Geometry-controlled conserving approximations for the t-J model

M. D. Johnson* and Claudius Gros†
Indiana University, Department of Physics, Swain Hall West 117, Bloomington, Indiana 47405

K. J. von Szczepanski
Max-Planck-Institut für Festkörperforschung, 7000 Stuttgart 80, Federal Republic of Germany
(Received 17 May 1990; revised manuscript received 5 October 1990)

We present in detail a Green's-function approach for studying charged-spin systems which preserves the local constraints prohibiting double occupancy. This approach satisfies Wick's theorem, uses a fermionic expansion around a singly occupied Néel state, and treats charge and spin degrees of freedom on an equal footing. For the antiferromagnetic Heisenberg model we recover gapless spin excitations (renormalized spin waves) in a straightforward real-space random-phase-approximation approach. This expansion is strictly controlled by a geometrical factor, $1/z$, where z is the coordination number. We describe the incoherent motion of charges (holes) in the t-J model by a self-retracing-path approximation and consider two competing contributions to the coherent hole propagation. These approximations are made conserving in a constructive fashion by mapping Feynman diagrams to an equivalent tight-binding model. To study the accuracy of this procedure, we have made a detailed numerical check against the results obtained by exact diagonalization of a 4×4 system with one hole, finding excellent agreement both in and near the Ising limit.

I. INTRODUCTION

The problem of strongly correlated spin-$\frac{1}{2}$ electrons on a lattice is one of the unsolved fundamental problems in condensed-matter physics that can be described by a simple Hamiltonian. It is characterized by just three parameters: the density n of particles per lattice site, a hopping amplitude t, and the strength of the Coulomb repulsion U. The simplest realization of this model is thought to be the Hubbard Hamiltonian, for which only one state is available at each lattice site for each spin species.

Shortly after the discovery of high-temperature superconductivity, Anderson[1] proposed that the electronic properties of the Cu-O planes might be described by such a model. This was later justified by Zhang and Rice,[2] who showed that the low-energy excitations of the Cu-O planes could be described by a coherent superposition of O holes bound to a Cu spin in a singlet state, resulting in a single-band model.

Despite the flurry of interest which has focused on this fundamental problem, progress has lagged. Indeed, only two undisputed facts are known so far.

The first is Nagaoka's theorem,[3] which states that the ground state for $U\rightarrow\infty$ (taken before $N\rightarrow\infty$), with all sites occupied but one, is ferromagnetic. Lately, several groups have examined a possible extension of this ferromagnetic instability to the case of several empty sites or even a finite density of these holes.[4] This has turned out to be a very subtle question, because of the extreme sensitivity to boundary conditions. This area of attention will be beyond the scope of the approach developed below.

It is well known[5] that in the limit of very strong correlation, these systems can be described by what is now called the t-J model. In this limit, only states with singly occupied sites are allowed in the Hilbert space. This, the famous constraint of no double occupancy, will play a fundamental role in what follows. The charge carriers are then the empty sites, the holes. The model has three parameters, the hopping matrix element t, the antiferromagnetic spin interaction J, and the hole density n. In the absence of charge carriers, the t-J model reduces to the antiferromagnetic Heisenberg model.

It is in this limit that a second, generally accepted statement can be made, namely, the existence of antiferromagnetic long-range order for the case of a hypercubic lattice in two and higher dimensions. Although this was hypothesized quite a while ago,[6] only recently has it been established beyond doubt[7] for the case of two dimensions. This is the limit that the method we will develop in detail in this paper is based on.

Naturally, one of the ultimate goals would be to understand the full phase diagram of the strongly correlated electron gas on, say, the square lattice. Many researchers have tried to head in this direction using various mean-field approaches.[8] Unfortunately, the approximations used in many of these approaches are not always controlled to the extent that one would like to see. The main obstacle is the difficulty in handling the local constraint of strong repulsion in a satisfactory way. Therefore, methods which can handle this constraint exactly, such as Monte Carlo methods at zero or finite temperature,[9] variational approaches with projected wave functions[10] and other variational methods and expansions for the two-dimensional (2D) antiferromagnetic Heisenberg model[11] have stimulated large interest.

For the case of the t-J model, the *supersymmetric* region $t\sim J$ is similar to the region of intermediate coupling $t\sim U$ in the Hubbard model.[12] In this region there is only one energy scale, and the question arises whether

any small parameter is left in the model. A small density of charge carriers might be identified as a small parameter, but to date it has proven impossible to exploit it as such systematically. Instead we will argue that, at least for the properties of a system with only one hole or no holes, geometry alone might provide a suitable expansion parameter. More specifically, we will examine the influence on the charge excitations of the coordination number z of a general hypercubic lattice.

Another approach to the problem of the local constraint is to attempt to make a connection to the weak-coupling limit.[13] Such work typically focuses on the intermediate-coupling regime by extending techniques valid in the small-U limit, invoking a principle of moderation—viz., that nature tends not to choose extreme limits.[14] Or, one might argue, properties vary only smoothly between the extreme limits. If so, it becomes important to find the correct interpolation procedure.

One of the goals of this paper is to show that it is possible to take the local constraint exactly into account, if one limits oneself to quite specific questions; e.g., the nature of a single spin or charge excitation in a correlated background. This problem has generated interest for two reasons. One is the hope that, even for a finite density of charge carriers, the propagation might be similar to that of a single hole. Ultimately, one would like to test one of the interesting concepts which have been popularized lately, such as the notion of a marginal Fermi liquid[15] or of a Luttinger liquid.[16]

A second reason for interest in the properties of a single hole is that it provides an opportunity to compare with the results of exact diagonalization of small systems (typically 16 or 18 sites with one vacancy[17]). Although one has to clarify the consequences of finite-size effects, this is a very powerful test. To date, none of the analytic theories has taken advantage of this possibility of a close comparison with computer experiment.

Now we will discuss briefly several historical approaches to the situation with one hole, and then outline the structure of this paper.

As early as 1970, Brinkman and Rice[18] examined an essential aspect of this problem. They introduced the now famous "self-retracing-path approximation," also called the Brinkman-Rice approximation. They showed that, even in a disordered spin background, a hole can lower its kinetic energy nearly to the bottom of what would be the free-carrier band, in this case by restricting motion to paths which self-retrace. This works because the self-retracing paths constitute a substantial fraction of all possible paths. One important result of this approximation is that the hole's motion becomes incoherent. Said another way, the hole is unable to propagate. Although there are many different paths which allow the hole to propagate to a different lattice site, generally these disturb the spin background. Viewed in terms of path integrals, most paths through which the hole can move end up in a state orthogonal to the original state. These paths only add incoherently.

A very nice approach to the one-hole problem has been suggested by Trugman.[19] It consist of building up a local basis of spin flips near the hole, and then diagonalizing the effective tight-binding problem which results. The strong points of this approach are that it takes both the local constraint and translational invariance exactly into account and that it is easy to use. On the other hand, a correct treatment of the quantum spin fluctuations in the spin background is not yet settled. We will be able to connect this to our more rigorous Green's-function approach.

An approach which has strongly influenced the development of this field is due to Kane, Lee, and Read.[20] Their method combines a slave-fermion with a Holstein-Primakoff spin-wave representation. The coupling of hole motion to the spin excitation (which is induced via the constraints) then can give rise to a coherent propagation of the hole. It would be interesting to test this approach against numerical results[21] obtained by the exact diagonalization of small cells. This is not an easy task. It turns out to be necessary to complement the original formulation of this theory with certain graphs which, as we will demonstrate in this paper, dominate for a weak spin coupling J. These are graphs in which a hole hops around closed loops so as to restore the antiferromagnetic spin background. The importance of these graphs was first stressed by Trugman. As we will show below, a consistent conserving approximation incorporating this motion is highly nontrivial.

A very different approach has been developed by Becker and Fulde,[22] based on a projection technique in Liouville space. This method is systematic and does not need a Wick's theorem. It has a broad range of applicability. On the other hand, for the Heisenberg case, rotational invariance is not restored naturally, as it is for a 1/(coordination number) expansion. The latter point will be demonstrated in this paper.

In the following we will set out to explain in detail a real-space Green's-function approach, which we will then use to calculate certain properties of single spin and charge excitations. This method has been presented previously in a compressed form.[23] It has by now been well established that the Heisenberg antiferromagnet on the hypercubic lattice has long-range order of the Néel type (for $D \geq 2$). It is therefore reasonable to describe the Heisenberg limit by a systematic expansion around a simple mean-field Hamiltonian, the ground state of which is the pure Néel state. In Sec. II we will explain how to do this in Green's-function language. For this we will use a formulation based on the underlying electrons, an approach which allows for Wick's theorem.

In Sec. III we will then describe the implementation of this real-space procedure to study the spin excitation. This will have several outcomes.

First, it turns out that the spin problem is controlled by $1/z$, i.e., there exists a systematic expansion parameter determined simply by geometry. Here z is the coordination number, i.e., the number of nearest neighbors, on a simple-hypercubic lattice; $z=4$ in two dimensions. Thus we can describe the spin-$\frac{1}{2}$ antiferromagnetic Heisenberg model by a rapidly converging series expansion around the Néel state.

Second, we find that our initial explicit breaking of the spin rotational symmetry (the Néel state is the starting

point of the perturbation expansion) becomes *spontaneous* at each order in the expansion in $1/z$. This occurs in the following sense. It is easy to label the contributing diagrams to the spin propagator's self-energy by their order in $1/z$. If one retains all contributing diagrams to a given order, the resulting spin excitation spectrum necessarily is gapless in the Heisenberg limit, with a linear dispersion. This gapless spectrum can be associated with the Goldstone mode of the spontaneously broken spin-rotation symmetry.[24] We will carry this through to the first two orders. In lowest order, the result is identical to results obtained by many other methods, including the Holstein-Primakoff transformation.[25] One might say that this seems to be a fixed point in approximation space.

We also calculate the first quantum correction to the spin excitation spectrum, here to the next order in $1/z$. It turns out that the first quantum correction only renormalizes the spin-wave velocity, so that only in some higher order is the dispersion relation changed. This is an interesting result because it explains why the usual spin-wave result agrees so well with numerical estimates of the staggered magnetization, even in two dimensions. The reason is that the sublattice magnetization is dimensionless, and therefore at zero temperature depends only on the dispersion relation and not on the energy scale. But, as mentioned above, the change in the dispersion relation is only of order $1/z^2$, which is small in two dimensions.

Third, we will be able to simulate by our approach the real-space basis-expansion method of Trugman, discussed above. We then can compare the contribution of different connected and disconnected diagrams to his effective-Hamiltonian formulation. It turns out that the subtraction of the spin fluctuation in the quantum ground state, which is the reason for identifying connected diagrams, is a nontrivial task in a nongraphical theory.

Fourth, our formalism is closely related to a $1/z$ expansion for some thermodynamical properties of the 3D Heisenberg ferromagnet developed independently by Fishman and Liu.[26] They perform a cluster expansion around the same mean-field Hamiltonian we use as the starting point for our perturbation expansion. For the free energy, they retain the *lattice-independent* terms, which we would call *self-retracing*.

Now, how is this technique generalized to case of a charge excitation? At first, this might seem difficult, since our expansion is based on the case of exactly half-filling. Indeed, this method has no straightforward generalization to the case of finite density of holes.

However, in the spirit of Brinkman and Rice, the properties of a single charge excitation can be obtained by studying the Green's function for the creation and propagation a hole in the half-filled ground state. In this formulation, the kinetic energy, which is the term which propagates the hole, does not perturb the half-filled ground state. The main difficulty is that, because of the constraint against double occupancy, the kinetic-energy operator is made up of six (instead of the usual two) fermion creation and destruction operators. As a result, at no point during the propagation is a doubly occupied site ever created. This kinetic energy leads naturally to the emission of a spin excitation every time the hole hops to a different site, as was pointed out by Kane, Lee, and Read. Therefore the charge motion is strongly coupled to the spin system.

Since it is so important to handle the constraints correctly, it will be convenient to work in real space. Furthermore, we will consider the isotropic spin interaction as a limiting case of a more general Hamiltonian, where the diagonal coupling between the z component of spins is given by J_z and the transverse coupling is given by $J_\perp$. Then to study the Green's function for the hole, we will proceed in three steps. First, we will consider the Ising limit, $J_\perp=0$. Second, we will examine the asymptotic Ising limit, $0<J_\perp \ll J_z$. The last step is to recover the Heisenberg point, $J_\perp=J_z$. We have earlier argued that the essential physics of the Heisenberg limit can be described by retaining all graphs contributing to the asymptotic Ising limit, and replacing all spin excitation lines by the full, gapless spin propagators. In a previous paper we laid down some essential aspects of this limit.[23] In this paper we will concentrate on the first two steps.

One of the main aims of this paper is to develop a systematic method which is checked wherever possible against numerical results, i.e., results obtained by exactly diagonalizing a 16-site system with one vacancy. Only by cross-checking different methods, we believe, can one control the approximations inherent in any analytic theory.

In order for such a comparison to be possible in detail, all approximations made must be conserving. Although nonconserving approximations might describe a similar physical picture, they generally generate some unpleasant side effects, such as a negative density of states. Such an outcome would seriously hamper any detailed comparison with computer experiments. Here we used a constructive technique to generate conserving approximations beginning with some low-order Feynman diagrams. The technique we have developed involves a mapping from the diagrams to an equivalent tight-binding problem which is then solved exactly. This procedure automatically shows what extra diagrams must be added to those already selected in order to gain a conserving approximation.

Exact diagonalization efforts have typically calculated the bandwidth, defined for the 4×4 lattice by the difference in the lowest eigenvalues for $\mathbf{k}=(0,0)$ and $\mathbf{k}=(\pi/2,\pi/2)$. $\mathbf{k}$ is a good quantum number. The outcome is that for large values of J/t, the minimum lies at $\mathbf{k}=(\pi/2,\pi/2)$, while for small values this switches to $\mathbf{k}=(0,0)$. The level crossing occurs at $J/t\sim\frac{1}{10}$ in the Heisenberg limit. Note that this ratio is near to what is proposed to be the relevant physical regime for the high-T_c superconductors. Now, slightly below this crossover, the transition to the fully polarized Nagaoka state takes place. (By Nagaoka's theorem, the ground state is ferromagnetic for $J=0$ and one hole.[3]) The common interpretation of this level crossing has therefore been that it is a precursor of the ferromagnetic transition to the Nagaoka state. In a finite system, the Nagaoka state disappears for J greater than some finite, critical J_C, But

in an infinite system the Nagaoka state disappears at infinitesimal $J>0$, i.e., $J_C \to 0$. We will show strong evidence that the crossover in the energy minimum is in fact not due to the onset of the Nagaoka phase. We can reproduce this crossover using a Green's function calculation for one hole in an infinite system: we find that the crossover occurs at finite J_C, not, as would be expected by the above argument, an infinitesimal.

It was earlier pointed out that even in a Néel state the hole can propagate coherently.[27] That is, it can move on certain paths to a different site during which the antiferromagnetic spin configuration of the background is completely restored. These paths, which were neglected in the self-retracing-path approximation, generally trace out a loop. Therefore we will refer to them as "loop paths" in what follows. The smallest such path, which traces out a single square, is the "plaquette path."

In Sec. IV we will examine the Ising limit, by considering both the self-retracing paths and the dominant plaquette graph. The latter can be viewed as the lowest-order vertex correction to the hole propagation in the Ising limit. We will show that one can map the problem exactly onto an effective tight-binding Hamiltonian. This Hamiltonian has one-dimensional rings in each unit cell, but can nevertheless be solved analytically. Because of this mapping, this approximation to the Green's function is guaranteed to be conserving. The essential physics of the plaquette graphs is that of a one-dimensional tunneling problem on a discrete lattice, where the barrier (of order J_z) is provided by the Ising energy of the disturbed spins during intermediate steps. Consequently, for low J_z the hole acquires a large dispersion, with the bottom of the band at $\mathbf{k}=(0,0)$. For lower and lower J_z, more and more (i.e., longer and longer) loop graphs begin to contribute. In our approach we take into account only the shortest, or plaquette, graph, plus one other that arises when the approximation is made conserving. This approximation will therefore break down when longer paths matter, i.e., $J_z \sim t/5$.

In Sec. V we consider the asymptotic Ising limit. In addition to all graphs retained in the Ising limit, a different type of motion is now possible. This is a process in which a hole hopes twice, leaving two flipped spins behind. These can then be flipped back to Néel order by a vacuum fluctuation. This process, proportional to $J_\perp$, we call the "string-erasure" graph. As is true for all our graphs, our formulation yields an explicit analytic formula for this process. It is quite remarkable that a straightforward extension of the conserving approximation used in the pure Ising limit is possible. The interesting fact is that this process competes with the plaquette graphs, i.e., the two channels interfere. The erasure graph alone yields a coherent band with a width proportional to $J_\perp$. The minimum of this band lies at $\mathbf{k}=(\pi/2,\pi/2)$. At sufficiently large $J_\perp$, this graph dominates. Thus it is possible to see how accurately this single graph represents the physics of string erasures. To check this in more detail we compare our results to small-cell values for $J_\perp/J_z$ equal to $\frac{1}{4}$ and $\frac{1}{20}$. We find $\sim 3\%$ agreement for $J_z \sim t$ (a J_z large enough to give small finite-size effects). This is one order of magnitude better than the 25% accuracy one might have expected for $J_\perp/J_z=\frac{1}{4}$. This strongly suggests that the higher-order corrections, beyond the string-erasing graph $\sim J_\perp$ considered by us, must be reduced by a geometrical expansion factor [perhaps $1/(2z)$] just as was the case for the spin waves.

Additionally, we will discuss in Sec. V a process which lets the hole move between the two sublattices. This process is the dynamical analog of the fact that the expectation value of the kinetic energy is linear in t whenever quantum fluctuations are present in the spin background. This in contrast to the case of a Néel background. But, as was suggested to us by Brenig,[28] we will show that this process only contributes to the spectral weight by influencing the intensity of certain poles, but not their positions.

II. FORMALISM

In this section we outline a fermionic real-space Green's-function method for studying the t-J model. In sum, we perturb about the half-filled mean-field Ising Hamiltonian, the ground state of which is the Néel state. We use ordinary fermion operators, replacing electron operators by operators which create vacant or doubly occupied sites to eliminate explicit reference to the two sublattices. The unique feature that allows a straightforward analysis is that we work in real space, so that it is very simple to ensure that the constraints against double occupancy are preserved. The price paid is that the Hamiltonian breaks into several perturbations, and, moreover, that the interesting Heisenberg limit can only be approached perturbatively.

We treat the t-J model generalized to include asymmetry in the spin coupling:

$$H_{t\text{-}J}=-t\sum_{\langle i,j\rangle,\sigma}\{(1-c^{\dagger}_{i,-\sigma}c_{i,-\sigma})c^{\dagger}_{i,\sigma}c_{j,\sigma}(1-c^{\dagger}_{j,-\sigma}c_{j,-\sigma})+\text{H.c.}\}+\sum_{\langle i,j\rangle}\{J_z S^z_i S^z_j+(J_\perp/2)(S^+_i S^-_j+S^-_i S^+_j)\}\ . \tag{1}$$

Here $c^{\dagger}_{i,\sigma}$ is the creation operator for an electron with spin $\sigma=\uparrow,\downarrow$ on the lattice site i, and $\langle i,j\rangle$ denotes pairs of nearest-neighbor (NN) sites. The spin operators are $S^z_i=\frac{1}{2}(c^{\dagger}_{i,\uparrow}c_{i,\uparrow}-c^{\dagger}_{i,\downarrow}c_{i,\downarrow})$, $S^+_i=c^{\dagger}_{i,\uparrow}c_{i,\downarrow}$, and $S^-_i=c^{\dagger}_{i,\downarrow}c_{i,\uparrow}$.

We want to study $H_{t\text{-}J}$ on the restricted Hilbert space from which states with doubly occupied sites are excluded. Viewed as a limit of the Hubbard model, this implies that doubly occupied sites are so high in energy, due to strong electron correlations (the Hubbard U), that they can be ignored entirely. Therefore in the t-J model the strong electron correlations are represented crucially by the constraint against double occupancy, as well as by the residual coupling between electron spins.

The first term in $H_{t\text{-}J}$ represents the kinetic energy of holes doped into the half-filled state. The second term is an antiferromagnetic coupling between the spins of elec-

trons on adjacent sites. (We assume that $t, J_z, J_\perp > 0$.) The operators in parentheses in the first term enforce the constraints against double occupancy.

In the following we will consider only bipartite lattices, focusing on hypercubic lattices (square, cubic, etc.). (A generalization to any lattice, such as the triangular lattice, which allows for a classical antiferromagnetic spin ordering is possible.) For dimensions $D \geq 2$, the ground state at half-filling (i.e., one electron per site) is known to possess Néel-type antiferromagnetic order. We will therefore take the Néel state as our starting point.[29] Split the Ising part of the Hamiltonian into a mean-field and an interaction piece:

$$S_i^z S_j^z = \tfrac{1}{2}(-S_i^z + S_j^z) + (S_i^z - \tfrac{1}{2})(S_j^z + \tfrac{1}{2}) + \tfrac{1}{4} \ . \tag{2}$$

Here i and j denote sites on the $A(\uparrow)$ and $B(\downarrow)$ sublattices, respectively. The first part of Eq. (2) is the mean-field part of the Ising interaction and, like the complete Ising interaction, has the Néel state as its ground state. Moreover, this mean-field piece is quadratic in fermion operators and so forms a useful unperturbed Hamiltonian.

It is convenient to transform from electron to quasihole and quasiparticle degrees of freedom. The creation operators $h^\dagger$ and $d^\dagger$ are defined as

$$\begin{aligned} &h_i^\dagger = c_{i\uparrow}, \quad d_i^\dagger = c_{i\downarrow}^\dagger, \quad i \in A \ , \\ &h_j^\dagger = c_{j\downarrow}, \quad d_j^\dagger = c_{j\uparrow}^\dagger, \quad j \in B \ . \end{aligned} \tag{3}$$

Thus, acting on the Néel, state, $h^\dagger$ creates a vacant site, a hole, while $d^\dagger$ creates a doubly occupied site. More accurately, in general $h^\dagger$ destroys a "correct"-spin electron, while $d^\dagger$ creates a "wrong"-spin electron.[30] Within a constant, the t-J Hamiltonian is exactly

$$\begin{aligned} H_{t\text{-}J} &= H_0 + H_z + H_\perp + T_K \ , \\ H_0 &= (\varepsilon_0/2) \sum_{i \in A,B} (h_i^\dagger h_i + d_i^\dagger d_i) \ , \\ H_z &= -(J^z/4) \sum_{\langle i,j \rangle} (h_i^\dagger h_i + d_i^\dagger d_i)(h_j^\dagger h_j + d_j^\dagger d_j) \ , \\ H_\perp &= (J^\perp/2) \sum_{\langle i,j \rangle} (d_i^\dagger h_i^\dagger d_j^\dagger h_j^\dagger + h_j d_j h_i d_i) \ , \\ T_K &= -t \sum_{\langle i,j \rangle} [(d_i d_i^\dagger) h_i d_j (h_j^\dagger h_j) \\ &\qquad + (h_i^\dagger h_i) d_i^\dagger h_j^\dagger (d_j d_j^\dagger) + \text{H.c.}] \ . \end{aligned} \tag{4}$$

Here $\varepsilon_0 = zJ^z/2$, where z is the number of nearest neighbors. H_0, the mean-field Ising interaction, will be our unperturbed Hamiltonian. H_z and $H_\perp$ are perturbations in the spin coupling. H_z, the remainder of the Ising $S_i^z S_j^z$ coupling, represents the energy decrease that occurs when two flipped spins are moved onto neighboring sites. $H_\perp$, the transverse spin coupling, lets antiparallel spins on neighboring sites flip simultaneously. T_K is the kinetic energy for hole motion. It is important to note that the operators within parentheses in T_K are the constraints which enforce single occupancy. For example, the first term in T_K can hop an up-spin from site j to site i, but only if there is no down-spin on either site. That is, T_K will never cause double occupancy.

