Z$_2$ gauge theory of electron fractionalization in strongly correlated systems

T. Senthil and Matthew P. A. Fisher
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106–4030
(Received 25 October 1999)

We develop a new theoretical framework for describing and analyzing exotic phases of strongly correlated electrons which support excitations with fractional quantum numbers. Starting with a class of microscopic models believed to capture much of the essential physics of the cuprate superconductors, we derive a new gauge theory—based upon a discrete Ising or Z$_2$ symmetry—which interpolates naturally between an antiferromagnetic Mott insulator and a conventional d-wave superconductor. We explore the intervening regime, and demonstrate the possible existence of an exotic fractionalized insulator, the nodal liquid, as well as various more conventional insulating phases exhibiting broken lattice symmetries. A crucial role is played by vortex configurations in the Z$_2$ gauge field. Fractionalization is obtained if they are uncondensed. Within the insulating phases, the dynamics of these Z$_2$ vortices in two dimensions is described, after a duality transformation, by an Ising model in a transverse field, the Ising spins representing the Z$_2$ vortices. The presence of an unusual Berry’s phase term in the gauge theory leads to a doping-dependent “frustration” in the dual Ising model, being fully frustrated at half filling. The Z$_2$ gauge theory is readily generalized to a variety of different situations, in particular, it can also describe three-dimensional insulators with fractional quantum numbers. We point out that the mechanism of fractionalization for $d>1$ is distinct from the well-known one-dimensional spin–charge separation. Other interesting results include a description of an exotic fractionalized superconductor in two or higher dimensions.

I. INTRODUCTION

Strongly interacting many-electron systems in low dimensions can exhibit exotic properties, most notably the presence of excitations with fractional quantum numbers. In these instances the electron is “fractionalized,” effectively split-tered into constituents which essentially behave as free particles. The classic example is the one-dimensional (1D) interacting electron gas, which exhibits many anomalous properties such as the separation of the spin and the charge of the electron. Electron “fractionalization” is also predicted to occur in two-dimensional (2D) systems in very strong magnetic fields that exhibit the fractional quantum Hall effect. Recent experiments have given strong supporting evidence of fractionalization both in quantum Hall systems and in carbon nanotubes. Motivated by these examples, several authors have proposed the possibility of electron fractionalization in various other experimental systems. Perhaps the most tantalizing was the suggestion by Anderson of “spin-charge separation” in cuprate high-$T_c$ materials. However, this suggestion is currently surrounded by considerable controversy, in part because the 1D electron gas and the fractional quantum Hall effect appear to be rather special situations which do not readily generalize. Indeed, in 1D the Fermi liquid breaks down even at weak coupling and in the quantum Hall regime the kinetic energy is strongly quenched by a time reversal breaking magnetic field.

In this paper, we will explore theoretically the possibility of electron fractionalization in strongly correlated systems in spatial dimensions $d>1$ in the presence of time reversal symmetry. Our primary motivation is the cuprates, although we expect our results to be of significance to a variety of other strongly interacting systems. Early attempts to implement theoretically Anderson’s suggestion of 2D spin-charge separation typically started with either a quantum spin model or the $t$-$J$ model. Slave boson/fermion representations of the spin and electron operators were employed to obtain a mean field “saddlepoint” exhibiting spin-charge separation. The slave boson/fermion representation introduces a gauge symmetry, $U(1)$ in the simplest formulations, and requires inclusion of a corresponding gauge field. Fluctuations about the mean field theory lead to a strongly interacting gauge theory about which very little is reliably known. It is then quite difficult to reach any definitive conclusions about the true low energy behavior, in particular whether spin-charge separation survives beyond the mean field level. An alternate more recent approach describes strongly correlated electron systems in 2D in a dual language where the vortices in the many-electron wave function are the fundamental degrees of freedom. In this approach, insulating phases can be obtained by condensing vortices. Fractionalized insulators arise upon condensing pairs of vortices.

In this work we introduce a new gauge theory approach which enables us to reliably address issues of fractionalization. In contrast to the slave boson/fermion representation, our gauge symmetry is discrete, in fact, an Ising or Z$_2$ gauge symmetry. This has several advantages. First, gauge theories with discrete symmetry are much simpler to analyze than those with continuous symmetries, so it is possible for us to make definitive statements about low energy physics. But in addition, the pure Z$_2$ gauge theory in 2+1 space-time dimensions is dual to the three-dimensional (3D) classical Ising model, which implies the existence of two distinct quantum phases. In one of these two phases “charges” are deconfined, in marked contrast to the pure 2+1 dimensional $U(1)$ gauge theory which is always in a confining phase. The presence of deconfinement allows us to demonstrate the existence of insulating phases exhibiting electron fractional-
fractionalization in our $Z_2$ gauge theory approach is physically equivalent to vortex pairing in the earlier dual formulation. We demonstrate this equivalence by combining the standard boson-vortex duality with the Ising duality mentioned above.

