_d - \frac{1}{2}\text{Tr}\ln\Delta_0 - \frac{i}{2g\sqrt{N}}\int_x \lambda(x)[1 - g\Delta_0(0)]$$

$$+ \frac{1}{4N}\int_{x,y} \Delta_0(x,y)\lambda(y)\Delta_0(y,x)\lambda(x) + \cdots. \quad \text{(A7)}$$

Thus we see that there is a stable saddle point giving the familiar gap equation

$$\frac{1}{g} = \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 + \langle\sigma\rangle}. \quad \text{(A8)}$$

Now, in the theory considered in this paper, we have a similar situation. In the Lorentzian signature, we have

$$Z = \int [\mathcal{D}c\mathcal{D}c^\dagger \mathcal{D}D\mathcal{D}D^\dagger \mathcal{D}\varphi\mathcal{D}\varphi^\dagger] \exp^{i\int d^d x(L_0 + H_{con})}, \quad \text{(A9)}$$

where

$$L_0 = \int d^2\theta \left\{ \bar{\theta}\theta \sum_{i,\sigma} (1 - n_{i,-\sigma})c_{i,\sigma}^\dagger \dot{c}_{i,\sigma} + \sum_i D_i^\dagger \dot{D}_i \right.$$

$$+ U\sum_j D_j^\dagger D_j - t\sum_{i,j,\sigma} g_{ij}[C_{ij,\sigma}c_{i,\sigma}^\dagger c_{j,\sigma} + D_i^\dagger c_{j,\sigma}^\dagger c_{i,\sigma}D_j$$

$$\left. + (D_j^\dagger \theta c_{i,\sigma}V_\sigma c_{j,-\sigma} + \text{h.c.})] \right\}. \quad \text{(A10)}$$

Here again, we have written the functional δ-function constraint as the integral of a complex exponential. In this case, however, φ_i as the Lagrange multiplier is a complex field such that $\varphi_i^* = \Re\varphi_i - i\Im\varphi_i$. After Wick rotation, we obtain the path integral in Euclidean signature,

$$Z = \int [\mathcal{D}c\mathcal{D}c^\dagger \mathcal{D}D\mathcal{D}D^\dagger \mathcal{D}\varphi\mathcal{D}\varphi^\dagger] \exp^{-\int d^d x(L_0 + H_{con})}. \quad \text{(A11)}$$

In either the Lorentz or Euclidean signatures, φ_i is a Lagrange multiplier. That is, the integral over φ_i must still yield a functional δ function even though the i is absent. This requirement dictates that we must integrate the fluctuations of both the real and imaginary parts of φ_i along the imaginary axis as in Eq. (A5). The result is a stable Gaussian integral (if D is integrated first and then φ_i) which can be evaluated using Eq. (22). Of course in the reverse order, the φ_i integrals simply yield δ functions. In both cases, one ends up with the $+U$ Hubbard model.

APPENDIX B: ABSENCE OF LOCAL SU(2) SYMMETRY

Here we consider the possibility of trivially extending our model at half-filling to a locally SU(2) invariant theory. Let us organize the electron operators in the form

$$\Psi = \begin{pmatrix} c_\uparrow & c_\downarrow \\ c_\downarrow^\dagger & -c_\uparrow^\dagger \end{pmatrix}, \quad \text{(B1)}$$

where the spatial index is suppressed. A local SU(2) transformation acts via left multiplication by an SU(2) matrix h so that $\Psi \to h\Psi$. Let us write h as

$$h = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix}, \quad \text{(B2)}$$

where $|\alpha|^2 + |\beta|^2 = 1$. The electron bilinear $c_\uparrow c_\downarrow$ is a member of the triplet of the local SU(2). We will now show that the correct "middle" term in the electron triplet is $(n_\uparrow + n_\downarrow - 1)/\sqrt{2}$. Transforming the prospective triplet we find

$$\begin{pmatrix} c_\downarrow c_\uparrow \\ \dfrac{n_\uparrow + n_\downarrow - 1}{\sqrt{2}} \\ c_\uparrow^\dagger c_\downarrow^\dagger \end{pmatrix} \to \begin{pmatrix} \alpha^2 & \sqrt{2}\alpha\beta^* & -\beta^{*2} \\ -\sqrt{2}\alpha\beta & |\alpha|^2 - |\beta|^2 & -\sqrt{2}\alpha^*\beta^* \\ -\beta^2 & \sqrt{2}\alpha^*\beta & \alpha^{*2} \end{pmatrix}$$

$$\times \begin{pmatrix} c_\downarrow c_\uparrow \\ \dfrac{n_\uparrow + n_\downarrow - 1}{\sqrt{2}} \\ c_\uparrow^\dagger c_\downarrow^\dagger \end{pmatrix}. \quad \text{(B3)}$$

By inspection it is clear that this matrix is unitary since h is unitary, and also that its determinant is $(|\alpha|^2 + |\beta|^2)^3 = 1$. Thus, we have an SU(2) transformation and we conclude that the electron triplet is

$$W = \begin{pmatrix} c_\downarrow c_\uparrow \\ \dfrac{n_\uparrow + n_\downarrow - 1}{\sqrt{2}} \\ c_\uparrow^\dagger c_\downarrow^\dagger \end{pmatrix}. \tag{B4}$$