By construction, the ground state of the unperturbed Hamiltonian H_0 is the Néel state (since $\varepsilon_0 > 0$). Because the h and d operators satisfy the usual fermion anticommutation rules and H_0 is quadratic, we can use Wick's theorem to generate a fermionic diagrammatic expansion of Green's functions.

The zero-temperature time-ordered Green's functions that are useful for us to consider are

$$\begin{aligned} G_{h,ij}(t) &= -i \langle \mathcal{T} \hat{h}_i(t) \hat{h}_j^\dagger(0) \rangle \ , \\ G_{d,ij}(t) &= -i \langle \mathcal{T} \hat{d}_i(t) \hat{d}_j^\dagger(0) \rangle \ , \\ D_{ij}(t) &= -i \langle \mathcal{T} \hat{h}_i(t) \hat{d}_i(t) \hat{d}_j^\dagger(0) \hat{h}_j^\dagger(0) \rangle, \quad i,j \in A \ , \end{aligned} \tag{5}$$

where $\langle \cdots \rangle$ is an expectation value within the exact half-filled ground state and the time dependences here are in the Heisenberg representation [indicated by the hat, $\hat{O}(t)$]. We can call $G_{h,ij}(t)$ and $G_{d,ij}(t)$ the Green's functions, respectively, for holes and doubly occupied sites, with one caveat—namely, that $G_{d,ij}(t = 0-)$ measures the density of doubly occupied sites only in a rather trivial sense, since, by construction, there are none in the half-filled state. The two-particle function D_{ij} corresponds to a flipped spin, since $h^\dagger$ removes a correct spin and $d^\dagger$ replaces it with a wrong spin (on the A sublattice $S_i^- = c_{i,\downarrow}^\dagger c_{i,\uparrow} = d_i^\dagger h_i^\dagger$).

In the interaction representation [indicated by no hat, $O(t)$], $h^\dagger(t) = \exp(i\varepsilon_0 t/2) h^\dagger$ and $d^\dagger(t) = \exp(i\varepsilon_0 t/2) d^\dagger$. The unperturbed Green's functions are

$$\begin{aligned} G_{d,ij}^{(0)}(t) = G_{h,ij}^{(0)}(t) &= -ie^{-i\varepsilon_0 t/2} \Theta(t) \delta_{ij} \\ &\equiv G^{(0)}(t) \delta_{ij} \ , \\ D_{ij}^{(0)}(t) &= -ie^{-i\varepsilon_0 t} \Theta(t) \delta_{ij} \equiv D^{(0)}(t) \delta_{ij} \ . \end{aligned} \tag{6}$$

Here $\Theta(t) = 1$ for $t > 0$ and $\Theta(t) = 0$ for $t < 0$. Notice that these are *site-diagonal* Green's functions—the unperturbed Green's functions are nonpropagating. This, and the fact that they are causal [proportional to $\Theta(t)$], makes the diagrammatic expansion very natural and simple in real space. Physically, the unperturbed hole Green's function is causal because no charge carriers are present in the half-filled case. The unperturbed spin Green's function $D^{(0)}(t)$ is causal because no quantum fluctuations are present in the Néel state. $D(t)$ will cease being causal, once quantum fluctuations are taken into account.

The Fourier transforms of the unperturbed Green's functions, defined as

$$f(\omega) = \int_{-\infty}^{\infty} dt \, e^{i\omega t} f(t) \ , \tag{7}$$

are

$$\begin{aligned} G^{(0)}(\omega) &= \frac{1}{\omega - \varepsilon_0/2 + i\delta} \ , \\ D^{(0)}(\omega) &= \frac{1}{\omega - \varepsilon_0 + i\delta} \ , \end{aligned} \tag{8}$$

where δ is a positive infinitesimal. In the rest of this paper, everywhere ω appears it should be understood that $i\delta$ is to be added; where useful, we have written the

infinitesimal explicitly.

Graphically we represent $G_h^{(0)}$ as a single line and $G_d^{(0)}$ as a single line with a tick mark (see Fig. 1). Then $D^{(0)}$, representing a flipped spin or h-d pair, is naturally drawn as a double line. (To make contact with the usual diagrams, one can expand the double line for $D^{(0)}$ into a particle-hole type of polarization insertion.) Because there are three separate interactions in $H_{t\text{-}J}$ when it is written in the form Eq. (4), we need three different interaction lines in the diagrammatics. These are shown as the basic vertices in Fig. 1.

Explanations of how to evaluate the various graphs that can arise are given in the Appendixes. Most of these rules are best understood by examples, and we will endeavor to make them clear, especially in the following section.

III. SPIN EXCITATIONS

In this section we will study the spin excitations of the Heisenberg Hamiltonian (i.e., the t-J model at half filling). We have two purposes here. We will show that on hypercubic (square, simple cubic, etc.) lattices our real-space analysis naturally leads to an expansion in $1/z$ (where z is the number of nearest neighbors), and that at each order in $1/z$ the rotational symmetry of the Heisenberg Hamiltonian is preserved—i.e., the excitation spectrum is gapless. First, however, we will use this problem to illustrate several different ways to compute graphs in our real-space analysis. This we will do by repeating the calculation of the same graph three ways.

One way to find the energy of spin excitations is to consider the time-ordered spin propagator

$$D_{ij}(t) \equiv -i\langle \mathcal{T}\hat{S}_i^+(t)\hat{S}_j^-(0)\rangle \ . \tag{9}$$

If both the initial and final sites are on sublattice A, this can be written in terms of h and d operators as shown in Eq. (5). Thus the propagation of a flipped spin is here represented by a particle-hole Green's function. The unperturbed $D_{ij}^{(0)}(t)$ is given in Eq. (6) and is represented by the double line as shown in Fig. 1.

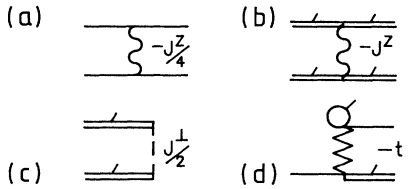

FIG. 1. Basic vertices representing the perturbations in Eq. (4). Here a single line represents $G_h^{(0)}$, the site-diagonal Green's function for a hole. A single line with a tick mark represents $G_d^{(0)}$, the propagator for a doubly occupied site. A double line represents $D^{(0)}$, the Green's function for a flipped spin. (a) H_z vertex between any single lines. (b) H_z vertex between double lines (i.e., spin propagators): the energy of two flipped spins is lowered if they are moved to neighboring sites. (c) $H_\perp$ can end a pair of double lines (as shown here) or begin a pair. This represents the flipping of two neighboring antiparallel spins, from wrong to right (as shown here) or from right to wrong. (d) T_k can hop a hole to a neighboring site, leaving a flipped spin behind. The loop with a tick-mark is the constraint forbidding double occupancy.

In the absence of perturbation, of course, the flipped spin cannot propagate, and $D_{ij}^{(0)}(t) \propto \delta_{ij}$. The perturbation $H_\perp$ permits the flipped spin to move. We will analyze the effect of $H_\perp$ (and also H_z) by summing over a particular infinite set of graphs. The basic point we will discover is that the resulting excitation spectrum is gapless when $J_\perp$ is set equal to J_z. To lowest order, in fact, the spectrum is identical to the spin-wave result obtained using Holstein-Primakoff operators.[31] Now, the true Heisenberg Hamiltonian must have a gapless Goldstone mode, because of the spontaneous breaking of the spin's continuous rotational symmetry. That our graphical analysis yields gaplessness is perhaps unexpected, since our starting point is the mean-field Ising Hamiltonian, which is manifestly not rotationally invariant;[24] the lowest-lying excitation in the mean-field Hamiltonian is a single spin flip, which has a finite energy ε_0.

The reason we obtain a gap less spectrum is that the spin calculation can be expressed as an expansion in $1/z$ or, on a hypercubic lattice in d dimensions, $1/(2d)$. It turns out, for example, that the lowest-order graph in $H_\perp$ also gives the lowest-order contribution to the $1/z$ expansion. Since the complete Heisenberg Hamiltonian is gapless for all z, so must each contribution be gapless in a $1/z$ expansion. We have carried this expansion out for the first two orders in $1/z$. [Mathematically an expansion in $1/z$ is equivalent to an expansion in $1/d$. But realistically the magnitudes of the prefactors are important. Here they are such, indications are, that we might in fact take $1/(2z)$ as an expansion parameter.]

At the end of this section we will show that this calculation is formally equivalent to a tight-binding problem, and exploit this equivalence to relate this approach to a method due to Trugman. These two apparently dissimilar approaches are linked by their common focus on real-space motion. A Trugman calculation does not precisely reproduce our Green's-function results; the difference appears to be the way in which diagrams can be identified as connected versus disconnected.

A. Lowest order: spin waves

Let us proceed to calculate the spin excitations, first to lowest order in $J_\perp$. The perturbation $H_\perp$ allows the flipped spin to move to a next-nearest-neighbor site by two steps, as shown in Fig. 2. We can completely analyze the spin problem by keeping only processes like these, in which the flipped spin propagates only on sublattice A. (We took this approach in earlier work;[23] see also below.) Here instead we present an alternative approach, based on the use of an anomalous Green's function, which proves simpler for the higher-order calculation. We define a 2×2 matrix $\bar{D}_{ij}^{\alpha\beta}$ where the row and column indices α and β specify the sublattice upon which, respectively, i and j lie (i.e., $i\in\alpha$, $j\in\beta$, and both α and β can

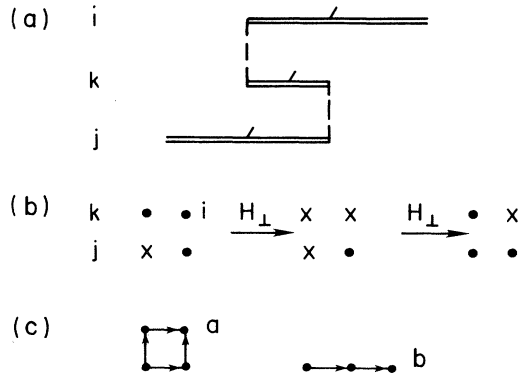

FIG. 2. (a) The simplest process by which a flipped spin can move around on one sublattice. (b) A pictorial representation of the same process. The flipped spin, represented by ×, is moved by two applications of $H_\perp$. (c) There are two kinds of sites that can be reached. Type "*a*" sites can be reached via two paths. Type "*b*" sites can only be reached via one path. [The graph in (a) can also return the flipped spin back to its original location on site j in z ways.]

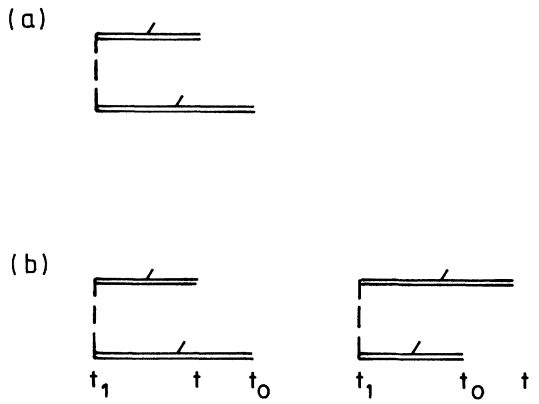

FIG. 3. (a) The basic step by which a flipped spin can move about the plane. This, and a partner in which both double lines are absorbed, leads to the off-diagonal self-energy $\Sigma^{(1)}$ [Eq. (24)]. This self-energy is of order ε_0. (b) Both time orderings of the graph shown in (a). Time runs from left to right. Because the unperturbed Green's functions are causal, $t_1 < t$ and $t_1 < t_0$. The time orderings of the two graphs shown here are, respectively, $t_1 < t < t_0$ and $t_1 < t_0 < t$.

be either A or B). Define

$$\begin{aligned}
\overline{D}_{ij}^{AA}(t) &= -i\langle \mathcal{T}[\hat{h}_i\hat{d}_i](t)[\hat{d}_j^\dagger\hat{h}_j^\dagger](0)\rangle, \quad i\in A,\ j\in A\ , \\
\overline{D}_{ij}^{BB}(t) &= -i\langle \mathcal{T}[\hat{d}_i^\dagger\hat{h}_i^\dagger](t)[\hat{h}_j\hat{d}_j](0)\rangle, i\in B,\ j\in B\ , \\
\overline{D}_{ij}^{AB}(t) &= +i\langle \mathcal{T}[\hat{h}_i\hat{d}_i](t)[\hat{h}_j\hat{d}_j](0)\rangle, \quad i\in A,\ j\in B\ , \\
\overline{D}_{ij}^{BA}(t) &= +i\langle \mathcal{T}[\hat{d}_i^\dagger\hat{h}_i^\dagger](t)[\hat{d}_j^\dagger\hat{h}_j^\dagger](0)\rangle, \quad i\in B,\ j\in A\ .
\end{aligned} \tag{10}$$

$\overline{\mathbf{D}}$ is almost the spin propagator given in Eq. (9), except that the sign of the off-diagonal elements is reversed (so that our Feynman rules, given in the Appendixes, are still valid). The notation $[\cdots](t)$ means that all of the operators within the bracket have the time dependence t. $\hat{O}(t)$ means that the time dependence is in the Heisenberg representation; hereafter, $O(t)$ means in the interaction representation. Using Eqs. (5) and (8), the unperturbed Green's function here is

$$\overline{\mathbf{D}}_{ij}^{(0)}(\omega) = \begin{bmatrix} D^{(0)}(\omega) & 0 \\ 0 & D^{(0)}(-\omega) \end{bmatrix} \delta_{ij} \equiv \overline{\mathbf{D}}^{(0)}\delta_{ij}\ . \tag{11}$$

The lowest-order process which lets the spin move is shown in Fig. 3(a). We will solve this in three ways to illustrate the different approaches possible. First for orientation is a direct calculation in time space. The graph in Fig. 3(a) is a particular diagram in the Wick's expansion of[32]

$$\overline{D}_{ij}^{AB}(t) = (+i)(-i)\int dt_1 \langle \mathcal{T}[h_i d_i](t)[h_j d_j](0) H_\perp(t_1)\rangle_{\text{connected}}\ . \tag{12}$$

In particular, Fig. 3(a) arises from a single term in the expansion of

$$\overline{D}_{ij}^{AB}(t) = \delta_{\langle ij\rangle}\left[\frac{J_\perp}{2}\right]\int dt_1 \langle \mathcal{T}([h_i d_i](t)[h_j d_j](0)[d_j^\dagger h_j^\dagger d_i^\dagger h_i^\dagger](t_1))\rangle_{\text{connected}}\ . \tag{13}$$

Here $\delta_{\langle ij\rangle}$ means that i and j must be nearest neighbors. By directly contracting the expression Eq. (13), one readily finds

$$\overline{D}_{ij}^{AB}(t) = (-1)\delta_{\langle ij\rangle}\left[\frac{J_\perp}{2}\right]\int dt_1\, D^{(0)}(-t_1)D^{(0)}(t-t_1)\ , \tag{14}$$

where the -1 can be traced to the choice of sign in the definition of D^{AB}. Fourier transforming this [Eq. (7)] gives

$$\overline{D}_{ij}^{AB}(\omega) = -\left[\frac{J_\perp}{2}\right] D^{(0)}(\omega)D^{(0)}(-\omega)\ . \tag{15}$$

The integrals are easy bearing in mind that our unperturbed $D^{(0)}(t)$, although time ordered, is also *causal:* nonzero only for positive time arguments.

The above calculation illustrates the origin of the Feynman rules in time space given in Appendix A. We can also simply write down the answer is frequency space using the rules in Appendix B. In Fig. 3(a), time is running from left to right. That is, lines connecting a vertex

from its left are *entering* the vertex, and those connecting to the vertex's right are *leaving* it. Conservation of energy in Fig. 3(a) thus requires the bottom and top double lines to have frequencies ω and $-\omega$, respectively. Then

$$\overline{D}_{ij}^{AB}(\omega)=\left[\frac{J_\perp}{2}\right]\delta_{\langle ij\rangle}D^{(0)}(\omega)D^{(0)}(-\omega)(-1)^{2+1} \quad (16)$$

which is equivalent to Eq. (15). There are two fermion loops in Fig. 3(a) (since each double line is equivalent to the usual bubble); and because our definitions of D and $\overline{D}$ contain an overall (-1) relative to ordinary rules, the sign is $(-1)^{2+1}$, as explained in Appendix B.

Finally, we obtain this result a third way, one which is unnecessarily complicated here but which makes higher-order diagrams very easy to solve. This method is explained in Appendix D. A basic parent diagram leads to a set of daughter graphs in which every possible time ordering is specified; it should be borne in mind that the unperturbed Green's functions are causal, so that lines cannot reverse direction (i.e., time runs from left to right in all of our diagrams). The possible time orderings of Fig. 3(a) are shown in Fig. 3(b). In this case there are only two, $t_0<t$ and $t<t_0$. The contribution from each time-ordered graph is obtained by a product of factors, one corresponding to each time interval. In Fig. 3(b), each graph has two time intervals; e.g., the second has intervals $[t_1,t_0]$ and $[t_0,t]$. To find the contribution from each time interval, figure out a total frequency $\omega_{\rm tot}$ (a vertical sum of the frequencies on each line crossing the interval) and a total on-site energy $\varepsilon_{\rm tot}$ (a vertical sum of the on-site energy ε_0 for each double line); this gives a factor $(\omega_{\rm tot}-\varepsilon_{\rm tot}+i\delta)^{-1}$. Multiply one such term from each time interval. For example, for the second graph in Fig. 3(b) this procedure gives

$$(\omega-\omega-2\varepsilon_0+i\delta)^{-1}(\omega-\varepsilon_0+i\delta)^{-1} . \quad (17)$$

Multiply the factors arising from each time interval, and add the contribution from each graph corresponding to a distinct time ordering. The sign and other prefactors come from the usual Feynman rules. For Fig. 3(b) we have

$$\overline{D}_{ij}^{AB}(\omega)=(-1)\left[\frac{J_\perp}{2}\right]\delta_{\langle ij\rangle}\frac{1}{(-2\varepsilon_0+i\delta)}\left[\frac{1}{\omega-\varepsilon_0+i\delta}+\frac{1}{-\omega-\varepsilon_0+i\delta}\right], \quad (18)$$

which as one can readily see, is equal to Eq. (15). It is simple to show why this works. The explanation, and a summary of this approach, are given in Appendix D.

The graph shown in Fig. 3(a) is a step which permits a flipped spin to move to a nearest-neighbor site. There is a second graph, leading to $\overline{D}^{BA}$, which can be obtained by reflecting Fig. 3(a) about a vertical line. By repeating these basic steps, the flipped spin can walk around the plane. This repetition is an RPA (random-phase approximation) or Dyson's resummation, as follows. From Eq. (16), the self-energy for Fig. 3(a) and its partner is

$$\Sigma_{ij}^{(1)}(\omega)=-\left[\frac{J_\perp}{2}\right]\delta_{\langle ij\rangle}\sigma_x \quad (19)$$

where

$$\sigma_x=\begin{bmatrix}0&1\\1&0\end{bmatrix} . \quad (20)$$

The possible ways to get from the origin (site 0) to any other site i, using only repetitions of the basic steps, combine to give the propagator

$$\overline{\mathbf{D}}_{i0}^{(1)}(\omega)=\overline{\mathbf{D}}^{(0)}\delta_{i0}+\overline{\mathbf{D}}^{(0)}\Sigma_{i0}^{(1)}\overline{\mathbf{D}}^{(0)}+\overline{\mathbf{D}}^{(0)}\Sigma_{ij}^{(1)}\overline{\mathbf{D}}^{(0)}\Sigma_{j0}^{(1)}\overline{\mathbf{D}}^{(0)}+\cdots , \quad (21)$$

where a repeated subscript implies a sum. The third piece above, for example, represents a two-step move from site 0 to j to i. The sum here is nothing but the usual Dyson's sum (or RPA sum for bubble graphs like these) written in the real-space language that we favor.

The Fourier transform of $\overline{\mathbf{D}}_{i0}^{(1)}$ is

$$\begin{aligned}\overline{\mathbf{D}}^{(1)}(\mathbf{k},\omega)&=\sum_i e^{-i\mathbf{k}\cdot\mathbf{R}_i}\overline{\mathbf{D}}_{i0}^{(1)}(\omega)\\&=\overline{\mathbf{D}}^{(0)}+\overline{\mathbf{D}}^{(0)}\Sigma^{(1)}(\mathbf{k},\omega)\overline{\mathbf{D}}^{(0)}\\&\quad+\overline{\mathbf{D}}^{(0)}\Sigma^{(1)}(\mathbf{k},\omega)\overline{\mathbf{D}}^{(0)}\Sigma^{(1)}(\mathbf{k},\omega)\overline{\mathbf{D}}^{(0)}+\cdots\\&=\overline{\mathbf{D}}^{(0)}+\overline{\mathbf{D}}^{(0)}\Sigma^{(1)}\overline{\mathbf{D}}^{(1)} ,\end{aligned} \quad (22)$$

where $\mathbf{R}_i$ is the coordinate position of site i, and so, as usual,

$$\overline{\mathbf{D}}^{(1)}(\mathbf{k},\omega)=[1-\overline{\mathbf{D}}^{(0)}(\omega)\Sigma^{(1)}(\mathbf{k},\omega)]^{-1}\overline{\mathbf{D}}^{(0)}(\omega) . \quad (23)$$

To evaluate this we need the Fourier transform

$$\Sigma^{(1)}(\mathbf{k},\omega)=\sum_i e^{-i\mathbf{k}\cdot\mathbf{R}_i}\Sigma_{i0}^{(1)}=\Sigma^{(1)}(\mathbf{k},\omega)\sigma_x , \quad (24)$$

where

$$\Sigma^{(1)}(\mathbf{k},\omega)=-\left[\frac{J_\perp}{2}\right]z\gamma_{\mathbf{k}}=-\Delta\varepsilon_0\gamma_{\mathbf{k}} ,$$

$$\Delta=J_\perp/J_z , \quad (25)$$

and

$$\gamma_{\mathbf{k}}=\frac{2}{z}\sum_{\mu=1}^{z/2}\cos k_\mu . \quad (26)$$

Here we are explicitly limiting ourselves to hypercubic lattices in $d=z/2$ dimensions. The sum in Eq. (26) is over the distinct Cartesian coordinates, so that $\gamma_k=\cos k$ in 1D, $\gamma_{\mathbf{k}}=(\cos k_x+\cos k_y)/2$ in 2D, etc., and thus $|\gamma_{\mathbf{k}}|\le 1$, regardless of the dimensionality.

Inserting Eq. (24) into Eq. (23) gives, for the AA component of our first-order spin propagator

$$D_{AA}^{(1)}(\mathbf{k},\omega)=\frac{1}{\omega-\varepsilon_0+i\delta-\dfrac{(\varepsilon_0\gamma_{\mathbf{k}}\Delta)^2}{-\omega-\varepsilon_0+i\delta}} . \quad (27)$$

(This Green's function for motion around the A sublattice can also be obtained directly from the graph in Fig. 2, without the use of a matrix Green's function;[23] see also below.) Because of the frequency dependence of the self-energy, this has poles at $\omega=\pm(\omega_{\mathbf{k}}-i\delta)$, where

$$\omega_{\mathbf{k}}=\varepsilon_0\sqrt{1-\Delta^2\gamma_{\mathbf{k}}^2}\ . \tag{28}$$

Notice that in the Heisenberg limit ($J_\perp=J_z=J$) this gives $\omega_{\mathbf{k}}=(zJ/2)\sqrt{1-\gamma_{\mathbf{k}}^2}$, which is gapless and, in fact, is exactly the result from spin-wave theory. Furthermore, Eq. (28) contains no explicit factors $1/z$. As we will show in the following subsection, it is the lowest (zero order) contribution to a $1/z$ expansion of the spectrum.