In addition to the fractionalized phases, our approach allows us to readily access the more conventional confined phases and the concomitant confinement transitions. Furthermore, the $Z_2$ gauge theory can be readily generalized to describe a variety of different situations, arbitrary spatial dimensions, spin-rotation noninvariant systems, etc. Some of these generalizations are explored towards the end of the paper. For the most part, we concentrate on spin-rotation invariant electronic systems in 2D. An overview and summary of our main results may be found at the end of Sec. I.

In the context of frustrated quantum spin models, Read and Sachdev have demonstrated the possibility of disordered phases with fractionalization of spin. Specifically, an $Sp(2N)$ antiferromagnet at large $N$ and the related quantum dimer model were shown to reduce to a $Z_2$ gauge theory when frustration was present. In the deconfined phase of the gauge theory free propagating spinons (spin 1/2 excitations) would be possible. Somewhat similarly, in the slave-fermion representation of the conventional Heisenberg magnet which introduces an $SU(2)$ gauge invariance, Wen proposed obtaining fractionalization of spin by pairing and condensing pairs of spinons. This reduces the gauge symmetry down to $Z_2$. In contrast, we show explicitly that the conventional Heisenberg spin model can be written as a $Z_2$ gauge theory coupled to fermionic spinons, even in the absence of any frustration. The key observation is that, with fermionic spinons, the local constraint of single occupancy is equivalent to the constraint of an odd number of fermions per site. This latter constraint can be implemented with a discrete $Z_2$ gauge field. Such a $Z_2$ gauge description may also be obtained with the Majorana fermion representation of Heisenberg spins.

The basic physics underlying our description of electron fractionalization is perhaps most readily understood in $d = 2$. At the heart of quantum mechanics is wave-particle duality. For a many-body system of interacting bosons (with charge $Q_e$, for example) this dualism implies that in addition to the conventional "particle" framework, a description developed in terms of wave functions is possible. In 2D this dual wave description focuses on point like singularities in the phase of the complex wave function—the familiar vortices with circulation quantized in units of $Q_e$. A fundamental property of such vortices is that the product of their quantum of circulation and the particle charge is a constant,

$$Q_e Q_v = \text{H.c.}$$

It is this simple identity which underlies the two known examples of fractionalization in two dimensions, and is at the heart of the $Z_2$ gauge theory developed in this paper. In a (BCS) superconductor, the pairing of electrons to form a Cooper pair with charge $Q_e = 2e$, implies a "halving" of the flux quantum, $Q_v = \frac{\pi}{2} (hc/e)$, which is tantamount to "vortex fractionalization." The second example of 2D fractionalization occurs in the fractional quantum Hall effect. In the $\nu = 1/3$ state three vortices bind to each electron forming a "composite boson" with total circulation $Q_v = 3(hc/e)$, which then condenses. The above identity implies the existence of topological excitations in this condensate with electrical charge $\frac{e}{3}$ the celebrated Laughlin quasiparticles.

The route to electron fractionalization that we explore in this paper is physically equivalent to a pairing of vortices, precisely as in earlier work by Balents et al. But the mathematical implementation is rather different. Balents et al. argued that a pairing and condensation of conventional $Q_v = hc/2e$ BCS vortices in a singlet superconductor results in an exotic fractionalized insulator. As Eq. (1) demonstrates, this insulator should support spinless charge $e$ excitations. Our analysis begins by noting that such an excitation can be thought of as "one half" of a Cooper pair. We implement this fractionalization by formally re-expressing the Cooper pair creation operator as the product of two "chargon" operators, $b^\dagger$, each creating a spinless, charge $e$ boson. This change of variables introduces a local $Z_2$ symmetry, since it is possible to change the sign of $b^\dagger$ on any given lattice site while leaving the Cooper pair operator invariant. This is the origin of a local Ising, or $Z_2$, gauge symmetry, described mathematically in terms of a $Z_2$ gauge field. In the exotic fractionalized insulator, there are strange gapped excitations which are vortices in the $Z_2$ gauge field. These excitations, which we refer to as "visons" because they can be represented in terms of Ising spins, are the remnant of the unpaired $hc/2e$ BCS vortices, which survive in the fractionalized insulator. As we shall see, when the visons condense they drive "confinement," thereby destroying fractionalization. These visons will play an absolutely central role throughout this paper, since any insulator with gapped visons is necessarily fractionalized.