Thus, to make our theory local SU(2) invariant, we must also take D and φ to reside in triplets as well. Thus, we posit a charge zero bosonic field φ_0 which we collect into a triplet $\Phi = (\varphi, \varphi_0, \widetilde{\varphi})^T$. SU(2) invariance would necessitate an additional constraint term of the form

$$H_{\mathrm{con}} \to H_{\mathrm{con}} + s\overline{\theta}\sum_i \varphi_{0i}^\dagger[D_{0i} - \theta(n_\uparrow + n_\downarrow - 1)]. \tag{B5}$$

Just as in the cases of the earlier boson fields, the field φ_0 enters the theory as a Lagrange multiplier corresponding to a dynamical field D_0, which will be integrated over. Now that we have determined the form of the electron triplet we can write the generalized Lagrangian, now including the φ_0 terms, in the form

$$\begin{aligned}
L = \int d^2\theta \Bigg[&\sum_i (aD_i^\dagger \dot{D}_i + b\dot{\widetilde{D}}_i^\dagger \widetilde{D}_i + cD_{0i}^\dagger \dot{D}_{0i}) \\
&+ \frac{1}{2}U\sum_j \left(\alpha D_j^\dagger D_j - \beta \widetilde{D}_j^\dagger \widetilde{D}_j + \frac{\delta}{2}D_{0j}^\dagger D_{0j} \right) \\
&- t\sum_{i,j,\sigma} g_{ij}\frac{1}{2}(D_j^\dagger \theta c_{i,\sigma}V_\sigma c_{j,-\sigma} - \overline{\theta}c_{i,\sigma}V_\sigma c_{j,-\sigma}\widetilde{D}_j \\
&+ \overline{\theta}D_{0i}c_{i,\sigma}^\dagger c_{j,\sigma} + \mathrm{h.c.}) + H_{\mathrm{con}} \Bigg],
\end{aligned} \tag{B6}$$

where the constraint is given by

$$\begin{aligned}
H_{\mathrm{con}} = s\overline{\theta}\sum_i \varphi_i^\dagger(D_i - \theta c_{i,\uparrow}c_{i,\downarrow}) + \mathrm{h.c.} + \widetilde{s}\overline{\theta}\sum_i \widetilde{\varphi}_i^\dagger(\widetilde{D}_i - \theta c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger) \\
+ \mathrm{h.c.} + s\overline{\theta}\sum_i \varphi_{0i}^\dagger[D_{0i} - \theta(n_{i,\uparrow} + n_{i,\downarrow} - 1)].
\end{aligned} \tag{B7}$$

The undetermined coefficients may be fixed by the condition that the theory reduces to the Hubbard model when the constraints are solved. Integrating over the φ_i and then D fields, yields

$$\begin{aligned}
L \to &\sum_i [an_{i-\sigma}c_{i\sigma}^\dagger \dot{c}_{i\sigma} - b(1 - n_{i-\sigma})c_{i\sigma}^\dagger \dot{c}_{i\sigma}] + \frac{U}{2}\sum_i \Bigg[\left(\beta - \frac{\delta}{2}\right) \\
&\times (n_{i\uparrow} + n_{i\downarrow} - 1) + (\alpha - \beta + \delta)n_{i\uparrow}n_{i\downarrow}\Bigg]
\end{aligned} \tag{B8}$$

up to total time derivatives. In order for this to yield the Hubbard model we must therefore have $\beta = 1 - \alpha$ and $\delta = 2(1 - \alpha)$. We note that $\alpha = \beta = 2\delta = 1/2$ solves these and is, in fact, the SU(2) symmetric point. Similarly, we find $a = -b = 1$. The coefficient c is unconstrained because D_0 has trivial dynamics. We have found then that the theory at half-filling may be extended to an SU(2) invariant theory. This SU(2) acts only globally, however, this may be plainly seen by examining the interaction terms on the second line of Eq. (B6). In order to make these terms local SU(2) invariant, an explicit SU(2) gauge field (a Wilson line) would have to be introduced. We conclude, then, that the local SU(2) symmetry of the Heisenberg model is broken by the presence of the bosonic degrees of freedom in our model. That a local SU(2) symmetric version of the theory cannot be constructed is not surprising as the Hubbard model lacks this symmetry—in that case it is broken by hopping terms (which again could only be made invariant by the introduction of explicit Wilson lines). In the case of the Heisenberg model, the local SU(2) symmetry appears strictly because of projection. Hence, the exact low energy theory constructed without using projection should not possess symmetries not found in the Hubbard model.

[1] K. G. Wilson and J. Kogut, Phys. Rep., Phys. Lett. **12**, 75 (1974).

[2] P. W. Anderson, *The Theory of Superconductivity in the High-T_c Cuprates*, Princeton Series in Physics (Princeton University Press, Princeton, NJ, 1997).

[3] B. Baskaran, Z. Zou, and P. W. Anderson, Solid State Commun. **63**, 973 (1987).