We can rewrite Eq. (27) to show its poles:

$$D_{AA}^{(1)}(\mathbf{k},\omega)=\frac{1}{2}\left[1+\frac{\varepsilon_0}{\omega_{\mathbf{k}}}\right]\frac{1}{\omega-\omega_{\mathbf{k}}+i\delta}+\frac{1}{2}\left[1-\frac{\varepsilon_0}{\omega_{\mathbf{k}}}\right]\frac{1}{\omega+\omega_{\mathbf{k}}-i\delta}\ . \tag{29}$$

In the Ising limit $\omega_k\to\varepsilon_0$ and the second term, which explicitly breaks causality, vanishes. Similarly, one can derive and expand the formula for $D_{AB}^{(1)}$:

$$D_{AB}^{(1)}(\mathbf{k},\omega)=\frac{\Delta\varepsilon_0\gamma_k}{2\omega_k}\left[\frac{1}{\omega-\omega_{\mathbf{k}}+i\delta}-\frac{1}{\omega+\omega_{\mathbf{k}}-i\delta}\right]\ . \tag{30}$$

B. $1/z$ expansion

This gapless spin-wave spectrum is a somewhat surprising result. Our starting point is the mean-field Ising Hamiltonian, which is far from rotationally invariant. Besides, the gaplessness is restored when $J_\perp/J_z$ is set to unity, even through we are performing perturbation theory in $J_\perp$. How then can this level of perturbation theory, resulting from the Dyson's sum arising from one low-order graph, restore rotational invariance? The answer is that there is an additional control parameter ($1/z$) in the spin calculation, and that at any given order in $1/z$ gaplessness, and so apparently rotational invariance, is restored. This is like the way a $1/N$ expansion retains the Hamiltonian's symmetries at each order in $1/N$, or the way a loop expansion does so. Since the true Heisenberg Hamiltonian is rotationally invariant for every z, then at each order in $1/z$ the spectrum must be gapless.

When $J_\perp=J_z$, the complete Heisenberg Hamiltonian is rotationally invariant, the true ground state spontaneously breaks this symmetry, and the Goldstone mode is gapless. If this solution can be expanded in powers of $1/z$, as we show by construction, then necessarily the contribution at each power must be gapless. Symbolically, expand the spectrum in a series in $1/z$:

$$E(\mathbf{k},z)=\sum_{n=0}^{\infty}E_n(\mathbf{k})z^{-n}\ . \tag{31}$$

Since the spectrum is gapless for all z [i.e., $E(\mathbf{0},z)=0$], then it must also be the case that $E_n(\mathbf{0})=0$. That is, the contribution to the spectrum at each order is gapless.

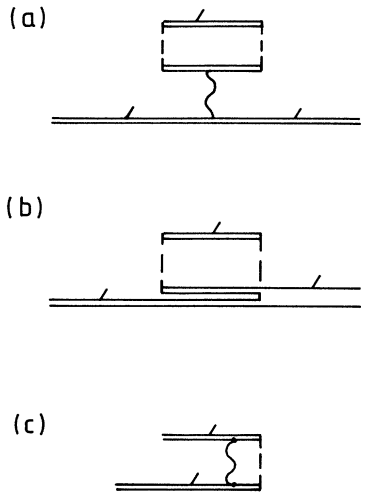

FIG. 4. Next-lowest order corrections in $1/z$ to the spin propagator. All of these have self-energies of order ε_0/z. (a) and (b) contribute to the diagonal part of Σ, and (c) is off-diagonal. (a) is a correction to the on-site energy, (b) derives from an interaction with a vacuum fluctuation, and (c) is a vertex correction.

We will argue that a $1/z$ expansion arises and learn how to identify the order of various graphs in the following. These arguments are very similar to those which lead to a $1/z$ expansion in a cluster expansion of the free energy.[26] The basic point is that the energy scale for spin excitations is $\varepsilon_0=zJ_z/2$, which contains a factor of z. This is the energy scale because our starting point is the mean-field Ising Hamiltonian, in which a flipped spin (an excitation) feels the average displeasure of z nearest neighbors; equivalently, the unperturbed spin Green's function $D^{(0)}=(\omega-\varepsilon_0+i\delta)^{-1}$ has ε_0 as its on-site energy. Feynman diagrams possess energy factors of order J (i.e., J_z or $J_\perp$) for each interaction line. Since $J\sim\varepsilon_0/z$, an nth order diagram is generally of order $1/z^n$ relative to the unperturbed case. Actually this is not quite true—there are also geometric factors of z which tend to offset this reduction. These are easy to handle, as we will now illustrate.

We have asserted that the Dyson's sum developed from Fig. 3 gives the lowest-order contribution to the $1/z$ expansion. [Notice that $\Sigma^{(1)}$, in Eq. (25), is of order ε_0, i.e., the same as the energy in the unperturbed Green's function, Eq. (8).] All of the next-order contributions are shown in Fig. 4. These are a subset of the graphs that are first or second order in $H_\perp$, and zeroth or first order in H_z. They turn out to be all of the graphs with self-energies of order ε_0/z, i.e., those which give corrections of order $1/z$. The first two graphs, in Figs. 4(a) and 4(b), contribute to the diagonals of Σ and hence renormalize the unperturbed on-site energy ε_0. The third graph, in Fig. 4(c), is off-diagonal and renormalizes $\Sigma^{(1)}$.

Let us estimate the order in $1/z$ of each graph in Fig.

4. The first diagonal graph [Fig. 4(a)] has a self-energy $\Sigma^{(a)}\sim J_\perp^2 J_z z^2/\varepsilon_0^2$. Here the $J_\perp^2 J_z$ factor arises because this graph is second order in $H_\perp$ and first in H_z, the $1/\varepsilon_0^2$ comes from the frequency dependence of the internal propagators (or by dimensional analysis), and—most importantly—the geometric factor of z^2 appears because there are z nearest neighbors upon which the upper double lines can lie. Hence $\Sigma^{(a)}\sim\varepsilon_0/z$. Similarly, the other diagonal graph [Fig. 4(b)] has $\Sigma^{(b)}\sim J_\perp^2 z/\varepsilon_0\sim\varepsilon_0/z$. The off-diagonal graph in Fig. 4(c) is a dressed version of the basic step shown in Fig. 3(a). Relative to this basic step, Fig. 4(c) picks up an extra factor of $J_z/\varepsilon_0\sim 1/z$ and hence $\Sigma_{\mathbf{k}}^{(c)}\sim\Sigma_{\mathbf{k}}^{(1)}/z\sim\varepsilon_0/z$. Because it is so simple to estimate the order in $1/z$ of any graph, we believe that those pictured in Fig. 4 exhaust the possible next-lowest-order contributions in $1/z$.

Now let us actually evaluate the graphs shown in Fig. 4. This is simple using the Feynman rules of Appendix D. For example, Fig. 4(a) has four possible time orderings, which give the frequency dependence

$$\frac{1}{\omega-\varepsilon_0}\left[\frac{1}{\omega-3\varepsilon_0}\right]^2\frac{1}{\omega-\varepsilon_0}+2\frac{1}{\omega-\varepsilon_0}\left[\frac{1}{\omega-3\varepsilon_0}\right]^2\frac{1}{-2\varepsilon_0}+\frac{1}{-2\varepsilon_0}\left[\frac{1}{\omega-3\varepsilon_0}\right]^2\frac{1}{-2\varepsilon_0}=\frac{1}{(\omega-\varepsilon_0)^2(-2\varepsilon_0)^2}\ . \tag{32}$$

Hence

$$\bar{D}_{ij}^{AA,(a)}=z^2(-J_z)\left[\frac{J_\perp}{2}\right]^2\frac{1}{(\omega-\varepsilon_0)^2(-2\varepsilon_0)^2}\delta_{ij}\ . \tag{33}$$

The factor of z^2 arises because of the z nearest neighbors upon which each of the upper double lines can lie. With a similar graph for $\bar{D}^{BB}$, this gives

$$\mathbf{\Sigma}^{(a)}(\mathbf{k},\omega)=-\frac{\Delta^2}{2z}\varepsilon_0\mathbf{1}\ , \tag{34}$$

where 1 is the 2×2 identity matrix. In a very similar way one finds

$$\mathbf{\Sigma}^{(b)}(\mathbf{k},\omega)=-\frac{\Delta^2}{2z}(\omega-3\varepsilon_0+i\delta)\mathbf{1} \tag{35}$$

and

$$\mathbf{\Sigma}^{(c)}(\mathbf{k},\omega)=\frac{1}{z}\mathbf{\Sigma}^{(1)}(\mathbf{k},\omega)\ . \tag{36}$$

Inserting the sum of $\Sigma^{(1)}$, $\Sigma^{(a),(b),(c)}$ [Eqs. (24) and (34)–(36)] into the Dyson's expression Eq. (23) then gives the propagator $\bar{D}^{(2)}$ accurate to second order in $1/z$. For example, the AA component of the result is

$$\bar{D}_{AA}^{(2)}(\mathbf{k},\omega)=\frac{1}{\tilde{\omega}-\tilde{\varepsilon}_0+i\delta-\dfrac{(\alpha\tilde{\varepsilon}_0\gamma_{\mathbf{k}}\Delta)^2}{-\tilde{\omega}-\tilde{\varepsilon}_0+i\delta}}\ , \tag{37}$$

where we have defined

$$\tilde{\omega}=\omega\left[1+\frac{\Delta^2}{2z}\right],\quad \tilde{\varepsilon}_0=\varepsilon_0\left[1+\frac{\Delta^2}{z}\right], \tag{38}$$

and $\alpha^2=2(1-\Delta^2)/z$. In the Heisenberg limit $J_\perp=J_z$, the propagator $D^{(2)}$ in Eq. (37) becomes identical to Eq. (27), with the substitution of the *renormalized* frequency $\tilde{\omega}$ and on-site energy $\tilde{\varepsilon}_0$. The dispersion is thus simply

$$\tilde{\omega}=\tilde{\varepsilon}_0\sqrt{1-\Delta^2\gamma_{\mathbf{k}}^2}\ , \tag{39}$$

and therefore $\omega=\omega'_{\mathbf{k}}-i\delta$, where, at the Heisenberg point,

$$\omega'_{\mathbf{k}}=\varepsilon_0\left[1+\frac{1}{2z}\right]\sqrt{1-\gamma_{\mathbf{k}}^2}\ . \tag{40}$$

Thus we have regained the gapless spin-wave spectrum, but with a renormalized spin-wave velocity $\varepsilon_0\to\varepsilon_0(1+1/2z)$. [Compare Eq. (40) with Eq. (28)]. As explained above, keeping all the graphs contributing to lowest or to first order in $1/z$ restores gaplessness. Notice that this is quite unlike the correction to the spin-wave velocity that occurs in an ordinary calculation of corrections to spin-wave theory, using, for example, Holstein-Primakoff operators. There the corrections are formally of order $1/S$, which is questionable for spin $\frac{1}{2}$, as here, and the spin-wave velocity to this order is found to be different.[31] The velocity renormalization is in rather good agreement with the result found by Singh using a series expansion around the Ising limit.[33]

C. Effective tight-binding problem

Let us return to the lowest-order calculation above. For the purposes of this section, it is convenient to redo this calculation using Fig. 2—that is, restricting the flipped spin to lie only on one sublattice (say, A). It is then not necessary to use a matrix formulation. Thus we calculate not the complete $\bar{\mathbf{D}}$ but $\bar{D}^{AA}$ or, equivalently, D as given in Eq. (5). Using the rules of Appendix D, one readily finds that Fig. 2(a) has the self-energy

$$\Sigma_{ij}^{AA}(\omega)=\left[\frac{J_\perp}{2}\right]^2 C_{\mathbf{R}_i-\mathbf{R}_j}\frac{1}{-\omega-\varepsilon_0+i\delta}\ . \tag{41}$$

The geometric factor $C_{\mathbf{R}}$ is explained in Fig. 2(c): $C_{\pm\hat{\mathbf{x}}\pm\hat{\mathbf{y}}}=2, C_{\pm2\hat{\mathbf{x}},\pm2\hat{\mathbf{y}}}=1$, and $C_0=z$. Fourier transforming this self-energy gives

$$\Sigma^{AA}(\mathbf{k},\omega)=\frac{(\varepsilon_0\gamma_{\mathbf{k}}\Delta)^2}{-\omega-\varepsilon_0+i\delta}\ , \tag{42}$$

which is identical to the result extracted from the matrix formulation [Eq. (27)].

It is illuminating to recast this solution into a tight-binding form. Let $|\mathbf{r}\rangle$ be a state with some particle (here, the flipped spin) on site $\mathbf{r}$. Then imagine a tight-binding Hamiltonian H_{TB} for which

$$H_{TB}|\mathbf{r}\rangle=(\varepsilon_0+zt_2)|\mathbf{r}\rangle+2t_2\sum_{\delta=\pm\hat{\mathbf{x}}\pm\hat{\mathbf{y}}}|\mathbf{r}+\delta\rangle+t_2\sum_{\delta=\pm2\hat{\mathbf{x}},\pm2\hat{\mathbf{y}}}|\mathbf{r}+\delta\rangle\ , \tag{43}$$

where t_2 is some hopping matrix element. Again we are restricting attention to the hypercubic lattice in $z/2$ dimensions. It is easy to diagonalize the Schrödinger's equation $H_{TB}|\psi\rangle=\omega|\psi\rangle$ here. The solution is $|\psi_{\mathbf{k}}\rangle=\sum_{\mathbf{r}}e^{i\mathbf{k}\cdot\mathbf{r}}|\mathbf{r}\rangle$, with

$$\omega=\varepsilon_0+z^2t_2\gamma_{\mathbf{k}}^2 , \tag{44}$$

where $\gamma_{\mathbf{k}}$ is defined in Eq. (26). This is precisely the form obtained by our lowest-order calculation [cf. Eq. (27)], if we identify for the hopping amplitude

$$t_2=t_2(\omega)=\left[\frac{J_\perp}{2}\right]^2\frac{1}{-\omega-\varepsilon_0+i\delta} . \tag{45}$$

That is, the basic step shown in Fig. 2, when used as the beginning of a Dyson's or RPA sum, is equivalent to an effective tight-binding problem, but with an energy-dependent hopping amplitude. Solving Eq. (44) for the root $\omega=\omega_{\mathbf{k}}$ then gives Eq. (28). In fact, the correspondence is exact—H_{TB} above is the effective Hamiltonian arising from the action of $H_\perp$ up to second order within the Hilbert space in which the flipped spin moves only on sublattice A.

A Hamiltonian cannot have the eigenenergies as a parameter, of course, so perhaps this formal analogy to a tight-binding model appears more confusing than clarifying. In fact, our calculation is closely related to an approach used by Trugman[19] to study single-hole properties in the t-J and t-t'-J models. The basic step in Fig. 2 lets a flipped spin move from some site to a next-nearest-neighbor site, and, as we shall see, the same basic step is the heart of a Trugman calculation. The Trugman method is very similar in spirit to our real-space Green's-function calculation, and to compare the two we will here sketch a calculation of the spin excitation using the Trugman method. The differences between the result from this calculation and our Green's-function approach shed some light on Trugman's method.

Trugman's approach is based on restricting the Hilbert space to a small set of important states—and *every translation* of these states. The Hamiltonian [in this case $H_\perp$, given in Eq. (4)] is diagonalized within this set. In some sense it is like a cluster calculation, except that the "cluster" moves with the object whose properties are being studied. In the present case we will use the set of seven states shown in Fig. 5, and every possible translation of these (on the same sublattice). These are the states which, by repeated application of $H_\perp$, allow the flipped spin the move. For the remainder of this section, we restrict ourselves to 2D.

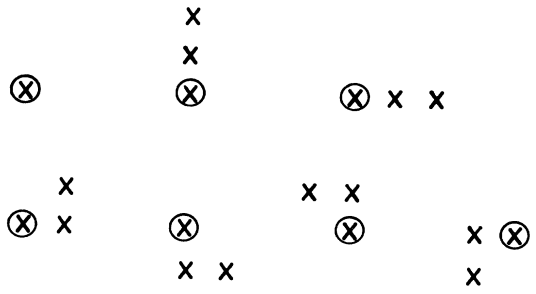

FIG. 5. Basic spin states used for a lowest-order Trugman-type calculation of the spin excitation. These states and all of their translations are used. Here an × represents a spin flip on some site. The circled × indicates the fiducial site $\mathbf{r}$. The states can be labeled $|\mathbf{r}m\rangle$ with $m=1$–7.

Following the approach described by Trugman,[19] here with the Hamiltonian $H_\perp$ and considering spin excitations, leads in lowest order to an effective tight-binding problem with seven basis states per unit cell. These states can be labeled $|\mathbf{r}m\rangle$, where $\mathbf{r}$ is a site on sublattice A, say, and $m=1,2,\ldots,7$ specifies one of the seven basis states (labeled by this fiducial site $\mathbf{r}$). We refer to Trugman's work for an explanation of the calculation, and present the results here. Expanding an eigenstate $|\psi_{\mathbf{k}}\rangle=\sum_{m=1}^{7}c_m\sum_{\mathbf{r}}e^{i\mathbf{k}\cdot\mathbf{r}}|\mathbf{r}m\rangle$, the Schrödinger equation yields for the Hamiltonian $H_\perp$ a set of seven linear equations $\mathbf{A}\mathbf{c}=\mathbf{0}$ for the c_m's and the eigenvalue ω, where $\mathbf{A}$ is

$$\begin{bmatrix}
(2J_z-\omega)/J_\perp & 1+e^{-2ik_x} & 1+e^{-2ik_y} & 1+e^{-i(k_x+k_y)} & 1+e^{i(k_x-k_y)} & 1+e^{i(k_x+k_y)} & 1+e^{i(-k_x+k_y)} \\
1+e^{2ik_x} & (4J_z-\omega)/J_\perp & 0 & 0 & 0 & 0 & 0 \\
1+e^{2ik_y} & 0 & (4J_z-\omega)/J_\perp & 0 & 0 & 0 & 0 \\
1+e^{i(k_x+k_y)} & 0 & 0 & (4J_z-\omega)/J_\perp & 0 & 0 & 0 \\
1+e^{i(-k_x+k_y)} & 0 & 0 & 0 & (4J_z-\omega)/J_\perp & 0 & 0 \\
1+e^{-i(k_x+k_y)} & 0 & 0 & 0 & 0 & (4J_z-\omega)/J_\perp & 0 \\
1+e^{i(k_x-k_y)} & 0 & 0 & 0 & 0 & 0 & (4J_z-\omega)/J_\perp
\end{bmatrix} . \tag{46}$$

Here $m=1$ labels the state with one spin flip, and $m=2,3,\ldots,7$ are the states with three flips in Fig. 5. These equations can be solved for the eigenvalue as usual by setting $\det(\mathbf{A})=0$; instead, as was pointed out to us by Trugman, it is more illuminating to solve them by row reduction.[34] The latter approach yields

$$\{(2J_z-\omega)(4J_z-\omega)-J_\perp^2[2+(\cos k_x+\cos k_y)^2]\}c_1=0 ,$$
$$(4J_z-\omega)c_m=0, \quad m=2,3,\ldots,7 . \tag{47}$$

Thus the low-energy solutions have $c_1\neq 0$, $c_{m>1}=0$. This eigenvalue equation for ω is equivalent to a tight-

binding problem with only one basis state per unit cell. That is, we can focus attention on the states with one flipped spin, $|\mathbf{r}1\rangle \equiv |\mathbf{r}\rangle$, and ignore the other states in Fig. 5. Then the eigenvalue equation Eq. (47) can be viewed as arising from the tight-binding Hamiltonian[34]

$$H'_{TB}|\mathbf{r}\rangle = (\varepsilon_0 + 12t'_2)|\mathbf{r}\rangle + 2t'_2 \sum_{\delta=\pm\hat{x}\pm\hat{y}} |\mathbf{r}+\delta\rangle + t'_2 \sum_{\delta=\pm 2\hat{x},\pm 2\hat{y}} |\mathbf{r}+\delta\rangle \ , \tag{48}$$

with the hopping amplitude

$$t'_2 = \left[\frac{J_\perp}{2}\right]^2 \frac{1}{\omega - 4J_z} \ . \tag{49}$$

This is extremely similar to the tight-binding expression Eqs. (43) and (45) derived from our Green's function calculation. Trugman's method and our real-space graphical analysis are formally equivalent to tight-binding problems which are similar but not identical. There are two basic differences. The first is unimportant: the hopping amplitude t'_2 found by Trugman's method has an on-site energy $4J_z$ in the denominator [Eq. (49)], whereas earlier in Eq. (45) we had an on-site energy $\varepsilon_0 = 2J_z$. That t'_2 has ω and not $-\omega$, as in t_2, may be related to this. Part of the difference in on-site energy is that in the Trugman method the Ising interaction is treated exactly, whereas in our lowest-order calculation it is included only at the mean-field level. As we pointed out earlier we can incorporate $H_\perp$ and H_z (the rest of the Ising interaction) on an equal footing in a $1/z$ expansion.

The second difference is more profound. The Trugman calculation gives an on-site frequency-dependent contribution to H'_{TB} [Eq. (48)] which has a multiplicity of 12, whereas the corresponding term in our approach has a multiplicity of 4 in 2D [Eq. (43)]. The difference is important. The on-site term in a Trugman calculation results from $H_\perp$ acting twice, first creating and then removing a pair of flipped spins on sites adjacent to the original spin. This is pictured in Fig. 6(a) and in our graphical language in Fig. 6(b). The factor of 12 arises in this 2D calculation because there are $z=4$ ways of choosing the nearest-neighbor site k and $z-1$ ways of choosing i. The on-site term in our graphical calculation, on the other hand, arises from setting $i=j$ in Fig. 2; this is pictured in Fig. 6(c). Here there are z ways to choose k.

Thus the basic difference between these two results [Eqs. (43) and (48)] is that Trugman's method, in our language, has a contribution from a graph [Fig. 6(b)] which we *exclude* because it is disconnected. Our calculation, on the other hand, has a contribution from a graph [Fig. 6(c)] which cannot be expressed simply as a sequence of spin flips, and so apparently cannot be included in Trugman's procedure. Keeping only connected graphs means that the poles of the Green's function given excitation energies relative to the perturbed ground state. In this case, unless only connected diagrams are kept, rotational invariance cannot be restored—i.e., we cannot get the gapless result. The approach used by Trugman (for hole problems) and imitated by us here to look at

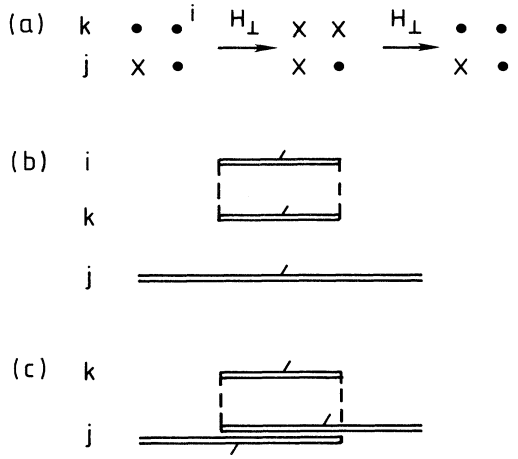

FIG. 6. (a) The on-site contribution in a Trugman calculation results from this process, in which a neighboring pair of spins is flipped and then flipped back. There are $z(z-1)$ ways to do this. (b) The same process shown graphically is a disconnected diagram. (c) The diagram which provides the on-site contribution in the graphical calculation. This diagram is a special case of Fig. 2(a). There are z ways to do this.

spin excitations is very appealing because it (1) emphasizes real-space properties, (2) focuses on the intuitively important states (those which represent fluctuations near the excitation), and (3) is simple. On the other hand, the approach lacks the kind of formal underpinning which makes interpretation of the results certain. Here we seem to see an example of that—the basic difference between our (formally justified) Green's function approach and the Trugman method appears to be the distinction between connected and disconnected graphs. A calculation of the properties of one hole by Trugman's method similarly contains contributions from disconnected graphs (see Sec. V).