Motivated by the cuprate superconductors, we will focus on a particular class of microscopic lattice models designed to capture much of the physics believed essential to these materials. (Our description of fractionalization is, however, more general and is not restricted to these models.) The models describe electrons hopping on a lattice with inclusion of strong spin and pairing fluctuations, and are quite similar to models introduced and analyzed numerically by Assaad et al. and to models considered more recently by Balents et al. Many microscopic models of the cuprates, such as the $t$–$J$ model, incorporate spin fluctuations from the outset. Our reasons for similarly incorporating "microscopic" pairing fluctuations are twofold. First, as the superconducting phase is a well-established and reasonably well-understood part of the high-$T_c$ phase diagram, just like the antiferromagnet, it serves as a useful point of departure to access more puzzling regions of the phase diagram. This point of view was also advocated in Ref. 21. But there are also more microscopic reasons to include pairing fluctuations from the outset. In particular, as emphasized, for instance in Ref. 6, a spin–spin interaction term as in the $t$–$J$ model can be suggestively rewritten in terms of electron operators as

$$S_\nu S_\nu = -\frac{1}{2}(c_\nu^\dagger c_\nu^\dagger c_\nu c_\nu^\dagger) (\text{H.c}) + \frac{1}{2} \rho_\nu \rho_\nu,$$

with $\rho_\nu = c_\nu c_\nu^\dagger$. For antiferromagnetic exchange the first term is an attractive pairing interaction in the $d_{x^2-y^2}$ (or extended-
s) wave channel. As in BCS theory, this interaction may be decoupled (in a functional integral) with a complex auxiliary pair field \( \eta_{ij} \) as

\[
\sum_{(rr')} J [2 \eta_{rr'}|^2 + \{ \eta_{rr'} a_{rr'} (c_{r'i} c_{r'i}^\dagger - c_{r'i} c_{r'i}) + c.c.],
\]

Here \( a_{rr'} = +1 \) for bonds along the x direction, and equals \(-1\) for bonds along the y direction. With \( \eta \neq 0 \), this corresponds to a superconducting phase with \( d_{x^2-y^2} \) symmetry. But more generally, \( \eta \) can be decomposed into an amplitude and a phase, \( \eta = \Delta e^{i\varphi} \). Ignoring fluctuations in the amplitude leads to a model of the type we consider below, with local fluctuating \( d \)-wave pairing correlations.

Further motivation for inclusion of such pairing fluctuations is provided by resonating valence bond (RVB) ideas. The wave function for a RVB Mott insulator can be obtained from the wave function of a superconductor by Gutzwiller projecting into a subspace with exactly one electron per site. Some mean field theories of the RVB state are equivalent to starting out with just the superconducting wave function. Gauge field fluctuations about the mean field solution are supposed to carry out this highly nontrivial projection and destroy the superconductivity. A natural physical route to achieve this end is to include strong fluctuations of the mean field order parameter. Indeed, in a recent preprint it was argued that fluctuations about the mean field theory of the \( d \)-wave RVB state are formally equivalent to a theory of a phase-fluctuating \( d \)-wave superconductor.

With these motivations, we consider generalized Hubbard type models of the form

\[
H = H_0 + H_J + H_u + H_\Delta,
\]

with

\[
H_0 = -t \sum_{(rr')} c_{r'a}^\dagger c_{r'a} + \text{H.c.},
\]

\[
H_J = J \sum_{(rr')} S_{r'} \cdot S_{r'},
\]

\[
H_u = \sum_r u (N_r - N_0)^2,
\]

\[
H_\Delta = \sum_r (e^{i\varphi} p_r + \text{H.c.}),
\]

with the local \( d \)-wave pair field defined as

\[
p_r = \sum_{r'} \Delta_{r'r} (c_{r'} c_{r'}^\dagger - c_{r'}^\dagger c_{r'}).
\]

Here, \( c_{r'a} \) denotes an electron operator at site \( r \) of (say) a 2D square lattice with spin polarization \( \alpha = \uparrow, \downarrow \). The electron density and spin operators are the usual bilinears: \( p_r = c_{r'a}^\dagger c_{r'a} \) and \( S_{r'} = \frac{1}{2} c_{r'}^\dagger \sigma \cdot c_{r'} \) with \( \sigma \) a vector of Pauli matrices.

The term \( H_u \) is an on-site repulsion. Strong local tendencies for \( d_{x^2-y^2} \) pairing are incorporated through the term \( H_\Delta \). In the definition of \( p_r \) in Eq. (9), the summation is over the four nearest neighbors of the site \( r \) and \( \Delta_{r'r} = \Delta \) for bonds along the \( x \) direction and \( \Delta_{r'y} = -\Delta \) for bonds along the \( y \) direction. With this choice, the operator \( p_r \) destroys a \( d_{x^2-y^2} \) pair of electrons centered at the site \( r \).