[4] S. A. Kivelson, D. S. Rokhsar, and J. P. Sethna, Phys. Rev. B **35**, 8865 (1987).

[5] N. Read and B. Chakraborty, Phys. Rev. B **40**, 7133 (1989); N. Read and S. Sachdev, Phys. Rev. Lett. **66**, 1773 (1991).

[6] C. Mudry and E. Fradkin, Phys. Rev. B **49**, 5200 (1994).

[7] T. Senthil and M. P. A. Fisher, Phys. Rev. B **62**, 7850 (2000).

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

[9] X.-G. Wen, Phys. Rev. B **44**, 2664 (1991); **65**, 165113 (2002).

[10] L. Balents and S. Sachdev, Ann. Phys. (N.Y.) **322**, 2635 (2007).

[11] B. I. Shraiman and E. I. Siggia, Phys. Rev. Lett. **60**, 740 (1988); **62**, 1564 (1989).

[12] A. H. MacDonald, S. M. Girvin, and D. Yoshioka, Phys. Rev. B **41**, 2565 (1990).

[13] A. L. Chernyshev, D. Galanakis, P. Phillips, A. V. Rozhkov, and A. M. S. Tremblay, Phys. Rev. B **70**, 235111 (2004).

[14] R. B. Laughlin, arXiv:cond-mat/0209269 (unpublished).

[15] E. Altman and A. Auerbach, Phys. Rev. B **65**, 104508 (2002). This procedure implements a recursive cluster scheme in which the four lowest eigenstates of a Hubbard plaquette are retained. The four lowest eigenstates were then partially Gutzwiller projected. The result is a theory with a hard-core boson system in which the bosons have a Fock space. The latter is absent from the exact theory.

[16] P. W. Anderson, P. A. Lee, M. Randeria, T. M. Rice, N. Trivedi, and F. C. Zhang, J. Phys.: Condens. Matter **16**, R755 (2004).

[17] T. C. Ribeiro and X.-G. Wen, Phys. Rev. Lett. **95**, 057001 (2005).

[18] Z.-Yu Weng, Int. J. Mod. Phys. B **21**, 773 (2007).

[19] S. E. Barnes, J. Phys. F: Met. Phys. **6**, 1375 (1976); **7**, 2637 (1977).

[20] P. Coleman, Phys. Rev. B **29**, 3035 (1984).

[21] G. Kotliar and A. E. Ruckenstein, Phys. Rev. Lett. **57**, 1362 (1986).

[22] S. Moukouri and M. Jarrell, Phys. Rev. Lett. **87**, 167010 (2001).

[23] M. B. J. Meinders, H. Eskes, and G. A. Sawatzky, Phys. Rev. B **48**, 3916 (1993).

[24] A. B. Harris and R. V. Lange, Phys. Rev. **157**, 295 (1967).

[25] H. Eskes and R. Eder, Phys. Rev. B **54**, R14226 (1996); H. Eskes, A. M. Oles, M. B. J. Meinders, and W. Stephan, *ibid.* **50**, 17980 (1994).

[26] S. L. Cooper, G. A. Thomas, J. Orenstein, D. H. Rapkine, A. J. Millis, S. W. Cheong, A. S. Cooper, and Z. Fisk, Phys. Rev. B **41**, 11605 (1990).

[27] S. Uchida, T. Ido, H. Takagi, T. Arima, Y. Tokura, and S. Tajima, Phys. Rev. B **43**, 7942 (1991).

[28] C. T. Chen, F. Sette, Y. Ma, M. S. Hybertsen, E. B. Stechel, W. M. C. Foulkes, M. Schulter, S-W. Cheong, A. S. Cooper, L. W. Rupp, Jr., B. Batlogg, Y. L. Soo, Z. H. Ming, A. Krol, and Y. H. Kao, Phys. Rev. Lett. **66**, 104 (1991).

[29] T.-P. Choy, R. G. Leigh, and P. Phillips (unpublished).

[30] R. G. Leigh, P. Phillips, and T.-P. Choy, Phys. Rev. Lett. **99**, 046404 (2007).

[31] In this expression for Q, $\varphi^{\dagger}$ should be interpreted as a creation operator with the appropriate charge.

[32] R. Friedberg and T. D. Lee, Phys. Rev. B **40**, 6745 (1989).

[33] J. Ranninger, J. M. Robin, and M. Eschrig, Phys. Rev. Lett. **74**, 4027 (1995).

[34] A complex HS transformation of the type in Eq. (22) is closely related to the exponentiation of a δ-functional constraint.

[35] P. W. Anderson, Phys. Rev. **124**, 41 (1961).

[36] E. Dagotto, E. Fradkin, and A. Moreo, Phys. Rev. B **38**, 2926 (1988); I. Affleck, Z. Zou, T. Hsu, and P. W. Anderson, *ibid.* **38**, 745 (1988).

[37] I. Dzyaloshinskii, Phys. Rev. B **68**, 085113 (2003).

[38] P. Phillips, Ann. Phys. (N.Y.) **321**, 1634 (2006); T. D. Stanescu, P. Phillips, and T.-P. Choy, Phys. Rev. B **75**, 104503 (2007).