The graph in Fig. 6(c) cannot be interpreted simply in terms of flipped spins. It occurs because we are expressing spins in terms of fermion operators; it is part of the price we pay for making Wick's theorem valid. Graphs like Fig. 6(c), which naively appear to violate the Pauli exclusion principle, occur in ordinary fermionic graphical calculations, not just here. They are simply visual representations of terms in the Wick's expansion of certain vanishing matrix elements. Although the complete matrix element vanishes, it is broken into a sum over pieces which do not individually vanish. For example, Fig. 6(c) plus the graphs in Figs. 6(b) (for $i=j$) and 4(b) all add to zero for the time ordering shown there, because they are the result of expanding the product

$$\langle d_j(t)h_j(t)[d_k h_k d_j h_j](t_2)[h_k^\dagger d_k^\dagger h_j^\dagger d_j^\dagger](t_1)h_j^\dagger(0)d_j^\dagger(0)\rangle \tag{50}$$

which is identically zero. Since the disconnected piece in Fig. 6(b) is used to cancel an S-matrix denominator, the remaining pieces in Figs. 6(c) and 4(b) do indeed contribute.

D. Summary of spin excitations

We have found that our graphical approach naturally leads to a $1/z$ expansion for the energy of the spin excitation, and that necessarily at each order in this expansion gaplessness is restored. Our lowest-order calculation is formally equivalent to a tight-binding problem, and we have exploited the equivalence to point out the similarities of this approach to that of Trugman; the difference between the two appears to be the identification of disconnected diagrams. In this case, unless this identification is made correctly, rotational invariance cannot be restored.

IV. CHARGE EXCITATIONS IN THE ISING LIMIT

In this section we will study the properties of charge excitations in the Ising limit ($J_\perp=0$). This is sometimes called the t-J_z model. The following section deals with the case of a small but nonzero $J_\perp$. The approach we take is to study the Green's function $G(\mathbf{k},\omega)$ for a single hole moving in the half-filled ground state. Ordinarily the spectral weight extracted from this Green's function, $-\pi^{-1}\mathrm{Im}G(\mathbf{k},\omega)$, has one of two basic forms. The spectral weight might have one or more quasiparticle peaks (a δ function, perhaps broadened, in ω) whose position is a function of $\mathbf{k}$. Instead, the spectrum might be incoherent, with a broadened distribution of weight and no sharp peaks. Generally one finds a combination of quasiparticle peaks and some incoherent portion.

The picture is different in the case of a strongly correlated system, as in the t-J_z model considered here. In this case, the motion of the charge (the hole) disturbs the background of spins, and as a result the charge and spin degrees of freedom are completely intertwined. In the lowest order *self-retracing-path* approximation, originally due to Brinkman and Rice, the hole is completely unable to propagate.[18] This is because a moving hole, as is by now well known, leaves a "string" of reversed spins which, in this approximation, is cleaned up by having the hole retrace its path. In the language of a diagrammatic expansion, the self-retracing path approximation is equivalent to a "noncrossing" approximation. When J_z is set to zero, the resulting spectral weight looks completely incoherent, in that it has no peaks.[18] When a nonzero J_z is turned on, but the self-retracing approximation is still made, the hole can be thought of as moving in a one-dimensional linear potential formed by the Ising energy of its string.[35] This confining potential gives a sequence of bound states for the hole; correspondingly one finds a sequence of δ functions in the spectral weight. Although sharp peaks, these are not quasiparticle peaks: they are independent of $\mathbf{k}$, because in this approximation the hole cannot propagate.

Besides the self-retracing paths, there are paths which, even in the Ising limit, clean up the string of reversed spins. These paths, which we call loop or plaquette paths, permit the hole to propagate, and hence give a quasiparticle peak in the spectral weight. The importance of such processes was first emphasized by Trugman.[19] Our formalism gives a very natural way to study the loop paths. They turn out to have a very simple interpretation in terms of one-dimensional tunneling through the Ising barrier, so that the bandwidth for hole propagation via this propagation falls off exponentially with J_z.

A very simple physical picture emerges from this approach. The self-retracing paths dress the hole, creating a composite object consisting of a hole plus a cloud of disturbed spins. This composite is then permitted to propagate by the loop path. In this section we will study each of these processes in turn, using the language of our real-space Green's-function approach.

A. The nonpropagating hole in the self-retracing-path approximation

To investigate the dynamics of a single hole in the correlated spin background we study the one-hole Green's function

$$G_{ij}(t)=-i\langle \mathcal{T}\hat{h}_j(t)\hat{h}_i^{\dagger}(0)\rangle \ , \tag{51}$$

where the inner product is taken with respect to the exact ground state at half-filling (i.e., with no holes present). In the Ising limit, the hole is injected into the Néel state. When the hole moves, it leaves behind a string of overturned spins, as shown in Fig. 7(a). Most paths through which the hole can move will leave some spins flipped. Such states are orthogonal to the Néel state, and so only

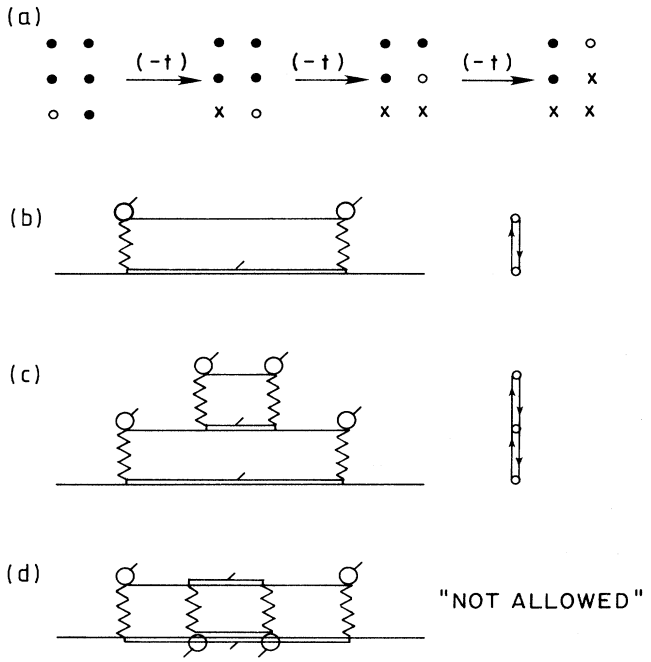

FIG. 7. (a) A moving hole, represented by ○, leaves a string of reversed spins (×) behind. Retracing the original path restores the Néel order. (b) The simplest retracing path [of order $(-t)^2$], from the site of injection to a nearest neighbor and back, in our graphical notation. (c) A retracing path to a next-nearest-neighboring site. (d) This graph is forbidden, since a spin excitation and a hole cannot be on the same site; more precisely, it is canceled by other connected graphs (see text).

paths which somehow *erase* the strings will contribute to G. The simplest paths of this sort are *self-retracing*, i.e., the hole reverses its motion at some point and retraces its original path back to the origin. The self-retracing path approximation includes only such paths, and, moreover, neglects the fact that certain paths close on themselves (i.e., a Cayley-tree approximation is made). We will briefly show how to handle this approximation, which has been considered in detail by Brinkman and Rice[18] and later Shraiman and Siggia,[35] both to illustrate our approach and to pave the way for the following subsection.

If only self-retracing paths are allowed, then the Green's function for a hole is site diagonal—that is, $G_{ij}=\delta_{ij}G(t)$. As a result, the Fourier transform $G(\mathbf{k},\omega)$ is independent of $\mathbf{k}$ and there cannot be any coherent hole motion. That is, even though it is possible for a peak to exist in the spectral weight, any such peak will be independent of $\mathbf{k}$ and hence will not represent hole propagation.

In our graphical language, self-retracing paths are of the form shown in Fig. 7(b). The case shown there is a piece of

$$iG(t)=\frac{(-i)^2}{2}\int dt_1\int dt_2\langle \mathcal{T}h_0(t)T_K(t_2)T_K(t_1)h_0^\dagger(0)\rangle \ . \tag{52}$$

Using the definition of the kinetic energy T_K in Eq. (4), the only nonzero contribution to the above is

$$iG(t)=(-i)^2(-t)^2z\int dt_1\int dt_2\langle h_0(t)[d_1d_1^\dagger h_1d_0h_0^\dagger h_0](t_2)[h_0^\dagger h_0d_0^\dagger h_1^+d_1d_1^\dagger](t_1)h_0^\dagger(0)\rangle \ . \tag{53}$$

For the graph in Fig. 7(b), which is the only nonzero $O(t^2)$ graph, this gives

$$G^{[1]}(t)=i^2zt^2\int dt_1\int dt_2[G^{(0)}(t_2-t_1)]^3G^{(0)}(t_1)G^{(0)}(t_1-t_2) \ , \tag{54}$$

from which we find the on-site self-energy

$$\Sigma^{(1)}_{\rm retr}(\omega)=i^2zt^2\int\frac{d\omega_1}{2\pi}\int\frac{d\omega_2}{2\pi}G^{(0)}(\omega_1)G^{(0)}(\omega_2)G^{(0)}(\omega-\omega_1-\omega_2)=zt^2\frac{1}{\omega-3\varepsilon_0/2+i\delta} \ . \tag{55}$$

Here the fact of z occurs because the hole can move this way to z nearest-neighboring sites. The sign is perhaps easiest to find by direct contraction of the matrix element in Eq. (53). It is also easy to evaluate the graph in Fig. 7(b) using the rules of Appendix D. The ordinary Dyson's sum based on the graph in Fig. 7(b) corresponds to the hole moving from its point of injection to a neighboring site and then back again, and repetitions of this. This gives $[G^{(1)}(\omega)]^{-1}=[G^{(0)}(\omega)]^{-1}-\Sigma^{(1)}_{\rm retr}(\omega)$.

Now consider the two-step processes in which the hole moves from the neighboring site to a next-nearest-neighbor site before retracing [see Fig. 7(c)]. In between it creates a string of length two. When the hole moves out and back, it first creates a string [the double lines in Fig. 7(c)] and then destroys the string in reverse order. This is why the self-retracing-path approximation might also be called a "noncrossing" approximation. There are only $z-1$ such NNN sites, since a return to the origin is already included in the Dyson sum above. The additional path returning to the origin shown in Fig. 7(d) is forbidden because it is not possible to have both a spin excitation and a hole on the same site. In our graphical representation, this is accomplished by the constraint loop on the kinetic energy vertex. In fact, this graph cancels others in which the lines on the site of insertion are connected differently. The easiest way to see this is to recognize that Fig. 7(d) is a connected piece in the Wick's expansion of a vanishing matrix element; and because the disconnected piece arising from the expansion vanishes, the sum of connected pieces must vanish also.

Allowing the hole to oscillate between the neighboring site and any next-nearest-neighbor site any number of times amounts to replacing the upper (internal) $G^{(0)}$ line in Fig. 7(b) by

$$G^{(0)}(\omega)\to\frac{1}{\omega-\varepsilon_0/2+i\delta-\dfrac{(z-1)t^2}{\omega-3\varepsilon_0/2+i\delta}} \ . \tag{56}$$

Except for the replacement of z by $z-1$, this is precisely the form of the $G^{(1)}$ derived from $\Sigma^{(1)}$. Inserting this into Eq. (55) gives

$$\Sigma^{(2)}_{\rm retr}=\frac{zt^2}{\omega-3\varepsilon_0/2+i\delta-\dfrac{(z-1)t^2}{\omega-5\varepsilon_0/2+i\delta}} \ . \tag{57}$$

Continuing in this way gives a continued fraction representation of the Green's function in the self-retracing-paths approximation.[18,35]

$$\Sigma^z_{\rm retr}=\cfrac{zt^2}{\omega-3\varepsilon_0/2+i\delta-\cfrac{(z-1)t^2}{\omega-5\varepsilon_0/2+i\delta-\cfrac{(z-1)t^2}{\omega-7\varepsilon_0/2+i\delta+\cdots}}} \ , \tag{58}$$

which is equivalent to

$$\Sigma^z_{\rm retr}(\omega)=\frac{zt^2}{\omega-3\varepsilon_0/2+i\delta-\frac{z-1}{z}\Sigma^z_{\rm retr}(\omega-\varepsilon_0)} . \quad (59)$$

Here we have made the usual Cayley tree approximation—i.e., assumed that at each step subsequent to the first there are $z-1$ equivalent sites to which the hole can move, and neglecting the possibility of loops. (We will study the contribution of loops in the next subsection.) The expression in Eq. (59) is not quite the correct result for the Ising limit, because we have only treated the Ising interaction in mean-field theory. However, as explained in Appendix E, it is easy to dress these graphs to infinite order in H_z, and the complete result in the Ising limit is

$$\Sigma^z_{\rm retr}(\omega)=z\Sigma_{\rm retr}(\omega-3\varepsilon_0/2+J_z/2) ,$$
$$\Sigma_{\rm retr}(\omega)=\frac{t^2}{\omega+i\delta-(z-1)\Sigma_{\rm retr}[\omega-(z-2)J_z/2]} . \quad (60)$$

Both Eqs. (59) and (60) give spectral weights which are a sequence of δ functions. These represent the discrete bound states formed by the linear confining string potential.[35]

B. Propagation via loop graphs

Besides the self-retracing paths, there are other paths in which, even in the Ising limit, the string of overturned spins is erased. These paths permit the hole to propagate and so give a nonzero bandwidth to, say, the lowest peak in the hole spectral weight. We call these loop graphs. The lowest-order loop graph is shown in Figs. 8(a) and 8(b). There the hole moves around the elementary plaquette $1\frac{1}{2}$ times (six steps), ending up diagonally across from its starting site. This we will call the plaquette graph.

This process can easily be included using our formalism. It turns out, however, that if this path alone is incorporated, the resulting approximation to the Green's function is nonconserving—i.e., the spectral weight is sometimes negative. But a conserving approximation can be regained by mapping the step in Fig. 8(a) to a particular tight-binding problem. In the process one naturally discovers that necessarily a second path—the loop in Fig. 8(c)—must also be included. (This is necessary in order to avoid a vast overcounting of the self-retracing paths leading to nearest-neighbor sites, which would result in too low a band edge.) Because the resulting approximation is equivalent to a tight-binding model, albeit a complicated one, we are guaranteed non-negative spectral weights. The calculation is simple but rather unusual, so we have outlined it here in some detail for the particular case of a square lattice in two dimensions. A more intuitive understanding of the loop graphs, and the results of our calculation, are all collected below.

The bare plaquette graph given in Fig. 8(a) is shown in our graphical notation in Fig. 9(a). Writing this in a more conventional form, as in Fig. 9(b), shows that the plaquette graph represents the lowest vertex correction to the simplest self-retracing path given in Fig. 7(b). We can easily evaluate the bare plaquette graph in Fig. 9(a) using the rules of Appendix D. They give

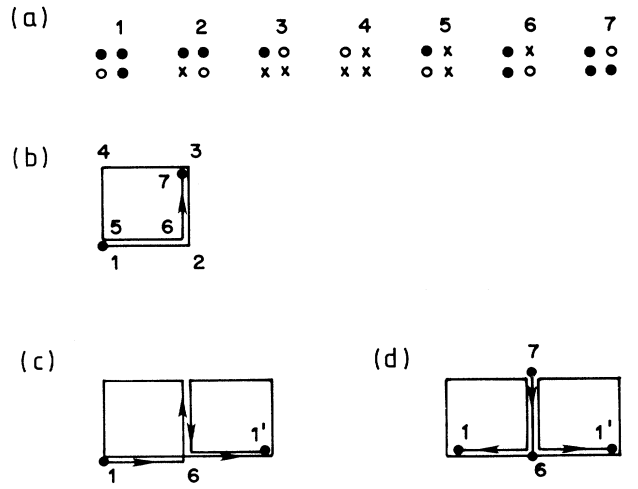

FIG. 8. (a) The plaquette path. The moving hole is represented by an ○, a reversed spin by a ×. When the hole moves $1\frac{1}{2}$ times around the plaquette, it moves diagonally across the plaquette while cleaning up its string. (b) The plaquette path shown schematically. (c) A longer loop path which permits the hole to move to a second-nearest-neighbor site on the same sublattice, while cleaning up its string. (d) Two plaquette paths which end up at different sites are joined in the middle.

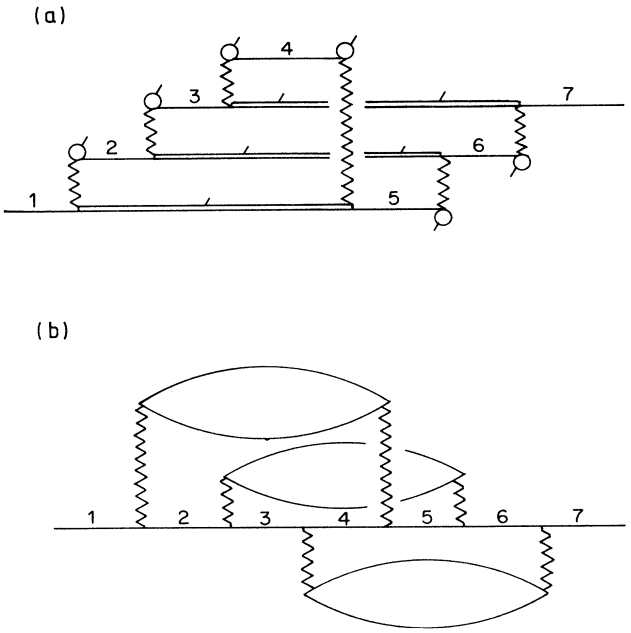

FIG. 9. (a) The plaquette path in our graphical notation. This is the direct path in which the hole moves from site 1 to site 7 without any other retracings. The labeling of the sites is the same as in Fig. 8. (b) The same diagram as (a) drawn more conventionally. Here the hole lines have been pulled together at the interaction node, and the spin lines have been separated by the interaction line. This illustration shows that the plaquette path gives the lowest-order vertex correction for the kinetic energy vertex.

$$\Sigma_{ij}^{(\text{bare})}(\omega)=(-t)^6\left[\frac{1}{\omega-3\varepsilon_0/2+i\delta}\right]^2\left[\frac{1}{\omega-5\varepsilon_0/2+i\delta}\right]^2\left[\frac{1}{\omega-7\varepsilon_0/2+i\delta}\right], \tag{61}$$

where i and j must be on opposite corners of a plaquette. Now, this graph by itself is not a conserving approximation. We will show how a conserving approximation can be made by building from this bare plaquette graph a type of effective tight-binding model which incorporates this process plus others. Among the additions are an extra loop graph, needed to make the approximation conserving. We can also completely dress the graph with H_z and also add all other retracings of the type discussed in the previous subsection.

To see how to handle the loop graph in Fig. 8(a), and its extensions, let us first consider a much simpler one-dimensional problem. Figure 10 shows an infinite chain with three sublattices (a one-dimensional Bravais lattice with a 3-site basis), with on-site energies $\varepsilon_0,\varepsilon_x,\varepsilon_x$. Suppose we want to integrate out the ε_x sublattices and find a Green's function for propagation of a particle along the ε_0 sublattice. We do so in two steps, first focusing on a 4-site segment $\varepsilon_0-\varepsilon_x-\varepsilon_x-\varepsilon_0$, and later restoring the complete chain. The Hamiltonian describing the complete 1D chain is

$$H=-t\sum_{\langle ij\rangle}c_i^\dagger c_j+\sum_i\varepsilon_i c_i^\dagger c_i\ . \tag{62}$$

First consider the segment consisting of the four numbered sites in Fig. 10. Suppose a particle starts at site 1. There are two distinct types of graphs we must consider—paths which return to 1 without ever reaching 4, and paths which go from 1 to 4 without ever returning to 1 and which stop as soon as 4 is first reached. This exhausts all possible motions without double counting. These paths are described by self-energies Σ_{11} and Σ_{14}. Σ_{11} arises from the self-retracing paths 1-2-1,1-2-1-2-1, 1-2-3-2-1, etc., evaluated just as in the previous subsection; then

$$\Sigma_{11}(\omega)=\frac{t^2}{\omega-\varepsilon_x-\dfrac{t^2}{\omega-\varepsilon_x}}\ . \tag{63}$$

Σ_{14} represents paths from 1 to 4 without double counting any of the self-retracing paths already included in Σ_{11}. That is, once the particle leaves 1 it does not return, nor, once 4 is reached, does the particle leave. This possible paths are 1-2-3-4, 1-2-3-2-3-4, etc., but not 1-2-1-2-3-4.

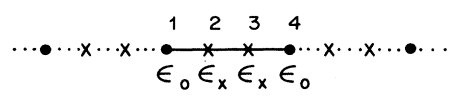

FIG. 10. A one-dimensional tight-binding model with three sublattices. One sublattice has on-site energies ε_0, and the other two have ε_x.

This simplest path, 1-2-3-4, gives

$$\Sigma_{14}^{(0)}=(-t)\frac{1}{(\omega-\varepsilon_x)}(-t)\frac{1}{(\omega-\varepsilon_x)}(-t)\ . \tag{64}$$

Allowing self-retracing paths of the form 2-3-2 changes this to

$$\Sigma_{14}(\omega)=(-t)\frac{1}{\omega-\varepsilon_x-\dfrac{t^2}{\omega-\varepsilon_x}}(-t)\frac{1}{\omega-\varepsilon_x}(-t)\ . \tag{65}$$

The two self-energies Σ_{11} and Σ_{14} were obtained for the 4-site cluster. Using these it is easy to write down the self-energy for propagation among ε_0 sites on the infinite chain:

$$\Sigma_{ij}(\omega)=2\Sigma_{11}\delta_{ij}+\Sigma_{14}(\delta_{i,j+3}+\delta_{i,j-3})\ . \tag{66}$$

Here the particle can start from i and return to i by initially moving either left or right; this gives the factor of 2 in the first term. Or, the particle can move to a neighboring ε_0 site three steps away to the left or right. Every possible path for the particle moving on the ε_0 sublattice is built up from repetitions of these basic paths:

$$G_{ij}=G^{(0)}\delta_{ij}+G^{(0)}\Sigma_{ij}G^{(0)}+G^{(0)}\Sigma_{ik}G^{(0)}\Sigma_{kj}G^{(0)}+\cdots\ . \tag{67}$$

Fourier transforming this and Eq. (66) over the ε_0 sites (the closed circles in Fig. 10) yields

$$G(k,\omega)=\frac{1}{\omega-\varepsilon_0+i\delta-\Sigma(k,\omega)}\ , \tag{68}$$

where

$$\Sigma(k,\omega)=2\Sigma_{11}(\omega)+2\Sigma_{14}(\omega)\cos(k)\ ,$$

with the ε_0-ε_0 separation set to unity. This is the exact Green's function for propagation within the ε_0 sublattice. One can check that the poles of this Green's function exactly reproduce the eigenenergies of the tight-binding model. Moreover, the spectral weight $-\pi^{-1}\mathrm{Im}G(k,\omega)$ will necessarily be non-negative, since it is the exact Green's function for a Hamiltonian.

A procedure very similar to this solves the plaquette path in Fig. 8(a). However, the problem in two dimensions is a bit harder. In the one-dimensional example just solved, replicas of the basic paths described by Σ_{11} and Σ_{14} join only at the end points—that is to say, only on ε_0 sites. (For example, the particle could move 1-2-1-2′-1 where 2′ is to the left of 1; the two retracing paths are connected only at the original site 1.) But in the two-dimensional problem built up from Fig. 8(a) on the square lattice, equivalent paths also can be joined at interior points. This is illustrated in Fig. 8(d). There, two pla-

quette loops, which end up at different final sites, share the same site 6. If this overlap is ignored, one is led to an approximation that greatly overcounts the self-retracing paths to nearest neighbors.

A way to handle this overlap is to draw an equivalent but expanded version of our original square lattice, including every possible plaquette loop. The result, an effective tight-binding model in spin-configuration space, is shown in Fig. 11(a). Suppose that we inject the hole on sublattice A at the site labeled 1. Then the plaquette graph allows the hole to move among sublattice A sites. In Fig. 11(a), each site at the center of a cross represents a state in our original problem consisting of a hole on sublattice A with no reversed spons. Each site at the point of a cross represents a state in which the hole has moved to a neighboring site, leaving a reversed spin on the original sublattice A site. The four points of each cross represent the four ways the hole can move to neighbors in two dimensions. In general, the numbering of sites in Fig. 11(a) corresponds in this way to the same numbers in Figs. 8(a) and 8(b) and Fig. 9(a); that is, the numbers in Fig. 11(a) represent both a hole position *and a particular spin configuration* in the original square lattice. Hopping from 1 to 7 in Fig. 11(a) is equivalent to the hole motion given by the plaquette loop.