As discussed above, this anomalous term can be obtained by decoupling a local spin exchange interaction—which is attractive in the \( d \)-wave pairing channel—with a complex Hubbard–Stratanovich field. Here, we keep the amplitude \( \Delta \) fixed, but include (quantum) fluctuations of the local pair field phase, \( \varphi_r \). This phase is canonically conjugate to the Cooper pair number operator, \( n_r \):

\[
[\varphi_r, n_{r'}] = i \delta_{rr'}.
\]

Due to the anomalous term in \( H_\Delta \), the two densities \( p_r \) and \( n_r \) are not separately conserved. The conserved electrical charge density is simply the sum of the Cooper pair and electron densities,

\[
N_r = 2n_r + p_r.
\]

It is this total density that enters into the local on-site Hubbard interaction term. The \( c \) number \( N_0 \) plays the role of a chemical potential, determining the overall electrical density.

This Hamiltonian describes interacting electrons in a system with strong local pairing and spin fluctuations. Since \( \varphi_r \) is a dynamical quantum field, these pairing fluctuations do not necessarily lead to a superconducting ground state. In addition to the pairing interaction terms, the above Hamiltonian includes interactions in the spin singlet (\( u \)) and spin triplet (\( J \)) particle/hole channels. The Hamiltonian retains the important global symmetries, corresponding to conservation of spin and electrical charge. It is worth emphasizing that the theoretical description of electron fractionalization that we develop below is not in the least restricted to this particular Hamiltonian.

### A. Overview

Due to the length of this paper, we first provide a brief synopsis of our approach and of the key results. We start with the observation of Kivelson and Rokhsar that, in an appropriate sense, a (singlet) superconductor already has separation of spin and charge. If one imagines inserting an electron into the bulk of a superconductor, its charge gets screened out by the condensate to leave behind a neutral spin-carrying excitation—a ‘‘spinon.’’ A mathematical implementation of this idea essentially amounts to binding half of a Cooper pair to an electron to produce a neutral spinon. Following these ideas, we first split the Cooper pair operator into two pieces, each piece creating an excitation with charge \( e \) but spin zero. These are the same quantum numbers as the ‘‘holon.’’ But since this object seems to be defined rather differently, and in any event is not directly tied to the doping of a Mott insulator, we prefer to refer to it as a ‘‘chargon.’’ The square of the chargon operator creates the Cooper pair. Next, we define a neutral spinon operator by multiplying the chargon and electron operators. Changing variables from the electrons and Cooper pairs to chargons and spinons introduces a degree of redundancy in the description. Specifically, all physical observables are invariant under a local change in the sign of the spinon and chargon operators. This implies that the resulting theory must have a local \( Z_2 \) gauge invariance.
In Sec. II, we carefully re-express the above model in terms of the chargon and spinon operators, paying special attention to the local $Z_2$ gauge symmetry. Following techniques familiar from slave boson/fermion theories, we derive an action in terms of the chargon and spinon fields coupled to a fluctuating $Z_2$ gauge field. This takes the form

$$S = S_c + S_s + S_B,$$  

(12)

$$S_s = -t_c \sum_{\langle ij \rangle} \sigma_{ij} (b_i^* b_j + c.c.),$$  

(13)

$$S_s = -\sum_{\langle ij \rangle} \sigma_{ij} (t_{ij}^a \bar{f}_{ia} f_{ja} + t_{ij}^d \bar{f}_{ij} f_{ji} + c.c.) - \sum_i \bar{f}_{ia} f_{ia}.$$  

(14)

Here $S_c$ describes the charge dynamics with $b_j = e^{-i \phi_j}$ the chargon field defined on a $d+1$ dimensional space-time lattice labeled by $i, j, \ldots$. The spin is carried by the (Grassmann-valued) spinon fields, $f_j$ and $\bar{f}_j$, also living on the lattice sites. The chargon and spinon fields are \textit{``minimally coupled''} to an Ising $Z_2$ gauge field $\sigma_{ij} = \pm 1$ living on the \textit{links} of the space-time lattice. The form of the charge and spin actions, $S_c$ and $S_s$, could have been guessed on symmetry grounds [the global charge $U(1)$, the global spin $SU(2)$ and the \textit{local} $Z_2$ gauge symmetry], but the derivation in Sec. II shows the presence of an additional term $S_B$. This is a \textit{``Berry phase''} term that takes the form

$$S_B = -i \sum_{i,j} N_0 \left[ 2\pi l_{ij} - \frac{\pi}{2} (1 - \sigma_{ij}) \right].$$  

(15)

Here $\tau$ refers to the time direction, and $l_{ij}$ is an integer on each temporal link defined in terms of the $\phi$ and $\sigma$ fields as

$$l_{ij} = \text{Int} \left[ \frac{\Phi_{ij}}{2\pi} + \frac{1}{2} \right],$$  

(16)

with $\Phi_{ij}$ the gauge invariant phase difference across the temporal link,

$$\Phi_{ij} = \phi_i - \phi_j + \frac{\pi}{2} (1 - \sigma_{ij}).$$  

(17)

The symbol $\text{Int}$ refers to the integer part. The Berry phase term simplifies considerably for integer $N_0$. For even integer $N_0$, we simply have $e^{-S_B} = 1$, while for odd integer $N_0$,$$

$$e^{-S_B} = \prod_{i,j=i-1}^{\tau} \sigma_{ij}, \quad N_0 \text{ odd.}$$  

(18)

A rough estimate of the dimensionless couplings $t_c, t^e, t^\Delta$ in terms of the parameters $t, u, J, \Delta$ of the original microscopic Hamiltonian may be obtained in the physically interesting limit of large $u$ and small $t$ near half filling:

$$t_c \sim \sqrt{\frac{t u}{J}} \quad t^e \sim \frac{J}{t} \quad t^\Delta \sim \frac{\Delta}{t} t_c.$$  

(19)

We will, however, regard these coupling strengths as phenomenological input parameters for the $Z_2$ gauge theory.

A great deal of physics is contained in the simple-looking action, Eq. (12). Consider varying the dimensionless chargon coupling, $t_c$, which represents the degree of charge fluctuations, and for simplicity specializing to half filling with $N_0 = 1$. Surprisingly, in the limit of vanishing chargon coupling, $t_c = 0$, the full $Z_2$ gauge theory action can be shown to be \textit{formally} equivalent (see Sec. IV) to the Heisenberg antiferromagnetic spin model. Increasing $t_c$ from zero introduces charge fluctuations into the Heisenberg model. In the limit of large $t_c$, the chargons will condense, resulting in a conventional $d_{x^2-y^2}$ superconductor. Thus, the above $Z_2$ gauge theory action has the remarkable property of interpolating between the Heisenberg antiferromagnet in one limit and a $d_{x^2-y^2}$ superconductor in the opposite limit. Determining the properties of this model in the intervening regime (with $t_c$ of order 1) is an extremely interesting question in the context of the cuprate materials, and will be one of the prime focuses of our analysis. Specifically, within the present $Z_2$ gauge theory we will explore the different possible routes between these two limits (which depend on the parameters in the action). Most important, for certain parameter regimes we will demonstrate the possibility of obtaining an exotic fractionalized insulating phase, dubbed the nodal liquid in previous work,\textsuperscript{21} an insulating phase, dubbed the nodal liquid in previous work,\textsuperscript{21} which in 2D exhibits Neél long-ranged order at zero temperature. The value of the parameter $K$ is determined by the chargon coupling, increasing monotonically with $t_c$. The full effective action appropriate to the insulating phases is simply

$$S = S_s + S_\sigma + S_B.$$  

(21)

Since the onset of superconductivity will occur at some critical value of order one, $t^e_\sigma \approx 1$, the validity of the effective action requires $t_c < t^e_\sigma$. Near this limit, but on the insulating side, $K$ will also be of order one. There are several limits in which the properties of this effective action may be reliably analyzed. A schematic phase diagram is shown in Fig. 1. As mentioned above, with $K = t_c = 0$ the action describes the Heisenberg antiferromagnet, which in 2D exhibits Neél long-ranged order at zero temperature. The opposite limit of large $K$ is far more interesting, though. Indeed, when $K = \infty$, fluctuations of the $Z_2$ gauge field $\sigma_{ij}$ are frozen, and one can set $\sigma_{ij} = 1$ on all the links. This results in a phase with deconfined spinons propa-
gating freely with the gapless "d-wave" dispersion, the "nodal liquid." Similarly, the chargons are also deconfined, and exist as gapped excitations in this insulating phase. The nodal liquid is thus a genuinely "fractionalized" insulator, within which the electron has splintered into two pieces that propagate independently. On reducing $K$ from $\infty$, the nodal liquid continues to be stable until a certain critical value $K_c$ of order one, where the gauge field undergoes a confinement transition. For $K<K_c$ the chargons and spinons are no longer legitimate excitations, but rather are confined together to form the electron (or other composites built from the electron such as magnons or Cooper pairs). This corresponds to a conventional insulating phase. As we argue in Sec. IV, the confinement transition is accompanied by breaking of translational symmetry leading to spin-Peierls order, at least for small spinon couplings $t_s^\alpha$. This may be understood from the limit when $t_s^\alpha=0$. Then, as we show in Sec. IV, we are left with a pure $Z_2$ gauge theory with the Berry phase term $S_B$ which is exactly dual to the fully frustrated Ising model in a transverse magnetic field. Ordering the Ising spins in this dual global Ising model leads to confinement. Physically, the Ising spins represent vortices in the $Z_2$ gauge field, namely, the vison excitations mentioned in the previous subsection. This same model also arose in the studies of Sachdev and Kternally, if conventional," with confinement of chargons and spinons. Alternately, if $K$ exceeds $K_c$ before the transition into the superconductor, the fractionalized nodal liquid phase will occur, sandwiched between the $d$-wave superconductor and a conventional insulator. Since both the superconducting and the deconfinement transitions occur at $t_c$ (and hence $K$) is of order one, the deconfinement boundary is expected to be "near" the onset of superconductivity. It is thus difficult to ascertain which of these two scenarios will be realized. The precise phase diagram interpolating between the antiferomagnet and superconductor will likely depend sensitively on various microscopic details.