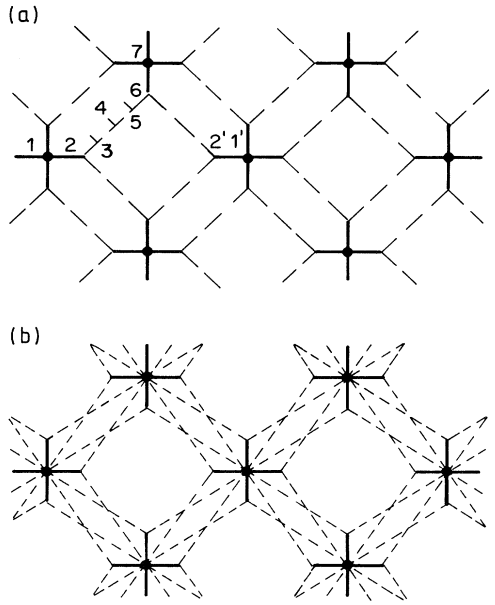

FIG. 11. (a) An equivalent representation of the plaquette graph, applied everywhere in the plane, as a tight-binding model in the space of hole position and spin configuration. The numbering here corresponds to that in Fig. 8(a). If the particle here is at the center of a cross, the corresponding state in the original problem has the hole on sublattice A, with no spins reversed. Site 2 here corresponds to the hole moving one step in the original problem, leaving a reversed spin behind. Thus motion along the path 1-2-3-4-5-6-7 here is identical to the plaquette path. All hops have amplitude $-t$. (b) An equivalent representation of the string-erasure ($J_\perp$) graph described in Sec. V. The sites on the cross are the same as in (a). The solid lines here represent hops with amplitude $-t$, and the dashed lines are hops with amplitude $(-t)J_\perp/(-2\varepsilon_0+J_z)$. Stepping from 1 to 2 to 7 [in the notation of (a)] represents the string-erasure process shown in Fig. 14. To avoid cluttering, we have left off one set of dashed lines which connect 1 to 2′ and all equivalent pairs.

In Fig. 11(a) each line connecting two sites represents a hop with amplitude $-t$. Each site has an on-site energy equal to the Ising energy of the corresponding hole and string configuration in Fig. 8(a). Thus, sites 1 to 7 have on-site energies, respectively, $\varepsilon_1,\varepsilon_2,\varepsilon_3,\ \varepsilon_4,\varepsilon_5,\varepsilon_6,\varepsilon_7$, where ε_n is the energy of a hole plus its string of $n-1$ reversed spins:

$$\begin{aligned}&\varepsilon_7=\varepsilon_1=\varepsilon_0/2,\quad \varepsilon_6=\varepsilon_2=\varepsilon_0/2+3J_z/2\ ,\\ &\varepsilon_5=\varepsilon_3=\varepsilon_0/2+5J_z/2\ ,\\ &\varepsilon_4=\varepsilon_0/2+6J_z/2\ .\end{aligned}\tag{69}$$

Here we have included the effect of complete dressing with H_z [compare this with the energies in Eq. (61)].

Thus we have mapped our original problem to an equivalent tight-binding model in configuration space. We are interested in propagation among sites at the centers of the crosses in Fig. 11(a); this problem can be solved exactly. The solution will include retracings among the sites in Fig. 11(a) (e.g., 1-2-3-2-1). This, however, only includes some but not all of the self-retracing paths described in the previous subsection. In fact, *all* of the self-retracing paths can be added to the lattice in Fig. 11(a), but for clarity we will postpone them until later.

The tight-binding problem in Fig. 11(a) is solved like the one-dimensional example given above. There we needed to consider a 4-site segment of sites. Here we need to consider the 16-site diamond in the center of 11(a), and compute G'_{22}, G'_{26}, and $G'_{22'}$ propagators for motion on the diamond only. In terms of these, the self-energies for propagation on the A sublattice (i.e., between cross centers) are

$$\Sigma_{11}=zt^2G'_{22},\quad \Sigma_{17}=2t^2G'_{26},\quad \Sigma_{11'}=t^2G'_{22'}\ .\tag{70}$$

(Hereafter primed quantities Σ',G' describe motion on the 16-site diamond, and unprimed quantities are for motion on the infinite lattice.) Here Σ_{11} represents every way the hole can leave site 1 and return to it without reaching any other cross center. During the motion, the hole can move freely around a diamond. Σ_{17} represents every way the hole can go from 1 to 7 without returning to 1 and without reaching any other cross center. $\Sigma_{11'}$ is similar.

Then the complete self-energy will be

$$\begin{aligned}\Sigma_{ij}^{\text{loop}}=\Sigma_{11}\delta_{ij}&+\Sigma_{17}\sum_{\mathbf{r}=\pm\hat{\mathbf{x}}\pm\hat{\mathbf{y}}}\delta_{\mathbf{R}_j,\mathbf{R}_i+\mathbf{r}}\\ &+\Sigma_{11'}\sum_{\mathbf{r}=\pm2\hat{\mathbf{x}},\pm2\hat{\mathbf{y}}}\delta_{\mathbf{R}_j,\mathbf{R}_i+\mathbf{r}}\ .\end{aligned}\tag{71}$$

Fourier transforming, the complete Green's function for the propagation of a hole among A-sublattice sites is

$$G(\mathbf{k},\omega)=\frac{1}{\omega-\varepsilon_1+i\delta-\Sigma(\mathbf{k},\omega)}\ , \tag{72}$$

where

$$\Sigma(\mathbf{k},\omega)=\Sigma_{11}(\omega)+4\Sigma_{17}(\omega)[\cos(k_x)\cos(k_y)] +2\Sigma_{11'}(\omega)[\cos(2k_x)+\cos(2k_y)]\ . \tag{73}$$

Here using Eq. (70) makes this the complete self-energy for the plaquette graph placed onto the tight-binding lattice.

Thus the problem has been reduced to solving for G'_{22}, G'_{26}, and $G'_{22'}$ on the 16-site diamond. This is a cyclic 1D ring, and we solve for its Green's functions in turn by pulling the smallest unit out of it: consider the cluster consisting of a single side of the diamond, i.e., the five sites 2-3-4-5-6. For this 5-site segment we need two self-energies: Σ'_{22}, representing all self-retracings from site 2 back to itself (without reaching site 6); and Σ'_{26}, representing all paths from 2 to 6 that never return to 2 and that stop as soon as 6 is reached. Just as in the 1D example above, these are [with reference to Eq. (63) and Eq. (65)]

$$\Sigma'_{22}=\cfrac{t^2}{\omega-\varepsilon_3-\cfrac{t^2}{\omega-\varepsilon_4-\cfrac{t^2}{\omega-\varepsilon_5}}}\ ,\quad \Sigma'_{26}=(-t)\cfrac{1}{\omega-\varepsilon_3-\cfrac{t^2}{\omega-\varepsilon_4-\cfrac{t^2}{\omega-\varepsilon_5}}}(-t)\cfrac{1}{\omega-\varepsilon_4-\cfrac{t^2}{\omega-\varepsilon_5}}(-t)\cfrac{1}{\omega-\varepsilon_5}(-t)\ . \tag{74}$$

In k space, the Green's function for the hole motion around the corners of this 1D diamond takes the form

$$G'(k,\omega)=\frac{1}{\omega-\varepsilon_2+i\delta-2\Sigma'_{22}-(2\cos k)\Sigma'_{26}}\ . \tag{75}$$

The allowed discrete k values are $-\pi/2$, 0, $\pi/2$, and π. Using the formulas for the inverse Fourier transform,

$$\begin{aligned} G'_{22}&=\tfrac{1}{4}[G'(-\pi/2,\omega)+G'(0,\omega) +G'(\pi/2,\omega)+G'(\pi,\omega)]\ ,\\ G'_{26}&=\tfrac{1}{4}[G'(0,\omega)-G'(\pi,\omega)]\ ,\\ G'_{22'}&=\tfrac{1}{4}[-G'(-\pi/2,\omega)+G'(0,\omega) -G'(\pi/2,\omega)+G'(\pi,\omega)]\ , \end{aligned} \tag{76}$$

we find

$$\begin{aligned} G'_{22}&=\frac{g^{-2}-2(\Sigma'_{26})^2}{g^{-2}-4(\Sigma'_{26})^2}g\ ,\\ G'_{26}&=\frac{\Sigma'_{26}}{g^{-2}-4(\Sigma'_{26})^2}\ ,\\ G'_{22'}&=\frac{2(\Sigma'_{26})^2 g}{g^{-2}-4(\Sigma'_{26})^2}\ , \end{aligned} \tag{77}$$

where

$$g=\frac{1}{\omega-\varepsilon_2+i\delta-2\Sigma'_{22}}\ . \tag{78}$$

Finally, the Greens' function for hole motion on the A sublattice, based on the plaquette loop, dressed with H_z, is given by Eq. (72), using Eqs. (73), (70), (77), and (74).

In this development we have included certain self-retracing paths, those which move around a single plaquette. But this is only a negligible fraction of the infinitely many self-retracing paths described earlier. However, it is very easy to add all other possible self-retracing paths, as long as they are treated in the Cayley-tree approximation. From any site equivalent to 2, a hole can move in one extra direction [besides those already included in Fig. 11(a)], and then the hole has $z-1$ choices at each further step. From sites 3, 4, or 5, the hole can go $z-2$ extra ways initially and $z-1$ ways thereafter. Thus in Eqs. (74) and (78) make the replacements

$$\begin{aligned} &\varepsilon_n\rightarrow\tilde{\varepsilon}_n=\varepsilon_n+\Sigma_{\rm retr}(\omega-\varepsilon_n),\quad n=2,6\ ,\\ &\varepsilon_n\rightarrow\tilde{\varepsilon}_n=\varepsilon_n+(z-2)\Sigma_{\rm retr}(\omega-\varepsilon_n),\quad n=3,4,5\ , \end{aligned} \tag{79}$$

where $\Sigma_{\rm retr}$ is given in Eq. (60) and ε_n in Eq. (69). This has the effect of hanging a Cayley tree off of each site (except for cross centers) in Fig. 11(a).

This, then, is our final result for the Ising limit: it combines the plaquette graph of Fig. 8(a) plus the self-retracing paths, dressing everything completely with H_z. The end result is that the self-retracing-path approximation has been partially corrected—certain paths which were erroneously treated in the Cayley-tree approximation have now correctly been identified as loops which let the hole move.

C. Results and interpretation

First let us point out that the spectral weight $-\pi^{-1}\mathrm{Im}G(\mathbf{k},\omega)$ computed from Eq. (72) is necessarily positive definite; i.e., the approximations made are conserving, because they are equivalent to a Hamiltonian—the tight-binding model in configuration space, in Fig. 11(a). Even when the remaining self-retracing paths are included, this is still true—in Fig. 11(a), each site except for those at the centers of crosses now becomes the root of a Cayley tree extending outward. The result is still a tight-binding model.

Let us look again at the tight-binding model. A path from one cross center to another along a diagonal in Fig. 11(a) is identical to the hole moving along the plaquette path in Fig. 8(a). However, it is also possible to move from site 1 to 2 to 6 to 2′ to 1′ in Fig. 11(a). This turns out to be equivalent to the larger loop path shown in Fig.

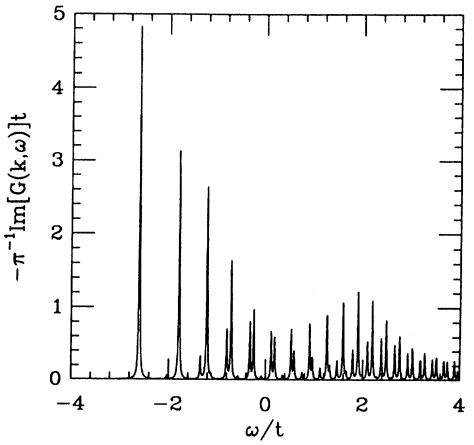

FIG. 12. An example of a spectral weight in the Ising limit arising from the loop paths, or, equivalently, the model in Fig. 11(a). Here $J_z=0.2$, $\mathbf{k}=(0,0)$, and the small imaginary bit $\delta=0.01$. The peaks are bound states in the confining potential produced by the string. Here the loop path lets the hole move, so, e.g., the lowest peak has dispersion.

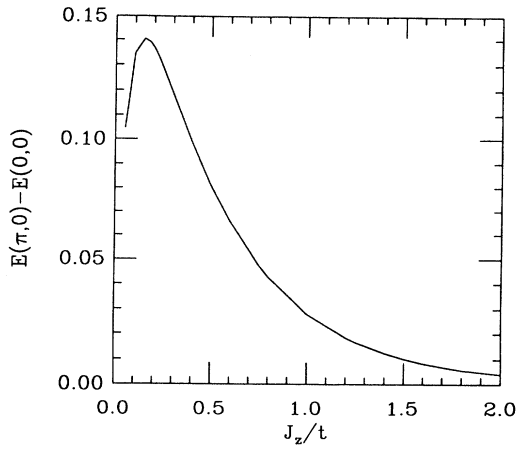

FIG. 13. Bandwidth in the Ising limit. For large J_z, the bandwidth decreases exponentially. Our approximations break down below $J_z \approx 0.2$; for smaller values, longer loop paths become important.

8(c). Thus, to get a conserving approximation (to have an approximation equivalent to some Hamiltonian) we needed to include both paths. This is something like the usual Ward's identity in which inclusion of one graph requires inclusion of another.

An example of the spectral weight computed in the Ising limit is shown in Fig. 12. The many sharp peaks represent the bound states of the hole, as in the previous subsection. Here, however, the peaks are dispersive. Let us focus on the lowest-energy peak; it is now a real quasiparticle peak, with a nonzero bandwidth provided by the loop graphs. The band minimum turns out to occur at $\mathbf{k}=(0,0)$, and the maximum at $\mathbf{k}=(\pi,0)$ or $(0,\pi)$ [because Σ_{17} is negative and dominates $\Sigma_{11'}$ in Eq. (73)]. We have plotted the difference between these, the bandwidth $W=\omega_{(0,\pi)}-\omega_{(0,0)}$, in Fig. 13. The qualitative features of the bandwidth are that it falls rapidly at large J_z, growing with decreasing J_z until a peak at $J_z \approx 0.2$ and then falling.

The large-J_z features of this bandwidth are easily understood in terms of effective one-dimensional tunneling. As the hole moves around the plaquette path in Fig. 8(a), the Ising energy of its string steadily grows until the hole reaches site 4, and then drops again. Shraiman and Siggia pointed out that the hole motion can be thought of as effectively one dimensional, with the length of the string acting, in a continuum approximation, as a Cartesian coordinate.[35] Here we see another example of that: for the hole to move around the plaquette loop, it must tunnel through the one-dimensional barrier created by the string's Ising energy. This picture is even clearer in the equivalent tight-binding model, in Fig. 11(a). There the particle must hop along the diagonal sites, which have a sawtooth on-site energy (rising up to a maximum $\sim 4J_z$ at site 4 and then dropping). This barrier gives an exponential suppression to the amplitude for the effective hop from site 1 to 7, and so the bandwidth must be exponentially small in J_z for large J_z. The continuum approximation,[35] applied to this sawtooth barrier, gives a tunneling amplitude of order $\exp[-5.3(J_z/t)^{1/2}]$. This gives rise to an effective hopping amplitude of the same order and hence controls the bandwidth. And, in fact, the large-J_z bandwidth for $J_z > t$ is well fit by $\omega_{(0,0)}-\omega_{(\pi,0)}=-3.0\exp[-4.6(J_z/t)^{1/2}]$, in reasonable agreement with the continuum approximation.

As J_z shrinks, on the other hand, the barrier height falls, the hole can propagate more freely, and the bandwidth grows. This is seen down to $J_z \approx 0.2$ in Fig. 13. What happens to suddenly stop this growth at this point? Something similar to a finite-size effect (even though we are working on an infinite lattice). As J_z falls, indeed it becomes highly probable that the hole will tunnel through the barrier and so traverse the plaquette path. However, at small J_z it can also go very far down the Cayley tree before it is forced by the Ising energy to retrace its path. Essentially, the hole gets lost in the infinitely many possible self-retracing paths, and cannot find its way around the plaquette. However, the vanishing of W seen in Fig. 13 is unphysical, a consequence of only considering the smallest loop paths.

The two loop paths shown in Figs. 8(b) and 8(c) are only the two smallest of an infinite set of related loop paths which permit the hole to move, while cleaning up its string. As long as J_z is sufficiently large, only these two will matter—longer paths will have higher barriers, and will be exponentially insignificant relative to the dominant plaquette graph. More quantitatively, the expected length of the strings in the ground state, in the continuum self-retracing approximation, is $1.87(J_z/t)^{-1/3}$.[35] When this length exceeds 3, it becomes very likely that the plaquette path will be traversed; this occurs at $J_z/t \approx 0.24$. However, as this expected length

becomes larger, then longer and longer loop paths will begin to contribute to the hole's propagation. The collapse of the bandwidth in Fig. 13 at $J_z \approx 0.2$ is a result of the expected string length growing to 3. In reality, at smaller J_z the longer loop paths (which we neglect) will begin to contribute to the propagation and hence the bandwidth. This becomes a geometric question—as J_z shrinks and the expected string length grows, the hole can move further down the Cayley tree; but simultaneously more and more loop paths can help it propagate. If the latter growth wins, it will keep the bandwidth increasing as J_z shrinks, to some finite value at $J_z=0$. If the dead ends dominate, then the bandwidth will vanish as $J_z \to 0$. Which is the case is not known. In any event, our results are only accurate down to $J_z \approx 0.2$. We will compare our results to those of exact diagonalizations in the next section.

The picture which develops from this analysis is the following. The basic excitation is the hole dressed by a cloud of disturbed spins; here this cloud results from the hole plus the many self-retracing paths which it can follow. This composite object can then propagate by a kind of one-dimensional tunneling through a barrier created by the Ising energy.

V. CHARGE EXCITATIONS NEAR THE ISING LIMIT

When $J_\perp$ is switched on, the string of reversed spins left by a moving hole can relax. This provides the charge a second way to propagate, distinct from the loop paths just treated. In earlier analytical studies of the t-J model (including our own), only this string erasure process was considered. In this section we will carefully examine this process near the Ising limit (i.e., for $J_\perp \ll J_z$).

One interesting result is that this erasure process and the loop path, both of which let the hole move, actually interfere and cause a level crossing. Consequently, as $J_\perp$ grows, the band minimum for the lowest-lying quasiparticle switches from $\mathbf{k}=(0,0)$ to $\mathbf{k}=(\pi/2,\pi/2)$. This interference explains some unusual features of numerical calculations.

There have been very few controlled calculations of hole properties in the t-J model; much of our understanding has resulted from small-cell numerical studies. One interesting result found in numerical work was that $\omega_{(0,0)}-\omega_{(\pi/2,\pi/2)}$ (where $\omega_{\mathbf{k}}$ is the energy at momentum $\mathbf{k}$) vanishes at a small but nonzero J in the Heisenberg model. However, this could not with certainty be ascribed either to real physics or to a finite-size effect [the value of J at which this occurs is small, and not too far from the value where the crossover to a ferromagnetic state (the Nagaoka effect) occurs]. Our calculation provides the first indication that this effect could be real, a consequence of the interference between two channels of propagation.

The picture that results here is similar to that of the previous section. The basic quasiparticle is the dressed hole, i.e., the hole plus a cloud of reversed strings. The cloud is given by the hole's virtual but retraced excursions. The composite object is allowed to move, now by string erasure as well as loops.

An interesting question is whether this picture survives in the Heisenberg limit. There the spin excitations become gapless, and one might imagine that the strings might effectively have zero energy, and perhaps could be neglected. We will not directly address this question here; whatever rigor is possessed by perturbation theory breaks down in this limit. Nonetheless, we have shown in earlier work that, using physically reasonable approximations, the same picture in fact does persist.[23] Even in the Heisenberg limit, that is, there is a low-lying quasiparticle peak given dispersion by the string erasure ($J_\perp$) process. The reason the low-lying peak must exist is simple. The string problem can be viewed as 1D motion in a linearly confining potential.[35(a)] As long as the gapless spin waves do not completely eliminate this confinement, then as in any 1D problem at least one bound state must exist. This appears as a low-lying peak in the spectral weight. In fact, small-cell calculations show that the string picture is better than this argument requires. In the Heisenberg limit, the numerical studies find, at least for moderate values of J/t, several bound states visible in the spectral weight.[17(g)] Our earlier analytical results, moreover, showed that at least one bound state persisted down to arbitrarily small J.[23]

A. String-erasure graph

The simplest string-erasure process is shown in Fig. 14. There the two flipped spins left by a moving hole are restored to their Néel configuration by the action of $H_\perp$. More precisely, the true ground state in the presence of $H_\perp$ adds fluctuations to the Néel state; it is just such a vacuum fluctuation that here erases the string and unbinds the hole. This basic step gives a self-energy (using the rules of Appendix D)

$$\Sigma_{ij}^{\perp}(\omega)=2t^2\left[\frac{J_\perp}{2}\right]\frac{1}{\omega-3\varepsilon_0/2+i\delta}\,\frac{1}{-2\varepsilon_0}\,c_{\mathbf{R}_i-\mathbf{R}_j}\ . \tag{80}$$

The factor of 2 arises from a second, equivalent graph in which the flipped spins are created before the hole moves (i.e., reflect Fig. 14 both horizontally and vertically).

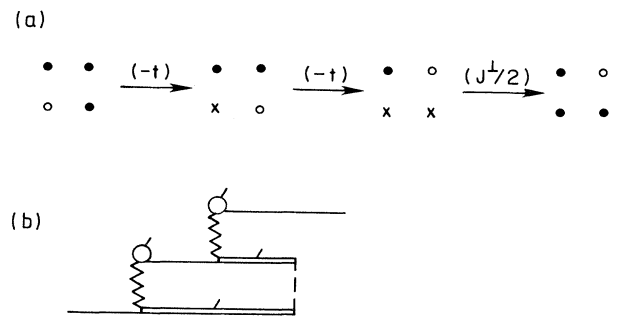

FIG. 14. (a) The basic string-erasure ($J_\perp$) graph. Here the hole hops twice, and the resulting pair of disturbed spins can relax to their Néel configuration by the action of $H_\perp$. (b) The erasure process in our graphical notation. Note that the intermediate state (the hole plus the flipped spin it left behind after the first hop) can be identified as a site labeled 2 or 6 in the effective tight-binding Hamiltonian (see Fig. 11).

Similar to Fig. 2 and Eq. (41), $c_{\pm\hat{x}\pm\hat{y}}=2$ and $c_{\pm 2\hat{x},\pm 2\hat{y}}=1$; but in this case, $c_0=0$. This graph can easily be dressed completely with H_z, as explained in Appendix E. Here we will partially dress with H_z, excluding certain vertex corrections (in which H_z lines connect the outgoing hole line to the double line). Then the frequency factors in Eq. (80) become $(\omega-\varepsilon_2)^{-1}(-2\varepsilon_0+J_z)^{-1}$, where ε_2 is given in Eq. (69). We can also easily include all retracing paths away from the internal line, as we did for internal sites in our calculations of the loop paths. This simply replaces ε_2 by $\tilde{\varepsilon}_2$ as in Eq. (79). Finally, Fourier transforming, the $J_\perp$ string erasure graph gives

$$\Sigma_\perp(\mathbf{k},\omega)=(J_\perp t^2)\frac{1}{\omega-\tilde{\varepsilon}_2+i\delta}\frac{1}{-2\varepsilon_0+J_z}[(z\gamma_{\mathbf{k}})^2-z] , \tag{81}$$

where $\gamma_{\mathbf{k}}$ is defined in Eq. (26). Notice that the $\mathbf{k}$ dependence of this expression, which arises quite naturally in this real-space analysis, was used to fit the numerical results of Stephan.[36]

The basic $J_\perp$ graph gives the hole an alternate route to hop, besides the loop path treated earlier. Like the latter it can be interpreted as a hop on an effective tight-binding Hamiltonian, from which we can then construct a conserving approximation. The corresponding tight-binding Hamiltonian is pictured in Fig. 11(b). There the sites on the crosses are identical to those in Fig. 11(a). The solid lines represent $-t$ hops, and the dashed lines at angles are hops with amplitude

$$-t'=(-t)\frac{J_\perp/2}{(-2\varepsilon_0+J_z)} . \tag{82}$$

We will give the exact Green's function for this tight-binding Hamiltonian below. Some dashed lines have been omitted from Fig. 11(b) to avoid overcrowding—there are $-t'$ hops between sites 1 and 6', and equivalent pairs, as well.