Considerable further insight is provided into the mechanism of electron fractionalization in an alternate dual formulation in which we trade the chargon fields for the $hc/2e$ vortices which occur in a conventional superconductor. In Appendix B, we show how this may be done following standard duality transformations for the classical three-dimensional $XY$ model. Starting with the full $Z_2$ gauge theory in Eq. (12), the resulting dual theory is a lattice action for the $hc/2e$ vortices coupled to the spinons. The vortices see a fluctuating $U(1)$ field $\sigma$ through its circulation is the total electrical three current. Further, the $hc/2e$ vortices have a long-ranged statistical interaction with the spinons: When a spinon encircles such a vortex, its wave function acquires a phase of $\pi$. In the present formulation, a $\pi$ flux of the $Z_2$ gauge field $\sigma$ is effectively attached to each vortex. As the spinons are minimally coupled to $\sigma$, they acquire the expected phase of $\pi$ upon encircling each vortex. Mathematically, this flux attachment is implemented by an analog of a Chern–Simons term for the Ising group. Quite remarkably, this Ising Chern–Simons term emerges automatically from the duality transformation in Appendix D.

This dual representation of the $Z_2$ gauge theory is in fact essentially identical to the vortex field theory introduced in Ref. 9 on a phenomenological basis starting with a BCS superconductor. In that work, the statistical interaction between spinons and vortices was put in by hand, employing a $U(1)$ Chern–Simons terms to attach flux to the spin of the spinons. An advantage of the Ising Chern–Simons terms is that it does not break spin-rotational invariance, and in fact is possible even for spinless electrons. Moreover, it enables the description of an exotic superconducting phase in which the Ising flux de-attaches from the vortices (see below). In this dual description, superconducting phases correspond to vortex vacua, while insulating phases correspond to vortex condensates. Simply condensing the $hc/2e$ vortices leads to confined insulating phases. Accessing deconfined insulating phases requires condensation of paired vortices, without condensation of single ones. In this way one obtains an alternate dual description of the fractionalized nodal liquid. The $Z_2$ gauge theory formulation suggests a mechanism for such vortex pairing: Since the chargons also have a long-ranged statistical interaction with $hc/2e$ vortices, their motion is "frustrated" in the presence of such vortices. Pairing the vortices reduces this frustration, allowing the charge to propagate more easily, and lowering the kinetic energy.

Superconducting phases are readily accessed in either the $Z_2$ gauge theory "particle" formulation of Eq. (12) or its dual vortex counterpart. In the particle formulation, when $t_c$ becomes large and the charge $e$ chargons condense, the result is a $d_{s-\gamma}$ superconductor, denoted $dSC$. This superconductor is conventional, perhaps surprising since BCS theory involves the condensate of a charge $2e$ Cooper pair. But as we demonstrate in Sec. III, the chargon condensate supports $hc/2e$ vortices, and shares all other properties with a conventional BCS superconductor. It is interesting to ask if it is possible to have a superconductor where the chargon pairs have condensed, while the single chargons have not. Such a
superconductor, which we label $dSC^*$, can be readily described with the present $Z_2$ gauge theory formulation. As detailed in Sec. VII, $dSC^*$ is a truly exotic superconducting phase with many unusual properties.

The $Z_2$ gauge theory is readily generalized to a wide variety of other situations. In particular, the particle formulation of Eq. (12) is valid in any spatial dimension. In 3D there again exist fractionalized insulating phases (and, of course, confined ones) which can be accessed by the theory. Remarkably, as we argue in Sec. VIII B, in contrast to the 2D case, a fractionalized insulator in 3D exists as a distinct finite temperature phase, separated by a classical phase transition from the high temperature limit. For an anisotropic layered three-dimensional material, it is also possible to have another 3D fractionalized phase consisting of weakly coupled 2D phases, but this phase is destroyed by thermal fluctuations. It is also of note that the $Z_2$ gauge theory formulation seems incapable of describing fractionalization in 1D. This indicates that the “solitonic” mechanism of fractionalization in $d=1$ is qualitatively different than “vortex pairing” which describes fractionalization in higher dimensions.