Moreover, it is simple to combine the two process, loop and erasure, as a conserving approximation, while treating both equally. Pictorially this is done by overlaying Figs. 11(a) and 11(b). The final result, the complete self-energy for the hole, including both the loop and erasure channels, is given by Eqs. (72) and (73). But instead of Eq. (70) we have (in 2D)

$$\begin{aligned}
\Sigma_{11}&=z[t^2G'_{22}+tt'(4G'_{26}+2G'_{22'})+t'^2(3G'_{22}+4G'_{26}+2G'_{22'})] ,\\
\Sigma_{17}&=2[t^2G'_{26}+tt'(2G'_{22}+2G'_{26}+2G'_{22'})+t'^2(2G'_{22}+5G'_{26}+2G'_{22'})] ,\\
\Sigma_{11'}&=t^2G'_{22'}+tt'(2G'_{22}+4G'_{26})+t'^2(2G'_{22}+4G'_{26}+3G'_{22'}) .
\end{aligned} \tag{83}$$

The G' are given by Eq. (77). As before, these self-energies come from paths which start on some A-sublattice site and end as soon as another A site is reached. For example, the first of these equations comes from the sum of all paths in which the particle starts at site 1, moves to any of 2, 6, 2' or 6', moves around the diamond to any of 2, 6, 2' or 6' again, and then returns to site 1. An example of the spectral weight arising from this expression is given in Fig. 15.

Let us investigate the limits. When t' is set to zero, so that the erasure graph is switched off, this reduces exactly to the expressions we derived earlier for the loop paths [Eq. (70)]. On the other hand, when J_z grows large, so that the loop process is suppressed, then Σ'_{22} and Σ'_{26} become small and from Eq. (77) we have $G'_{22}\to 1/(\omega-\tilde{\varepsilon}_2+i\delta)$ and $G'_{26},G'_{22'}\to 0$. Then from Eq. (83) we find

$$\begin{aligned}
\Sigma_{11}&=z(t^2+3t'^2)G'_{22} ,\\
\Sigma_{17}&=2\Sigma_{11'}=4(tt'+t'^2)G'_{22} ,
\end{aligned} \tag{84}$$

and hence

$$\Sigma(\mathbf{k},\omega)=\left[\frac{1}{\omega-\tilde{\varepsilon}_0+i\delta}\right]\{z(t^2+3t'^2)+2(tt'+t'^2)[(z\gamma_{\mathbf{k}})^2-z]\} . \tag{85}$$

Here the $\mathbf{k}$-independent terms come from retracings of forms 1-2-1 and 1-6-1. The $\mathbf{k}$-dependent term has a tt' piece equivalent to Eq. (81) (i.e., due to paths of the form 1-2-7), and another piece representing paths of the form 1-6'-7. This last expression is the solution of Fig. 11(b) alone. To develop a conserving approximation using the erasure graph, we have put the graph onto a tight-binding model. The exact Green's function for the latter then teaches us that, to get a conserving approximation, we must include several graphs in addition to the simple

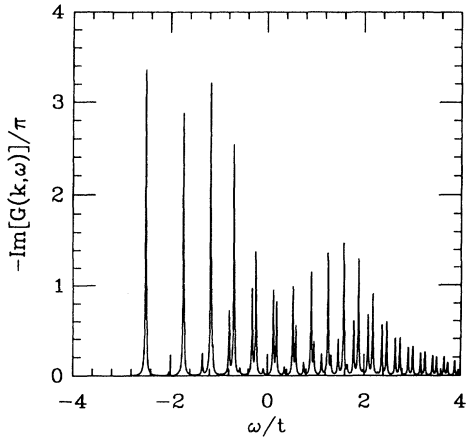

FIG. 15. An example of a spectral weight, including both the plaquette and string-erasure graphs, for $J_z=4J_\perp=0.2$, $\mathbf{k}=(0,0)$, and $\delta=0.01$.

erasure graph. These extra graphs are representations of the extra terms (particularly the **k**-dependent t'^2 term). In the same way, the many contributions to Eq. (83) represent a collection of graphs, including the plaquette and erasure graphs, which combine to give a conserving approximation.

The most interesting property of the hole arising from this analysis is that the two propagation channels turns out to interfere, giving a level crossing in the ground state. This occurs because t and t' have opposite signs. The loop path alone gives a band minimum for the low-lying quasiparticle at $\mathbf{k}=(0,0)$. The $J_\perp$ graph, however, favors $\mathbf{k}=(\pi/2,\pi/2)$ [see Eq. (81)], and so the two channels compete. Let us make this idea quantitative. When both channels are included, the energy $\omega_{\mathbf{k}}$ of the low-lying peak will be a function of both J_z and $J_\perp$ (as well as **k**, of course). For very small $J_\perp/J_z$, the loop path dominates and so $\omega_{(00)}<\omega_{(\pi/2,\pi/2)}$. For large enough $J_\perp/J_z$, conversely, the $J_\perp$ graph wins and $\omega_{(00)}>\omega_{(\pi/2,\pi/2)}$. At some intermediate value of $J_\perp/J_z$, a level crossing occurs in which the band minimum switches from $\mathbf{k}=(0,0)$ to $\mathbf{k}=(\pi/2,\pi/2)$. This crossover value of $J_\perp/J_z$ is shown as a function of J_z in Fig. 16.

At large J_z, the loop path's amplitude is exponentially small in J_z. The erasure graph has an amplitude proportional to $J_\perp$ [Eq. (81)]. Actually, both the loops and the erasures are exponentially small in J_z. But in the former case the hole moves farther and so the constant in the exponent ($\sim$ number of steps) is larger. Hence for large J_z, the erasure graph dominates even for quite small $J_\perp$. This can be seen in Fig. 16. But as J_z shrinks toward our limiting value of $J_z\approx 0.2$, the loop path's amplitude grows and it takes a larger $J_\perp/J_z$ to cross over to the erasure-dominated phase.

These calculations are in very good agreement with, and help explain, small-cell exact diagonalizations. We will compare our results with exact calculations[37] on a 4×4 lattice, the largest size generally treated. Interpreting these results requires one caution: the 4×4 lattice is somewhat special. It turns out that this particular lattice has an extra degeneracy between $\omega_{(\pi/2,\pi/2)}$ and $\omega_{(\pi/0)}$ which must be kept in mind.[17(b)] The "bandwidth" usually quoted for numerical work is $\Delta\omega=\omega_{(0,0)}-\omega_{(\pi/2,\pi/2)}$. In fact, in the infinite system, the band minimum changes from (0,0) to $(\pi/2,\pi/2)$ at some small J_z/t, and the band maximum switches from $(\pi,0)$ to (0,0) at some intermediate value of J_z, so some caution is needed.

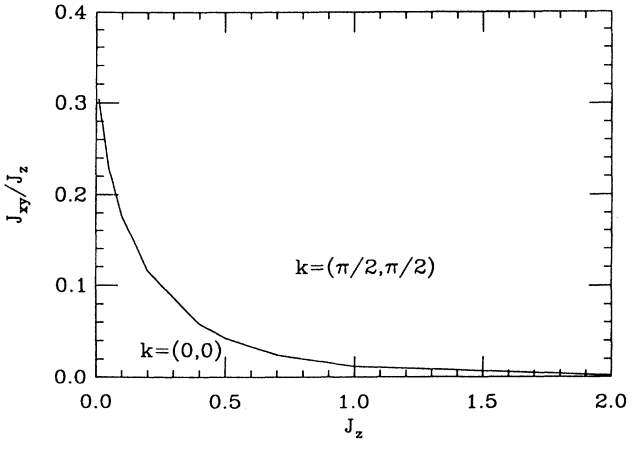

FIG. 16. Crossover values of $J_\perp/J_z$. Above the line, the erasure graph dominates and the hole's band minimum is at $\mathbf{k}=(\pi/2,\pi/2)$. Below the line, the loop paths dominate and the minimum is at $\mathbf{k}=(0,0)$.

In Fig. 17 we have compared several energy differences obtained on the one hand from the low-lying pole in the Green's function obtained using Eq. (83), and on the other from exact diagonalizations on a 4×4 lattice. Shown are results at three values of $J_\perp/J_z$. First consider the pure Ising case, Fig. 17(a). Here the band minimum is at $\mathbf{k}=(0,0)$ and the maximum is at $(\pi,0)$. The lower two curves shown in Fig. 17(a) show a very good agreement between our results and those from exact diagonalization on a 4×4 lattice, for J_z/t greater than 0.4 or so. [The band minimum is at $\mathbf{k}=(0,0)$ and the maximum at $(\pi,0)$ in the infinite system, so the curve marked with diamonds gives the (negative) bandwidth.] In the numerical work, finite-size effects set in when the expected length of a self-retracing path grows to be of order the linear dimension of the cell. In the Ising limit, this length $\sim 2(J_z/t)^{-1/3}$, so roughly one expects calculations on the 4×4 lattice to be valid for $J_z/t>0.2$ or so. (At some sufficiently small J_z, the hole's preference for ferromagnetic order in the finite system drives the spins into the Nagaoka phase.) Similarly, our calculation is valid as long as loop paths longer than those in Fig. 8 can be neglected, which is true at least down to $J_z/t=0.2$. Given the good agreement between our theoretical results and those of diagonalizations, and the understanding of why agreement breaks down at small J_z/t, we are confident that the loop path gives the dominant coherent behavior when $J_\perp=0$.

Consider now the case of a small $J_\perp$. In Fig. 17(b) energies are plotted for different values of J_z/t, with $J_\perp/J_z$ fixed to be $\frac{1}{20}$. Fig 17(c) is similar with $J_\perp/J_z=\frac{1}{4}$. First let us focus on the large-J_z behavior. For large J_z/t, the loop is exponentially suppressed, as described in Sec. IV, and the erasure graph dominates. Any differences between the analytical and numerical results then are a measure of how completely we can describe the string erasure processes using only the low-order graph (Fig. 14) and the partners which join it when the approximation is made conserving. It is evident from Figs. 17(b) and 17(c) that the agreement is quite good.

It is very encouraging that the large-J_z analytical results are in such very good agreement with the numerical bandwidth. We have included only the lowest-order string-erasure process, of order $J_\perp$, and have neglected higher-order erasures of order, e.g., $J_\perp^2$. At $J_\perp/J_z=\frac{1}{4}$, one might then expect our calculation to be in error by about 25%. In fact, the agreement is much better than that—about 3% for this case. We expect, but have not proven, that there is an extra geometric control factor present here, as in the spin case, and that in fact the convergence

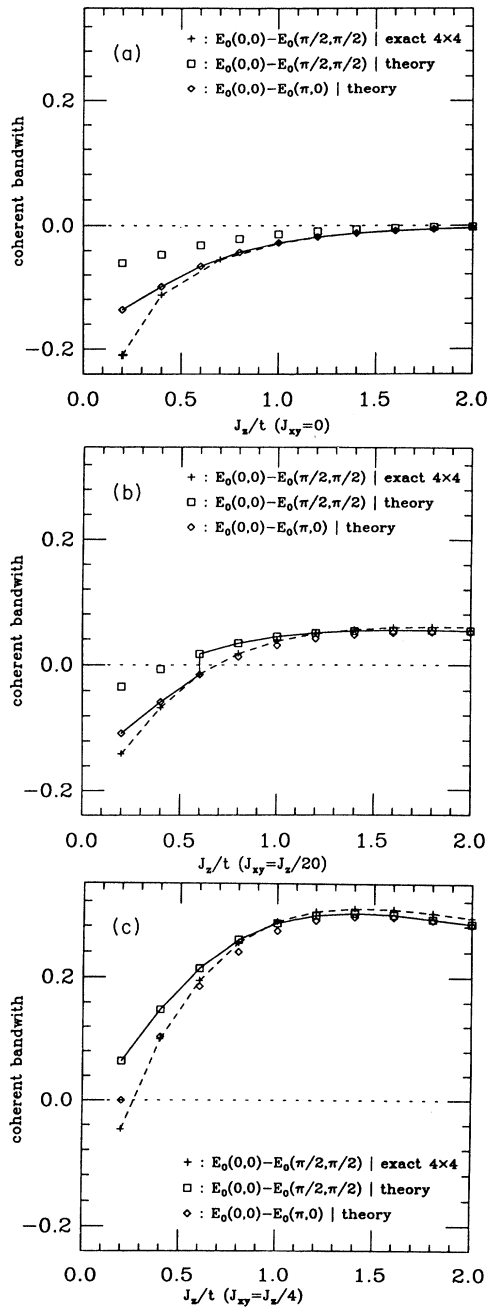

FIG. 17. Differences in energies of the low-lying quasiparticle peak at several values of $\mathbf{k}$, as a function of J_z. The solid lines are the results of our theory, and the dashed lines are obtained by exact diagonalization with one hole on a 4×4 lattice with periodic boundary conditions. Note that the theory contains no adjustable parameters. The theory reproduces some fine structure in the exact diagonalization results, including, e.g., the maximum in (c). (a) Ising limit ($J_\perp \equiv J_{xy}=0$). The energy differences are always negative, since the minimum is at $\mathbf{k}=(0,0)$ and the maximum at $(\pi,0)$. (b) For a constant ratio $J_\perp/J_z=\frac{1}{20}$. Here the band minimum crosses over from $\mathbf{k}=(0,0)$ (for $J_z<0.5$) to $\mathbf{k}=(\pi/2,\pi/2)$ (for $J_z>0.5$). The band maximum changes from $(\pi,0)$ to (0,0) at $J_z=0.7$. (c) For a constant ratio $J_\perp/J_z=\frac{1}{4}$. Note, as for (a) and (b), the very good agreement for large J_z, where the finite-size corrections are small, between theory and numerical data.

is controlled by $J_\perp/(2zJ_z)$.

Let us look more closely at Fig. 17(b). For small $J_z/t \leq 0.5$, the loop process dominates, and so the band minimum is at $\mathbf{k}=(0,0)$ while the maximum is at $(\pi,0)$. Thus in Fig. 17(b) the lower line (diamonds) gives the (negative) bandwidth for small J_z/t. For large $J_z/t>0.7$ the erasure path dominates, the band minimum is at $(\pi/2,\pi/2)$ and the maximum is at (0,0). Thus the upper theoretical line (squares) gives the large J_z/t bandwidth. In both of these limits there is good agreement with the results of exact diagonalization. (For intermediate values of J_z/t, the bandwidth is given by $\omega_{(\pi,0)}-\omega_{(\pi/2,\pi/2)}$.) Our results for the larger $J_\perp/J_z=\frac{1}{4}$, in Fig. 17(c), are qualitatively similar. Again there is a level crossing (here not shown) above which the erasure graph dominates and the band minimum is at $(\pi/2,\pi/2)$

One interesting result of the numerical calculations is that the bandwidth appears to vanish at some finite J_z. We can completely understand this result from our analytical work. Plotted in Fig. 17(b) are two analytical energy differences: $\omega_{(0,0)}-\omega_{(\pi/2,\pi/2)}$ and $\omega_{(0,0)}-\omega_{(\pi,0)}$. Also plotted is the single energy difference obtained from the 4×4-site diagonalization. (As mentioned above, $\omega_{(\pi/2,\pi/2)}$ and $\omega_{(\pi,0)}$ are degenerate for this cell.) Notice first the results from the diagonalization: there the bandwidths shrinks as J_z/t shrinks, passing through zero at $J_z/t\approx0.7$. This vanishing of the bandwidth, as mentioned above, has also been seen in calculations at the Heisenberg point. There it was taken to be a precursor of the Nagaoka transition. Here we see reason to doubt that interpretation. In our analytic calculation the quantity $\omega_{(0,0)}-\omega_{(\pi/2,\pi/2)}$ is also seen to pass through zero and become negative (at $J_z/t\approx0.5$). In our calculation the system is of infinite extent and no Nagaoka transition is possible. In fact, the correct interpretation of this point is that it is simply a level crossing, a consequence of the interference between the loop and erasure processes. The bandwidth does not vanish, but the ground state switches from $\mathbf{k}=(0,0)$ to $(\pi/2,\pi/2)$.

Such a vanishing of $\Delta\omega=\omega_{(0,0)}-\omega_{(\pi/2,\pi/2)}$ at small J was first seen in numerical calculations of the Heisenberg limit. Does the same explanation work there? We cannot with certainty answer this question, since our calculation breaks down at very small J_z. However, the vanishing is consistent with our results. From Fig. 16, clearly we predict that the band minimum lies at $\mathbf{k}=(\pi/2,\pi/2)$ in the Heisenberg limit at large J. (Go up toward $J_\perp/J_z=1$ at large J_z.) However, the region in which the band minimum lies at $\mathbf{k}=(0,0)$ grows as J_z shrinks. Our calculation becomes suspect below $J_z\approx0.2$. Nonetheless, it is quite plausible that at some sufficiently small J_z the (0,0) region extends all the way up to $J_\perp/J_z=1$. If so, a calculation in the Heisenberg limit would exhibit crossover at this value of $J=J_z$. This might explain the numerical result even in the Heisenberg limit.

Figure 18 summarizes the behavior of the low-lying coherent states of the hole as J_z/t is varied for some typical value of $J_\perp/J_z$. The lines schematically denote constants of energy near the band minimum. The shape of these "hole pockets" changes qualitatively near the crossover point discussed above. The values $J/t\sim1.0$, 0.1,

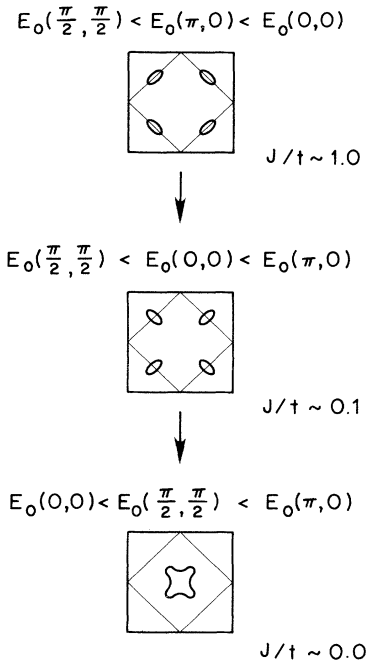

FIG. 18. Schematic illustration of evolution of the shape of the low-energy band as J_z/t varies. The values $J/t=1.0$, 0.1, and 0.01 are order-of-magnitude guidelines. For $J/t \gg 0.1$, when the string-erasure graph is dominant, the minimum lies at $\mathbf{k}=(\pi/2,\pi/2)$ and the maximum at $\mathbf{k}=(0,0)$. At $J/t \sim 0.1$, the maximum has shifted to $\mathbf{k}=(\pi,0)$. For $J/t \ll 0.1$, when the loop graphs are dominant, the minimum has shifted to $\mathbf{k}=(0,0)$.

and 0.01 are to be considered only order of magnitude guidelines.

B. Intersublattice motion

Up until now, we have concentrated on the dispersion of the lowest energy state, by focusing on processes in which the hole is injected onto and removed from the same sublattice. Now we wish to turn to a discussion of the total spectral weight of the charge excitation. To do so, we have to take into account an additional process which lets the hole be injected onto one sublattice and removed from the other. It turns out that this process does not change the position of the poles, but it does redistribute their spectral weights.

Naively one might expect that the intersublattice process might be more important than it is. This is suggested by the useful relation[17(g)]

$$\sum_{\sigma} \langle \psi | c_{i,\sigma}^{\dagger} T_K c_{j,\sigma} | \psi \rangle = t \langle \psi | (2\mathbf{S}_i \cdot \mathbf{S}_j + \tfrac{1}{2} n_i n_j) | \psi \rangle \ , \tag{86}$$

where $\langle i,j \rangle$ are nearest neighbors, $n_i = c_{i,\uparrow}^{\dagger} c_{i,\uparrow} + c_{i,\downarrow}^{\dagger} c_{i,\downarrow}$ and the kinetic-energy operator T_K is given by Eq. (1). Here $|\psi\rangle$ denotes some half-filled spin state.

In the classical antiferromagnetic state, the Néel state, the expectation value of the kinetic energy of an injected hole, as defined by Eq. (86), vanishes. But when quantum fluctuations are present in $|\psi\rangle$, then this expectation value becomes nonzero, of order t. It has therefore been suggested that one might inject a charge into a fluctuating antiferromagnet or perhaps a quantum spin liquid and propagate it with a coherent bandwidth of order t.[17(g)] We believe that this is in fact not correct.

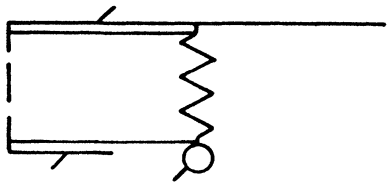

FIG. 19. The intersublattice process. The occurrence of a vacuum fluctuation in the half-filled state (represented by the $J_\perp/2$ vertex on the left side of the diagram) allows a spin-↑ hole to be created on the B site by the destruction of the d line. This changes the flipped spin (double line) into a single hole line in the lower part of the diagram.

The above relation involves propagation between the two sublattices, and so to address this issue we consider the Green's function for the hole created by removing an electron of definite spin (say, ↑):

$$G_{ij}(t) = -i \langle \mathcal{T} c_{i\uparrow}^{\dagger}(t) c_{j\uparrow}(0) \rangle \ . \tag{87}$$

Now, depending on the sublattice, creating a hole can be thought of as either creating a vacant or destroying a doubly occupied site. That is, G_{ij}, written in terms of h and d operators can have one of four forms. These can be labeled by AA, AB, BA, and BB, where the second letter denotes the sublattice on which the hole is injected and the first where it is taken out again:

$$\begin{aligned} G_{AA}(t) &= -i \langle \mathcal{T} \hat{h}_i(t) \hat{h}_j^{\dagger}(0) \rangle \ , \\ G_{AB}(t) &= -i \langle \mathcal{T} \hat{h}_i(t) \hat{d}_j(0) \rangle \ , \\ G_{BA}(t) &= -i \langle \mathcal{T} \hat{d}_i^{\dagger}(t) \hat{h}_j^{\dagger}(0) \rangle \ , \\ G_{BB}(t) &= -i \langle \mathcal{T} \hat{d}_i^{\dagger}(t) \hat{d}_j(0) \rangle \ . \end{aligned} \tag{88}$$

It is very easy to combine all of these contributions and make a connection to the effective tight-binding Hamiltonian discussed before. The lowest-order contribution to G_{AB} is shown in Fig. 19. There a vacuum fluctuation occurs in the Néel state so that a d operator can create a hole. (In Fig. 19 the hole is created when the double line in the lower half of the figure is changed into a single line.) The kinetic energy vertex then transfers this hole to the NN site, destroying the flipped spin originally created there by the vacuum fluctuation. This process is the dynamical generalization of Eq. (86). Using the rules of Appendix D, this graph (dressed with H_z) gives

$$G_{ij}^{AB}(\omega) = \left[\frac{-J_\perp/2}{-2\varepsilon_0 + J_z} \right] \frac{1}{\omega - 3\varepsilon_0/2 + i\delta} (-t) \frac{1}{\omega - \varepsilon_0/2 + i\delta} \delta_{\langle ij \rangle} \ . \tag{89}$$

This process creates as an intermediate configuration a state with one hole plus a flipped spin on a NN site, as shown by Fig. 19. We can interpret this configuration as a state labeled 2 or 6 on the lattice for the effective tight-binding Hamiltonian (see Fig. 11). Therefore we can handle this intersublattice process in a conserving way using the tight-binding lattice. The hole can be created or destroyed at either an A site, such as 1 or 7, or a B site, such as 2 or 6. Once the hole is created it is free to propagate and can be removed from either sublattice. Graphs like that in Fig. 19 can only occur at the beginning or the end of the propagation. Hence they do not change the position of the poles in the self-energy, but only influence their relative spectral weight.