We conclude Sec. I with an outline of the rest of the paper. Section II contains the formal derivation of the $Z_2$ gauge theory from the microscopic models. For ease of presentation, and as it is simpler, we will first provide the technical details of the derivation for situations with local $s$-wave pairing. (In Appendix B, we show how situations with $d$-wave pairing, the case of interest for the cuprates, can be readily handled). We next describe, in Sec. III, the physics of fractionalization and confinement in the simplest possible context, that of $s$-wave pairing with an even number of electrons per unit cell. We then consider in Sec. IV the more interesting situation of $d$-wave pairing with an odd number of electrons per unit cell. In Sec. V we formulate and develop the dual description in terms of vortices. The results of Sec. IV are reobtained in this representation. We then move on in Sec. VI to show how doping away from half filling may be incorporated into the formalism. In Sec. VII, we discuss the possibility of other exotic fractionalized phases, in particular the superconductor $SC^*$ mentioned above, in both the particle and vortex formulations. In Sec. VIII we discuss various generalizations of the theory, including spatial dimensions other than two, finite temperature, and situations with no spin rotational invariance. We also briefly discuss a useful analogy with $Z_2$ lattice gauge theories of classical nematic systems. In Sec. IX, we discuss the relationship between this work and several other previous approaches to fractionalization in strongly correlated systems. Contact will be made, when possible, with the earlier dual vortex descriptions of the nodal liquid, and with the slave boson/fermion approaches. Section X contains a discussion of the experimental signatures of the various novel phases described in earlier sections. We conclude with a summary of our main results. Various appendices contain technical details not presented in the main text.

II. MODELS AND $Z_2$ GAUGE THEORY

To describe our techniques in the simplest possible context, we will start with a microscopic model that has local $s$-wave pairing correlations. This can be readily generalized to other symmetries such as $d$ wave (see the end of Sec. II, and Appendix B). Of course, with strong local on-site repulsion (positive $u$ above) $d$-wave pairing fluctuations are presumably more energetically viable, and also of central interest in the context of cuprate superconductivity.

Consider then a generalized Hubbard type model,

$$H = H_0 + H_a + H_J + H_\Delta,$$  \hspace{1cm} (22)

with

$$H_0 = -t \sum_{\langle rr' \rangle} c_{r, \alpha}^\dagger c_{r', \alpha} + H.c.,$$  \hspace{1cm} (23)

$$H_a = \sum_r u(N_r - N_0)^2,$$  \hspace{1cm} (24)

$$H_J = J \sum_{\langle rr' \rangle} \left[ S_r \cdot S_{r'} + \frac{1}{4} \rho_r \rho_{r'} \right],$$  \hspace{1cm} (25)

$$H_\Delta = \Delta \sum_r (e^{i \varphi_r} c_{r, \uparrow}^\dagger c_{r, \downarrow}^\dagger + H.c.).$$  \hspace{1cm} (26)

As earlier, $c_{r, \alpha}$ denotes an electron operator at site $r$ with spin $\alpha$ and the electron density and spin operators are the usual bilinears: $\rho_r = c_{r, \alpha}^\dagger c_{r, \alpha}$ and $S_r = \frac{1}{2} \sum_{\alpha \neq \beta} \alpha c_{r, \alpha}^\dagger \alpha c_{r, \beta}$. This Hamiltonian is essentially the same as Eq. (4) in Sec. I, except that it has local $s$-wave pairing rather than $d$ wave, and we have added a term proportional to $\rho_r \rho_{r'}$ in $H_J$. These modifications have been made to simplify both the derivation and the subsequent analysis of the $Z_2$ gauge theory. We return later to the more physically interesting case of local $d$-wave pairing.

Here, $\varphi_r$ is the phase of a local $s$-wave Cooper pair field and is canonically conjugate to the Cooper pair number operator, $n_r = [\varphi_r, n_r'] = i \delta_{r, r'}$. As before, since $\varphi_r$ is a dynamical quantum field, these pairing fluctuations do not necessarily lead to a superconducting ground state. The conserved electrical charge density is the sum of the Cooper pair and electron densities,

$$N_r = 2n_r + \rho_r.$$  \hspace{1cm} (27)

A. Split the Cooper pair

We now proceed to split the Cooper pair into two pieces. Consider an operator $b_r^\dagger$ defined as

$$b_r^\dagger = s_r e^{i \varphi_r / 2} = e^{i \phi_r},$$  \hspace{1cm} (28)

with $s_r = \pm 1$ an Ising ‘‘spin’’ variable. With this definition the new field,

$$\phi_r = \frac{\varphi_r}{2} + \frac{\pi}{2}(1 - s_r),$$  \hspace{1cm} (29)

can be treated as a phase lying in the interval zero to $2\pi$, with $b_r^\dagger$ invariant under the transformation: $\varphi_r \rightarrow \varphi_r + 2\pi$ and $s_r \rightarrow -s_r$. The square of $b_r^\dagger$ creates a Cooper pair,

$$e^{i \varphi_r} = (b_r^\dagger)^2,$$  \hspace{1cm} (30)
so that $b_r^\dagger$ creates a spinless excitation with charge $e$, essentially one half of a Cooper pair. We refer to this operator as a chargon operator.