The above expression for G_{ij}^{AB} can be interpreted as the (bare) Green's function for propagation from site 2 to site 1 on the tight-binding lattice, multiplied by $(J_\perp/2)/(2\varepsilon_0-J_z)=-t'/t$. That teaches us what to do in general: compute the Green's function for motion on the tight-binding lattice, not only among sites equivalent to 1, but among all sites equivalent to 1 and 2 (i.e., sites on both sublattices A and B). To get the corresponding electron Green's function in Eq. (87), multiply by $-t'/t$ if i is on sublattice B and multiply by $-t'/t$ again if j is on B. For clarity in what follows, let us label by $G^{\rm TB}$ the Green's function for motion on the tight-binding lattice, without any extra multipliers, and use G for the electron Green's function Eq. (87).

Suppose the hole is injected on site 2. Then the hole could move to, say, site 1 and then to any other site $i\in A$ in the lattice. This gives a contribution to the tight-binding Green's function

$$G_{i2}^{\rm via\,1}=(-tG'_{22}-2t'G'_{26}-t'G'_{22'})G_{i1}^{\rm TB}(\omega)\ . \tag{90}$$

$G_{i1}^{\rm TB}$ is the propagator between site 1 and i worked out in the previous subsections. Here the hole can move around the diamond and exit to site 1 from any of the diamond's four corners. In addition, the hole could exit to any other A site adjacent to the diamond, and moreover could have been injected at any site within the diamond. Including all such contributions, allowing the hole to be injected on and removed from either A or B, and including the factor $-t'/t$ as appropriate, the complete Green's function is

$$G(\mathbf{k},\omega)=G^{\rm TB}(\mathbf{k},\omega)\left[1+z\gamma_{\mathbf{k}}\left[\frac{t'}{t}\right](t+3t')(G'_{22}+2G'_{26}+G'_{22'})\right]^2+4\left[\frac{t'}{t}\right]^2[G'_{22}+2G'_{26}+G'_{22'}]\ . \tag{91}$$

Here $G^{\rm TB}$ is the Green's function for motion on the A sublattice only worked out in the previous subsections.

In Fig. 20 we have plotted the spectral weight $-\pi^{-1}{\rm Im}(\mathbf{k},\omega)$ for a single charge, for two choices of parameters, using Eq. (91). This Green's function includes the Brinkman-Rice self-retracing paths, the plaquette path, the string-erasure graph, and the intersublattice motion discussed above. All of these contributions have been included in a conserving approximation, and consequently, the spectral weight is strictly non-negative.

Also shown in Fig. 20 are spectral weights obtained by the exact diagonalization of a 4×4 cluster with periodic boundary conditions and one hole. The quantitative agreement of the position of both the lower band edge (i.e., the propagating pole) and the upper band edge is evident. In addition, the spectral weight of the lowest pole (i.e., the integrated area) agrees very well between theory and computer experiment. This is a stringent test of a theory containing no adjustable parameters. (The values for $t,J_z,J_\perp$ and the small imaginary part of ω in Fig. 20 are the same for both theory and numerical data.)

For intermediate energies, the position and intensity of the peaks obtained from the two calculations do not coincide. In particular, the theoretical peaks are spaced much more regularly than are the numerical peaks. This regularity is not surprising, since the theoretical poles are essentially the excited states of the hole moving in the linear confining string potential of the self-retracing paths. Thus, without being able to prove it, we suspect that the discrepancy at intermediate values of ω might be an indication of the finite-size limitations of the exact diagonalization results. The excited string states are more extended spatially and therefore more sensitive to the finite-size boundary conditions than is the ground state.

We might ask ourselves: How systematic are the above calculations for the properties of the charge excitation in the t-J model? In particular, is the good agreement which we find between the theory and the numerical results due to the fact that we have limited ourselves to the asymptotic Ising limit, i.e., a regime where the transverse spin coupling $J_\perp$ can be treated perturbatively? We have already noticed in the previous subsection that the agreement is in fact much better than what one might naively expect based on the ratios $J_\perp/J_z=0$, $\frac{1}{20}$, and $\frac{1}{4}$ used in our calculations. Here we want to propose another possibility, that an additional geometrical control factor, namely, $1/z$, might be coming into play. To be sure, the following argument completely lacks the mathematical rigor of the $1/z$ expansion that we obtained for the spin excitation. Nevertheless, there are indications that a generalization of the present work to the Heisenberg limit (which we have left for the future) might give quantitatively controlled results as well.

Let us consider Fig. 21. In Fig. 21(a) the basic string-erasure graph is shown another time. Figure 21(b) is an illustration of the fact that this string-erasure graph does not let the hole come back to the original site, due to the constraint prohibiting a hole and spin excitation from occupying the same site. [This is the source of the $\mathbf{k}$ dependence of the graph in Eq. (81).]

Clearly, this string-erasure process is only one of many which let the hole propagate coherently. This situation is similar to that of the plaquette graph. That graph was also only the shortest of an infinite set of more general

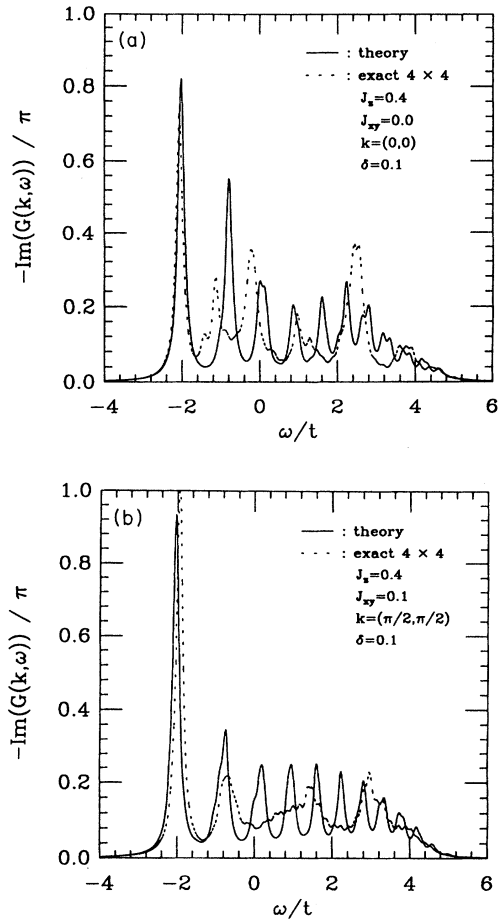

FIG. 20. The spectral weight as a function of ω (in units of t) for some selected values of $\mathbf{k}$ and J_z. The imaginary part of ω is $0.1t$. The solid line is the theoretical curve obtained by the conserving inclusion of the self-retracing paths, the plaquette path, the string-erasure graph and the intersublattice contribution. The dashed line is the data obtained by the exact diagonalization of a 4×4 lattice with one hole. (a) For $J_z=0.4$, $J_\perp\equiv J_{xy}=0$. (b) For $J_z=0.4$, $J_\perp\equiv J_{xy}=0.1$. Note that both the lowest dominant pole and the upper band edge are accurately reproduced. The Brinkman-Rice-type poles at intermediate energies are excited states which are quite extended and expected to be most affected by finite-size corrections.

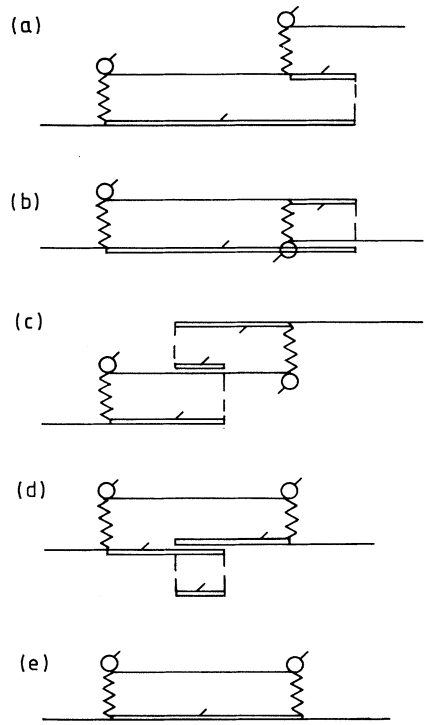

FIG. 21. This series of processes illustrates why we believe that the corrections to the basic string-erasure graph shown in (a) may be reduced by a geometrical factor $1/z$. (b) shows that the string-erasure graph has no site-diagonal version, because of the constraint prohibiting a hole and spin excitation from sharing the same site. (c) Is an example of a higher-order contribution to the string-erasure graph. (d) Is the (allowed) on-site version of this contribution. As such, it is evident that it is reduced by a factor $1/z$ with respect to the basic self-retracing path, shown in (e).

loop graphs, all of which contribute to the coherent hole motion in the Ising limit. In that case we understood that the contribution of the longer loop graphs is suppressed exponentially relative to the shortest, since the hole has to tunnel through a barrier whose height $\sim J_z$ increases with increasing loop length.

Does something similar hold for the string-erasure graph? In Fig. 21(c) we have drawn one simple correction to the basic process of Fig. 21(a). It is surely not the only correction, but serves here to illustrate our argument. In Fig. 21(d) the same process is redrawn as a site-diagonal contribution [this is now allowed, in contrast to Fig. 21(b)]. How important is this process in the Heisenberg limit? The two $J_\perp/2$ vertices, together with a geometrical multiplicity factor, gives this process the weight $z(J_\perp/2)^2/\varepsilon_0^2\sim 1/z$, following the arguments of Sec. III. That is to say, Fig. 21(d) is a $1/z$ correction to the simple self-retracing step shown in Fig. 21(e). This suggests that (quite independent of the relative magnitude of t versus J) the contribution of the process of Fig. 21(c) is suppressed by one order of magnitude in the coordination number with respect to the basic string-erasure graph.

Can we generalize this argument? If we consider the supersymmetric case, $t\sim J$, then all energy scales are of the same order of magnitude, namely, $\sim zt\sim zJ$. Each higher-order graph has more vertices than the string-erasure graph of Fig. 21(a). Therefore, following the arguments of Sec. III, it should be suppressed by some factors $1/z$, since $J\sim t\sim(zJ)/z$.

In Sec. III we showed that, while both Trugman's method and the graphical calculation lead to effective tight-binding problems, the two approaches are not equivalent. Trugman's method, for example, included terms that were, in the graphical calculation, discarded

for arising from disconnected diagrams. This seems to be a general lesson. For example, we can repeat the Trugman calculation of one-hole properties based on the process shown in Fig. 14.[19] This leads to a 17-state manifold. The graphs corresponding to the matrix elements between these states include the simplest erasure graph (Fig. 14), plus the shortest two self-retracing paths and several others. The results of these two calculations differ in a number of details. The simplest difference is that the Trugman result, interpreted in terms of a self-energy, includes a piece $12(J_\perp/2)^2/(\omega-\varepsilon_3)$. This is the contribution of a disconnected graph similar to that in Fig. 6(b). There are many other differences, but this illustrates that the basic result in Sec. III continues to hold true. While our calculation leads to a tight-binding problem, it is not identical to Trugman's method; e.g., every link in our effective tight-binding Hamiltonian corresponds to some connected diagram.

In summary, we have developed a method to study single-hole properties, preserving the constraints against double occupancy. Near the Ising limit, this method is quite controlled. An analysis that combines real-space Feynman-type diagrams and mappings to tight-binding models leads to a conserving approximation from which we can extract the complete spectral weight of the hole. The results are in good agreement with numerical calculations for large J_z, and show an interesting interference effect that leads to a crossover in the band minimum at certain values of $J_\perp$.

VI. SUMMARY AND CONCLUSIONS

We have developed a real-space Green's-function approach to study the properties of spin and charge excitations of the t-J model near half-filling. The t-J model can be viewed as a prototypical model of strongly correlated electrons in which the correlations enter the model in two ways: via a residual antiferromagnetic coupling between nearest-neighbor spins, and by a prohibition against the occupancy of any site by more than a single electron. By working in real space, we can ensure that the restriction against double occupancy, a subtle requirement in other approaches, is exactly satisfied. Similarly, a real-space analysis lets us identify and include processes—such as the loop path for a moving hole—which would otherwise be difficult. (For example, the loop paths give a set of vertex corrections that here can easily be evaluated.) Philosophically, this method builds on the work of Brinkman and Rice.

The price we pay for this real-space approach is that the perturbation problem is moderately complicated—for instance, our Hamiltonian has three separate perturbations and hence the graphs have three different interaction lines. On the other hand, the graphs are extremely simple to evaluate. Using the simple rules included here in Appendix D, each graph can be evaluated algebraically without any internal momentum or energy integrals

In this paper we have studied the properties of a single spin or charge excitation. We perturb about the Néel state (justified on the hypercubic lattice in two or more dimensions), using the underlying fermion operators to represent both charge and spin. Then Wick's theorem is valid, and we have developed a fairly ordinary graphical method, except that by working in real space the graphs are simpler than usual to evaluate.

In this approach the transverse spin coupling $J_\perp$ enters as a perturbation, and hence we can only approach the interesting Heisenberg limit perturbatively. However, convergence is actually better than this seems to imply. In the spin analysis we can explicitly identify a different control parameter, $1/z$ (where z is the coordination number). By grouping all of the graphs at a given order in $1/z$, we find that when the Heisenberg limit is taken, rotational invariance is completely restored. This is nontrivial, and explains why, for example, the contribution lowest order in $J_\perp$ gives the usual gapless spin-wave spectrum in the Heisenberg limit. We have extended this calculation to the next highest order in $1/z$, where we find a renormalization of the spin-wave velocity.

There is evidence that the charge properties also are controlled by something like a factor of $1/z$—for example, the bandwidth calculated for $J_\perp/J_z=\frac{1}{4}$ is within 3% of exact numerical results (for a 4×4 lattice), rather than within 25% as one would naively expect.

The charge excitation shows a very interesting interference between two competing processes which allow propagation. First one loop paths in which, for example, a hole moves around a fundamental plaquette $1\frac{1}{2}$ times; this permits the hole to move even in the Ising limit. This path, a correction to the kinetic energy vertex, can be thought of as a tunneling problem in one dimension; it is exponentially small in the Ising energy. The second propagation route is the erasure process in which the transverse spin coupling permits a string of disturbed spins to relax. It is of order $J_\perp$. These different processes interfere, changing the band minimum from $\mathbf{k}=(0,0)$ at small J_z to $\mathbf{k}=(\pi/2,\pi/2)$ at larger values. This appears to explain a surprising vanishing of the bandwidth found in numerical calculations—the bandwidth does not truly vanish, but what was seen in the numerical work was the level crossing in the ground state.

Both the loop and erasure processes, plus the self-retracing paths of Brinkman and Rice, have been included here in a conserving approximation. This was constructed by a novel method in which the original graphs were mapped to a tight-binding Hamiltonian. Solving the tight-binding problem gives the conserving approximation, and studying the tight-binding model lets us identify what graphs are needed (in addition to our original choices) for the approximation to be conserving. By comparison with the results of exact diagonalization on the 4×4 lattice we find good evidence that we have included the essential physics of the charge excitation for $J_z/t>0.2$ or so.

In summary, by considering only a single spin or charge excitation in a correlated background, it is possible to keep the constraint against double occupancy exactly and systematically learn about the strongly-correlated electron problem.

Our purpose here was twofold—to illustrate our method and to use it to study the above two interesting questions. We believe this method has the potential to be

useful in other problems, and to that end we have explained its use in considerable detail. Among interesting problems for which this method looks useful are calculations of the Raman spectrum, the renormalization of the spin excitation in a finite density of holes, the optical conductivity, and finite temperature properties.

ACKNOWLEDGMENTS

This was was supported by the NSF under Grant No. DMR-8802383 and by the Chester Davis Foundation. We are grateful to Allan MacDonald and Steve Girvin for interesting and helpful discussions. K.v.S. appreciates many helpful and revealing discussions with W. Stephan and P. Horsch.

APPENDIX A: FEYNMAN RULES FOR $H_\perp$; TIME SPACE

Here we describe the possible types of graphs that arise under the perturbation $H_\perp$, and present Feynman rules to aid in their evaluation.

1. Graphs

(i) Draw graphs with time increasing from left to right, and label the internal time t_i on each vertex. Generally let different sites be represented by different horizontal lines.

(ii) Note that $H_\perp$ only couples adjacent sites.

(iii) $H_\perp$ at some intermediate time will consist either of all creation or all destruction operators. Thus, at a given vertex, either all four lines enter from the left or exit to the right [Fig. 22(a)].

There are similar but slightly different rules for evaluating graphs leading to single-particle Green's functions G and two-particle functions D.

2. Feynman rules for G

(i) Each line gives $G^{(0)}(t_j-t_i)$ where t_i is the left-hand time and t_j is the right-hand time.

(ii) Integrate over internal times $\int_{-\infty}^{\infty} dt_i$ and multiply by $(i/\hbar)^n(J_\perp/2)^n$. Here we suppose the graph is of order n in $H_\perp$.

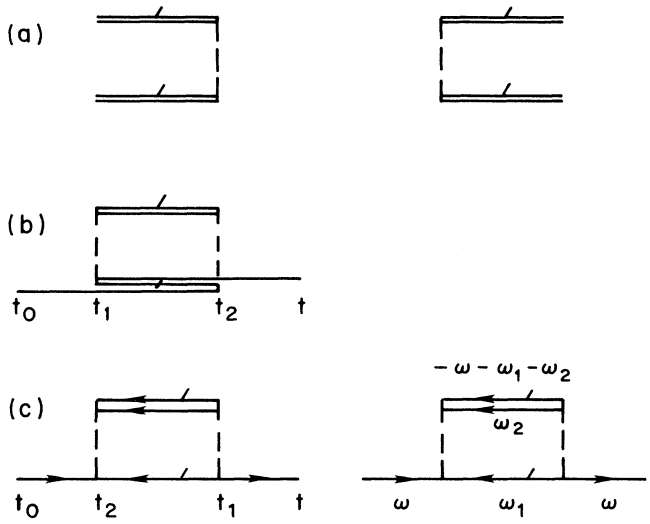

FIG. 22. (a) At an $H_\perp$ vertex, all four lines must enter together from the left or exit to the right. Increasing time is to the right. (b) A graph for G, second order in $H_\perp$, drawn as usual with time increasing to the right. (c) The same graph as in (b), here with the direction of increasing time explicitly indicated by arrows. In this form it is clear that there is one fermion loop. Also shown is the same graph labeled with frequencies.

(iii) Count the number of identical graphs to get a geometric multiplier. This can arise when sites can be connected by more than one path of nearest-neighbor links. (This is the C_R of Sec. III.)

(iv) The sign is given by $(-1)^F$, where F is the number of fermion loops. Notice that this includes a (-1) for each double line.

To count loops it is occasionally useful to draw graphs in a different way: put in the direction of increasing time explicitly with an arrow, and straighten out "bent" lines. See, e.g., Figs. 22(b) and 22(c).

(v) One can always replace $[G^{(0)}(t)]^2$ from a double line by the equivalent $-iD^{(0)}(t)$. For example, for the graph in Fig. 22(b) these rules yield

$$G(t)=(-1)i^2\left[\frac{J_\perp}{2}\right]^2 z\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2\,G^{(0)}(t_2)G^{(0)}(t-t_1)[G^{(0)}(t_2-t_1)]^3\ . \tag{A1}$$

3. Feynman rules for D

One can simply follow the above rules for G, and then, at the end, multiply by an extra factor of $-i$. Or, when the graph consists entirely of double lines, it can be convenient to use the following, equivalent rules.

(i) Each double line gives $D^{(0)}(t_j-t_i)$.

(ii) Integrate over internal times $\int_{-\infty}^{\infty} dt_i$ and multiply by $(J_\perp/2\hbar)^n$.

(iii) Count the number of identical graphs to get a multiplier, as in rule (iii) in subsection 2.

(iv) The sign is given by $(-1)^{F+1}$, where F is the number of double lines.

These rules give the lowest-order spin graph evaluated in Sec. III, Eq. (14).

APPENDIX B: FEYNMAN RULES FOR $H_\perp$; FREQUENCY SPACE

1. Graphs

Draw graphs as in Appendix A 1. Assign an energy ω_i to each internal line using conservation of energy at each

vertex. Note that lines enter a vertex from the left and leave it to the right. However, it can be convenient to redraw graphs with explicit "time" arrows, as in Fig. 22 (c), to determine the ω_i's.

2. Feynman rules for G

(i) Each line gives $G^{(0)}(\omega_i)$.

(ii) Integrate over internal frequencies $\int_{-\infty}^{\infty} d\omega_i/2\pi$ and multiply by $(i/\hbar)^n(J_\perp/2)^n$. Here we suppose the graph is of order n in $H_\perp$.

(iii) Count the multiplicity of identical graphs as in rule (iii), Appendix A 2.

(iv) The sign is given by $(-1)^F$, where F is the number of fermion loops (including double lines).

(v) One can always use

$$\int_{-\infty}^{\infty}(d\omega'/2\pi)G^{(0)}(\omega')G^{(0)}(\omega-\omega')=-iD^{(0)}(\omega)$$

to represent double lines.

Applying these rules to Fig. 22(b) gives

$$\begin{aligned}G(\omega)&=(-1)i^2\left[\frac{J_\perp}{2}\right]^2 z\int_{-\infty}^{\infty}\frac{d\omega_1}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega_2}{2\pi}[G^{(0)}(\omega)]^2G^{(0)}(\omega_1)G^{(0)}(\omega_2)G^{(0)}(-\omega-\omega_1-\omega_2)\\&=\left[\frac{J_\perp}{2}\right]^2 z\int_{-\infty}^{\infty}\frac{d\omega_1}{2\pi}[G^{(0)}(\omega)]^2G^{(0)}(\omega_1)[-iD^{(0)}(-\omega-\omega_1)]\\&=-\left[\frac{J_\perp}{2}\right]^2 z[G^{(0)}(\omega)]^2G^{(0)}(-\omega-\varepsilon_0)\ .\end{aligned}\tag{B1}$$

This same result can be obtained by Fourier transforming Eq. (A1).

3. Feynman rules for D

One can follow the above rules for G, then, at the end, multiply by an extra $-i$. Or, when the graph consists entirely of double lines, it can be convenient to use the following, equivalent rules.

(i) Each double line gives $D^{(0)}(\omega_i)$.

(ii) Integrate over internal frequencies $\int_{-\infty}^{\infty} d\omega_i/2\pi$, and multiply by $(J_\perp/2\hbar)^n$.

(iii) Count the number of identical graphs to get a multiplier, as in rule (iii) in Appendix A 2.

(iv) The sign is given by $(-1)^{F+1}$, where F is the number of double lines.

These rules were used in Sec. III to evaluate the lowest-order spin graph, Eq. (16).

APPENDIX C: RULES FOR H_z

In this appendix we present rules for graphs containing H_z as well as $H_\perp$.

1. Graphs

(i) H_z can only join lines on neighboring sites, without ending them. This is because H_z consists of terms such as $h_1^\dagger h_1 h_0^\dagger h_0$ which annihilate and then immediately recreate particles. Thus at either end of an H_z interaction line, a propagator enters and then leaves [Fig. 23(a)].

(ii) H_z interaction lines cannot cross. As explained in Fig. 23(b), this is because the unperturbed Green's function is casual [i.e., $G^{(0)}(t)\sim\Theta(t)$].

2. Feynman rules for H_z and $H_\perp$

Follow the time- or frequency-space rules given for $H_\perp$ in Appendix A or B, with the following additions to handle H_z.

(i) Each line gives a $G^{(0)}$ whether its end points are $H_\perp$ vertices or H_z vertices or a mixture.

(ii) Integrate over internal frequencies or times, as before. For a graph of order n in $H_\perp$ and m in H_z, multiply by a factor $(i/\hbar)^{n+m}(J_\perp/2)^n(-J_z/4)^m$.

(iii) An extra geometric multiplier can arise if different $G^{(0)}$ lines can be connected by H_z without changing the graph. This can occur because either h or d lines can be connected to either h or d lines. For example, an H_z connection between double lines give a factor of 4. Also include a geometric multiplier for $H_\perp$, as in Appendix A 2, rule (iii).

Everything else is as given in the $H_\perp$ rules, including, in particular, the sign $(-1)^F$ and an extra factor $-i$ if the diagram is for D.

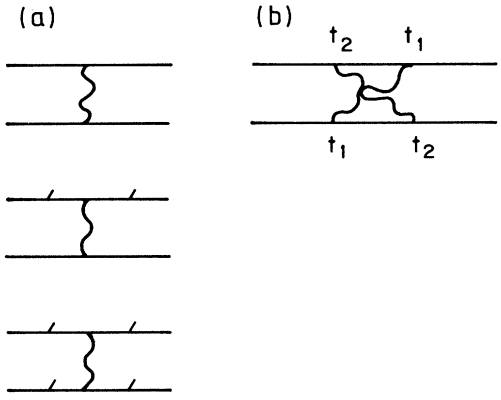

FIG. 23. (a) The H_z interaction line can only join two propagators without ending them. Both h and d lines can be connected to either h or d lines. (b) H_z interaction lines cannot cross. The bottom line requires $t_1<t_2$, while the top requires $t_2<t_1$ [since $G^{(0)}(t)$ is nonzero only for $t>0$].

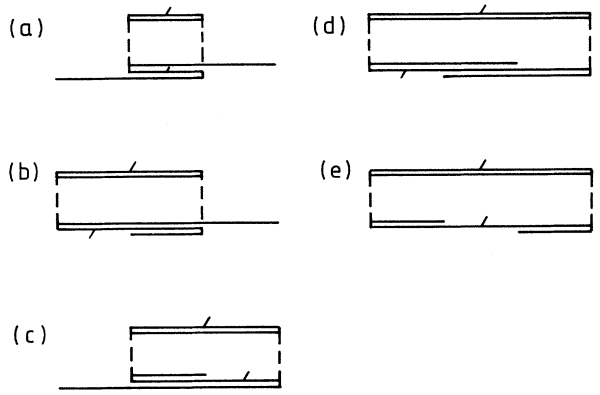

FIG. 24. All possible time orderings of the diagram in Fig. 22(b). Here the orderings are (a) $t_0 < t_1 < t_2 < t$, (b) $t_1 < t_0 < t_2 < t$, (c) $t_0 < t_1 < t < t_2$, (d) $t_1 < t_0 < t < t_2$, (e) $t_1 < t < t_0 < t_2$.

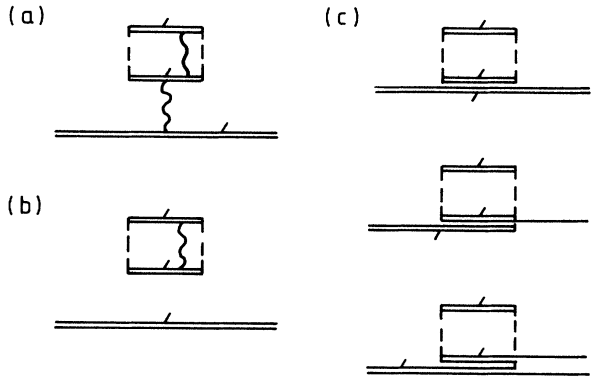

FIG. 25. (a) A graph connected by virtue of the H_z dressing. (b) Without part of the H_z coupling, the graph becomes disconnected. (c) A set of graphs which naively appear to violate the Pauli exclusion principle. Adding these together gives zero for the time ordering shown here.

APPENDIX D: ALTERNATE RULES FOR $H_\perp$, H_z, AND T_K

Here we summarize a less standard set of rules for evaluating Feynman diagrams containing any of $H_\perp$, H_z, and T_K. These rules tend to be the easiest to apply, especially for complicated graphs.

Rules for drawing the $H_\perp$ and H_z graphs are given in Appendixes A and C. Graphs involving T_K are of the form shown in Fig. 7, i.e., they hop a hole line to a neighboring site, leaving a double line behind. At one end of the T_K interaction line there is also a closed loop from the constraint pieces in T_K.

These rules focus on time *ordering* and *intervals*. A parent graph has certain time orderings unspecified; corresponding to each parent graph is a set of daughter time-ordered graphs. For example, all of the possible orderings of the parent given in Fig. 22(b) are shown in Fig. 24. On these graphs, find the frequency of each line using energy conservation.

(i) The frequency dependence of each parent graph is the sum of contributions from each time-ordered daughter; a daughter's contribution is a product over factors $(\omega_{\rm tot}-\varepsilon_{\rm tot}+i\delta)^{-1}$, one such factor per time interval. Here, for a given time interval, $\omega_{\rm tot}$ is the sum of the frequencies on each $G^{(0)}$ line passing through the interval and $\varepsilon_{\rm tot}$ is the sum of $\varepsilon_0/2$ for each line passing through the interval. For example, for the time-ordered daughter in Fig. 24(a), this gives

$$\left[\frac{1}{\omega-\frac{1}{2}\varepsilon_0+i\delta}\right]^2\frac{1}{\omega-\frac{5}{2}\varepsilon_0+i\delta}\ . \tag{D1}$$

Closed loops with tick marks (e.g., as in Fig. 7) give $+1$; these come from $\langle dd^\dagger\rangle$ factors. Without tick marks they give 0 (i.e., they force the graph to vanish).

(ii) Multiply by $(J_\perp/2\hbar)^n(-J_z/4\hbar)^m(-t/\hbar)^l$ for a graph of order n in $H_\perp$, m in H_z, and l in T_K.

(iii) An extra geometric multiplier is given by the number of identical graphs: (a) for $H_\perp$, this is the number of different ways sites can be joined by nearest-neighbor links; (b) for H_z, this is the number of different pairs of lines that can be connected by the interaction [e.g., there are four possibilities in Fig. 4(a)].

(iv) For graphs leading to G, the sign is $(-1)^F$ where F is the number of fermion loops (including double lines). For D, the sign is $(-1)^{F+1}$. These rules apply if there are no factors of T_K. We do not have succinct rules for the spin when the graph has T_K vertices; in every example thus far, the sign (found by direct Wick's expansion) is positive unless the graph is an intersublattice process, such as Fig. 19.

Let us complete the example of Fig. 22(b). The frequency dependence arising from all five time-ordered daughters in Fig. 24, is from rule (i),

$$\frac{1}{\omega-\frac{5}{2}\varepsilon_0+i\delta}\left[\left[\frac{1}{\omega-\frac{1}{2}\varepsilon_0+i\delta}\right]^2+2\left[\frac{1}{\omega-\frac{1}{2}\varepsilon_0+i\delta}\right]\left[\frac{1}{-2\varepsilon_0}\right]+\left[\frac{1}{-2\varepsilon_0}\right]^2\right]+\left[\frac{1}{-2\varepsilon_0}\right]^2\frac{1}{-\omega-\frac{3}{2}\varepsilon_0+i\delta}$$

$$=\left[\frac{1}{\omega-\frac{1}{2}\varepsilon_0+i\delta}\right]^2\frac{1}{-\omega-\frac{3}{2}\varepsilon_0+i\delta}\ , \tag{D2}$$

so that

$$G(\omega)=(-1)\left[\frac{J_\perp}{2}\right]^2 z\left[\frac{1}{\omega-\frac{\varepsilon_0}{2}+i\delta}\right]^2\frac{1}{-\omega-\frac{3\varepsilon_0}{2}+i\delta}\ , \tag{D3}$$

in agreement with Eq. (B1).

Now let us sketch the derivation of these rules, starting from the perturbation expansion of the Green's function.[32] Fourier transforming, a graph for $G(\omega)$ generally yields

$$G(\omega)=(-i)^{n+m+l+1}\left[\frac{J_\perp}{2}\right]^n\left[-\frac{J_z}{4}\right]^m(-t)^l S\int dt\, e^{i\omega t}\int\prod dt_i\, iG^{(0)}\cdots iG^{(0)} \tag{D4}$$

times geometric factors like those in rule (iii) above. The sign $S=\pm1$ arises from the Wick's expansion. There are $2n+2m+2l+1$ factors of $G^{(0)}$ in the integrand; we have associated a factor i with each. The integral can be broken into a sum over particular time orderings, as explained pictorially by Fig. 24. We signify this by $G=\sum_j G_j$, where G_j results from a particular time ordering. For a particular ordering we can collect the $iG^{(0)}$'s in the integrand into one factor $e^{i(\omega^i_{\rm tot}-\varepsilon^i_{\rm tot})(t_{i+1}-t_i)}$ for each interval $[t_i,t_{i+1}]$. Thus

$$G_j(\omega)=\left[\frac{J_\perp}{2}\right]^n\left[-\frac{J_z}{4}\right]^m(-t)^l S\int\prod dt_i\, e^{i\omega t}(-ie^{i(\omega^i_{\rm tot}-\varepsilon^i_{\rm tot})(t_{i+1}-t_i)})\ , \tag{D5}$$

where there is one term in the integrand for each of the $n+m+l+1$ time intervals. This gives the prefactor in rule (ii) and the sign in rule (iv). Evaluating the time integral gives the frequency rule above. We will illustrate this using as an example the time ordering in Fig. 24(a). There the integrand is

$$iG^{(0)}(t-t_1)[iG^{(0)}(t_2-t_1)]^3 iG^{(0)}(t_2)=[iG^{(0)}(t-t_2)iG^{(0)}(t_2-t_1)][iG^{(0)}(t_2-t_1)]^3[iG^{(0)}(t_2-t_1)iG^{(0)}(t_1)]\ , \tag{D6}$$

using Eq. (6). The integral for this time ordering is therefore

$$\begin{aligned}&\int_0^\infty dt_1\int_{t_1}^\infty dt_2\int_{t_2}^\infty dt\, e^{i\omega t}[iG^{(0)}(t_1)][iG^{(0)}(t_2-t_1)]^5[iG^{(0)}(t-t_2)](-i)^3\\&\quad=\int_0^\infty dt_1\int_0^\infty d(t_2-t_1)\int_0^\infty dt(t-t_2)e^{i[\omega-(1/2)\varepsilon_0]t_1}e^{i[\omega-(5/2)\varepsilon_0](t_2-t_1)}e^{i[\omega-(1/2)\varepsilon_0](t-t_2)}(-i)^3\\&\quad=(\omega-\tfrac{1}{2}\varepsilon_0+i\delta)^{-1}(\omega-\tfrac{5}{2}\varepsilon_0+i\delta)^{-1}(\omega-\tfrac{1}{2}\varepsilon_0+i\delta)^{-1}\ .\end{aligned} \tag{D7}$$

This agrees with Eq. (D1) above. Notice that the example chosen here would be a problem if it arose in our earlier analysis—the ordering shown in Fig. 24(e) allows a doubly occupied site. This ordering, however, was prohibited in Fig. 4(b) by the causality of the extra unbent h line

APPENDIX E: H_z TO INFINITE ORDER

Here we show that it is possible to compute the contribution of H_z to infinite order for each graph. By this we mean that a given graph in $H_\perp$ can be dressed by adding all possible H_z interaction lines.

The basic reason that this can be done is that $[H_0,H_z]=0$. The contribution of H_z to all orders would be given by letting the Hamiltonian $H_0\to H_0'=H_0+H_z$. If this substitution were made, then in the new interaction representation the time dependence of operators would become

$$O'(t)=e^{i(H_0+H_z)t}O_S e^{-i(H_0+H_z)t}=e^{iH_z t}[e^{iH_0 t}O_S e^{-iH_0 t}]e^{-iH_z t}=e^{iH_z t}O(t)e^{-iH_z t}\ , \tag{E1}$$

where $O(t)$ is in the original interaction representation, that is, in terms of H_0 in Eq. (4).

Let us see the consequences of this replacement by an example. The graph in Fig. 2(a) is a result of integrals over several time orderings, as explained in some detail in Appendix D. Consider the particular time-ordered daughter with the same ordering as Fig. 2(a), with times $0<t_1<t_2<t$. This bare time-ordered daughter is a term in the Wick's expansion of

$$\langle d_2(t)h_2(t)H_{01}(t_2)H^\dagger_{12}(t_1)h^\dagger_0(0)d^\dagger_0(0)\rangle\ , \tag{E2}$$

where we define $H_{ij}=d_j h_j d_i h_i$, and here the time dependence is in the original (H_0) interaction representation. The contribution of H_z to infinite order is given by replacing Eq. (E2) by

$$\begin{aligned}&\langle e^{iH_z t}d_2(t)h_2(t)e^{-iH_z(t-t_2)}H_{01}(t_2)e^{-iH_z(t_2-t_1)}H^\dagger_{12}(t_1)e^{-iH_z t_1}h^\dagger_0(0)d^\dagger_0(0)\rangle\\&\qquad=\langle d_2(t)h_2(t)H_{01}(t_2)e^{-i(-2J_z)(t_2-t_1)}H^\dagger_{12}(t_1)h^\dagger_0(0)d^\dagger_0(0)\rangle\ .\end{aligned} \tag{E3}$$

That is, the on-site energy in the time interval $[t_1,t_2]$ is changed from $3\varepsilon_0$ to $3\varepsilon_0-2J_z$.

The rule here is thus a modification of Appendix D: H_z can be incorporated to infinite order if *a pair of single lines on adjacent sites get an added "on-site" energy* $-J_z/4$. The contribution of all five daughters arising from Fig. 2(a), dressed in this way, is

$$D_{ij}^{(1)}(\omega)=\left[\frac{J_\perp}{2}\right]^2 C_{\mathbf{R}_i-\mathbf{R}_j}\left[\left[\frac{1}{\omega-\varepsilon_0}+\frac{1}{-2\varepsilon_0+J_z}\right]\left[\frac{1}{\omega-3\varepsilon_0+2J_z}\right]\left[\frac{1}{\omega-\varepsilon_0}+\frac{1}{-2\varepsilon_0+J_z}\right]+\left[\frac{1}{-2\varepsilon_0+J_z}\right]^2\left[\frac{1}{-\omega-\varepsilon_0}\right]\right] \tag{E4}$$

which will yield a dressed version of Eq. (41).

Sometimes this procedure accidentally incorporates a disconnected graph which therefore must be subtracted off. For example, graphs of the sort shown in Fig. 25(a), dressed all possible ways with H_z, include disconnected graphs like that in Fig. 25(b). Therefore the connected, dressed contribution arising from Fig. 25(a) is

$$D_{00}(\omega)=\left[\frac{J_\perp}{2}\right]^2 z(z-1)\left[\frac{1}{\omega-\varepsilon_0}+\frac{1}{-2\varepsilon_0+J_z}\right]^2\left[\frac{1}{\omega-3\varepsilon_0+2J_z}-\frac{1}{\omega-3\varepsilon_0+J_z}\right]. \tag{E5}$$

The piece subtracted in the final term in parentheses arises from the disconnected piece in Fig. 25(b).

It is more complicated when the graph is of a type that naively violates the Pauli exclusion principle, e.g., those in Fig. 25(c). Here we modify the rule: as far as the above added on-site energy is concerned, *count no more than one h line and one d line per site,* and only include graphs which are connected before H_z dressing. [This can be understood by thinking about the contribution from all three graphs in Fig. 25(c) for some fixed time ordering; e.g., for $0<t_1<t_2<t$, the sum of these three vanishes.] The contribution of the two connected graphs in Fig. 25(c), for all time orderings, completely dressed with H_z, is

$$D_{00}(\omega)=\left[\frac{J_\perp}{2}\right]^2 z\left[-\left[\frac{1}{\omega-\varepsilon_0}+\frac{1}{-2\varepsilon_0+J_z}\right]^2\left[\frac{1}{\omega-3\varepsilon_0+2J_z}\right]+\left[\frac{1}{-2\varepsilon_0+J_z}\right]^2\left[\frac{1}{-\omega-\varepsilon_0}\right]\right]. \tag{E6}$$

---

*Address after August 1, 1990: Department of Physics, University of Central Florida, Orlando, FL 32816-0385.

†Address after September 1, 1990: Institut für Physik, Universität Dortmund, Postfach 500 500, 4600 Dortmund 50, Germany.

[1]P. W. Anderson, Science **235**, 1196 (1987).

[2]F. C. Zhang and T. M. Rice, Phys. Rev. B **37**, 3759 (1988).

[3]Y. Nagaoka, Phys. Rev. **147**, 392 (1966).

[4](a) Y. Fang, A. E. Ruckenstein, E. Dagotto, and S. Schmitt-Rink, Phys. Rev. B **40**, 7406 (1989); (b) J. A. Riera and A. P. Young, *ibid.* **40**, 5285 (1989); (c) B. Doucot and X. G. Wen, *ibid.* **40**, 2719 (1989); (d) A. Barbieri, J. A. Riera, and A. P. Young, *ibid.* **41**, 11 697 (1990); (e) S. E. Barnes, *ibid.* **41**, 11 701 (1990).

[5]See, e.g., C. Gros, R. Joynt, and T. M. Rice, Phys. Rev. B **36**, 381 (1987).

[6]J. Oitmaa and D. D. Betts, Cand. J. Phys. **56**, 897 (1978).

[7]For a review, see Nandini Trevedi and David Ceperly, Phys. Rev. B **41**, 4552 (1990).

[8](a) G. Kotliar, Phys. Rev. B **37**, 3664 (1987); (b) T. Kopp, F. J. Seco, S. Schiller, and P. Wölfle, *ibid.* **38**, 11 835 (1988); (c) M. Drzazga, A. Kampf, E. Müller-Hartmann, and H. A. Wischmann, Z. Phys. B **74**, 67 (1989); (d) C. Jayaprakash, H. R. Krishnamurthy, and S. Sarker, Phys. Rev. B **40**, 2610 (1989); (e) B. I. Shraiman and E. D. Siggia, *ibid.* **40**, 9162 (1989).

[9](a) J. E. Hirsch and H. Q. Lin, Phys. Rev. B **39**, 4548 (1989); (b) S. Sorella, S. Baroni, R. Car, and M. Parinello, Europhys. Lett. **8**, 663 (1989); (c) M. Imada, J. Phys. Soc. Jpn. **58**, 2650 (1989); (d) I. Morgenstern, Z. Phys. B **73**, 299 (1989); **77**, 267 (1989); (e) S. R. White, D. J. Scalapino, R. L. Sugar, and N. E. Bickers, Phys. Rev. Lett. **63**, 1523 (1989); (f) S. R. White, D. J. Scalapino, R. L. Sugar, N. E. Bickers, and R. T. Scalettar, Phys. Rev. B **39**, 839 (1989); (g) T. Barnes, E. Dagotto, A. Moreo, and E. S. Swanson, *ibid.* **40**, 10 977 (1987).

[10](a) H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **57**, 2482 (1988); (b) C. Gros, R. Joynt, and T. M. Rice, Z. Phys. B **68**, 425 (1987); (c) For an overview, see, C. Gros, Ann. Phys. (N.Y.) **189**, 53 (1989); (d) T. Giarmarchi and C. Lhuiller (unpublished); (e) T. K. Lee and S. Feng, Phys. Rev. B **38**, 11 809 (1988).

[11](a) S. Liang, B. Doucot, and P. W. Anderson, Phys. Rev. Lett. **61**, 365 (1988); (b) D. Huse and V. Elser, Phys. Rev. Lett. **60**, 2531 (1988); (c) P. Horsch and W. von der Linden, Z. Phys. B **72**, 181 (1988); (d) R. R. P. Singh, M. Gelfand, and D. A. Huse, Phys. Rev. Lett. **61**, 2484 (1988).

[12]D. Förster, Phys. Rev. Lett. **63**, 2140 (1989).

[13](a) Z. Y. Weng and T. K. Lee, Phys. Rev. B **38**, 6561 (1988); (b) G. Vignale and K. S. Singwi, Phys. Rev. B **39**, 2956 (1989); **42**, 2611(E) (1990). (c) N. E. Bickers, D. J. Scalapino, and S. R. White, Phys. Rev. Lett. **62**, 961 (1989).

[14]J. R. Schrieffer, Bull. Amer. Phys. Soc. **35**, 274 (1990).

[15]C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, E. Abrahams, and A. E. Ruckenstein, Phys. Rev. Lett. **63**, 1996 (1989).

[16]P. W. Anderson, Phys. Rev. Lett. **64**, 1839 (1990).

[17](a) J. Bonča, P. Prelovšek, and I. Sega, Phys. Rev. B **39**, 7074 (1989); (b) E. Dagotto, A. Moreo, and T. Barnes, *ibid.* **40**, 6721 (1989); (c) Y. Hasegawa and D. Poilblanc, *ibid.* **40**, 9035

(1989); (d) K. Szczepanski, P. Horsch, W. Stephan, and M. Ziegler, *ibid.* **41**, 2017 (1990); (e) S. A. Trugman, *ibid.* **41**, 892 (1990); (f) V. Elser, D. A. Huse, B. I. Shraiman, and E. D. Siggia, *ibid.* **41**, 6715 (1990); (g) E. Dagotto, R. Joynt, A. Moreo, S. Bacci, and E. Gagliano, *ibid.* **41**, 9049 (1990).

[18]W. F. Brinkman and T. M. Rice, Phys. Rev. B **2**, 1324 (1970).

[19](a) S. A. Trugman, Phys. Rev. B **37**, 1597 (1988); (b) **41**, 892 (1989).

[20]C. L. Kane, P. A. Lee, and N. Read, Phys. Rev. B **39**, 6880 (1989).

[21]Work in this direction is currently in progress. F. Marsiglio (private communication).

[22](a) R. Eder and K. W. Becker, Z. Phys. B **78**, 219 (1990); (b) K. W. Becker, W. Hyekyung, and P. Fulde, *ibid.* **75**, 335 (1989).

[23]C. Gros and M. D. Johnson, Phys. Rev. B **40**, 9423 (1989).

[24]Spin rotational invariance can be understood in two distinct ways, as a property of the ground state or of the excited states. Together with a spontaneously broken symmetry, the latter leads to a gapless Goldstone branch. With respect to the Heisenberg model, see the discussion in C. Gros, Phys. Rev. B **42**, 6835 (1990).

[25]P. W. Anderson, Phys. Rev. **86**, 694 (1952).

[26]R. S. Fishman and S. H. Liu, Phys. Rev. B **40**, 11 028 (1989).

[27]S. Trugman and W. Stephan (private communication).

[28]W. Brenig (private communication), has also considered the intersublattice process in one dimension, where it is more important.

[29]One might be interested in the interaction of the charge motion with a long-wavelength spin distortion. In this case the appropriate starting point would be a classical spiral state.

[30]In our earlier work, Ref. 23, we used a notation with $h \rightarrow b$ and $d \rightarrow a$.

[31]T. Oguchi, Phys. Rev. B **117**, 117 (1960).

[32]A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, 1971), Sec. 9. See especially Eq. 9.3.

[33]R. R. P. Singh, Phys. Rev. B **39**, 9760 (1989).

[34]S. Trugman (private communication).

[35](a) B. I. Shraiman and E. D. Siggia, Phys. Rev. B **39**, 6880 (1989); (b) B. D. Simons and J. M. F. Gunn, *ibid.* **41**, 7019 (1990).

[36]W. Stephan, K. J. von Szczepanski, M. Ziegler, and P. Horsch, Europhys. Lett. **11**, 675 (1990).

[37]For the diagonalization method used, see Ref. 17(d) and also their Refs. 15 and 39.