In order to separate out the charge and spin degrees of freedom it will be extremely useful to define an electrically neutral but spin carrying fermion operator (a spinon):

$$f_{ra}^\dagger = b_r c_{ra}^\dagger.$$  

(31)

This operator carries the spin of the electron, but is electrically neutral as verified by noting that it commutes with the total electrical charge density $N_r$. On the other hand, the chargon is electrically charged, and its phase is canonically conjugate to the total electrical charge density,

$$[\phi_r, N_r] = i \delta_{rr'}.$$  

(32)

At this stage it is legitimate to implement an operator change of variables in the full Hamiltonian, replacing the electron and Cooper pair operators $(\varphi, n, c, c^\dagger)$ by chargons and spinons $(\phi, N, f, f^\dagger)$. This gives

$$H = H_0 + H_u + H_J + H_\Delta,$$  

(33)

with

$$H_0 = -t \sum_{(r \alpha)} b_r^\dagger b_r f_{ra}^\dagger f_{ra}^\dagger + \text{h.c.},$$  

(34)

$$H_\Delta = \Delta \sum_r (f_r f_r^\dagger + \text{H.c.}),$$  

(35)

with $H_u$ unchanged and $H_J$ of the same form as in Eq. (25) but with spinon operators replacing the electron operators: $\rho_r = f_{ra}^\dagger f_{ra}$ and $S_r = \frac{1}{2} f_{ra}^\dagger \alpha f_{ra}$.

There are several extremely important points to stress about this seemingly innocuous change of variables. First, one can change the sign of both the chargon and spinon operators on any given site $r$,

$$b_r \rightarrow -b_r, \quad f_{ra} \rightarrow -f_{ra},$$  

(36)

without affecting the original Cooper pair or electron operators. This implies that quite generally the transformed Hamiltonian must also be invariant under this local Ising $Z_2$ symmetry, as can be readily checked in Eqs. (34) and (35). As we shall shortly see, in a path integral formulation this local $Z_2$ symmetry will be manifest in terms of a $Z_2$ gauge field. Second, because of this redundancy introduced into the change of variables, a constraint must be imposed on the Hilbert space spanned by the spinon and chargon operators.

To understand the origin of this constraint, consider first the Hilbert space of the original Hamiltonian. In a number-diagonal basis, the Hilbert space on each site $r$ is a direct product of states with an arbitrary integer number of Cooper pairs $(n_r)$ and the four electron states consistent with Pauli empty, doubly occupied or singly occupied with an electron of either spin. Since the chargon has only one half the charge of the Cooper pair, the full Hilbert space spanned by the chargon and spinon operators is actually twice as large, and it is essential to project down into the physical Hilbert space of electrons and Cooper pairs. From Eq. (27), it is clear that this can be achieved by imposing a constraint that the sum (or difference) of the number of chargons ($N_r$) and spinons ($\rho_r = f_{ra}^\dagger f_{ra}$) on each site is an even integer:

$$(-1)^{N_r + \rho_r} = 1.$$  

(37)

This implies, for example, that a site with a single chargon but no spinon is unphysical and forbidden, whereas a spinon and chargon together (an electron) is allowed.

### B. Path integral and $Z_2$ gauge theory

The most convenient way to implement the constraint on the spinon and chargon Hilbert space is in a (Euclidean) path integral representation of the partition function. To this end we define a projection operator,

$$\mathcal{P} = \prod_r \mathcal{P}_r,$$  

(38)

with

$$\mathcal{P}_r = \frac{1}{2} [1 + (-1)^{N_r + \rho_r}] = \frac{1}{2} \sum_{\sigma_r = \pm 1} e^{i(\pi/2)(1-\sigma_r)(N_r + \rho_r)},$$  

(39)

which projects into the physical Hilbert space. Here, $\sigma_r = \pm 1$ is an Ising-like field and $\rho_r = f_{ra}^\dagger f_{ra}$. As can be verified directly from Eq. (33), this projection operator commutes with the chargon–spinon Hamiltonian,

$$[\mathcal{P}, H] = 0,$$  

(40)

so that the Hamiltonian does not cause transitions out of the physical Hilbert space.

The partition function can be written as

$$Z = Tr[e^{-\beta H \mathcal{P}}],$$  

(41)

where the trace is over the full Hilbert space spanned by the chargon and spinon operators $(\phi, N, f, f^\dagger)$. A Euclidean path integral representation can be obtained as usual by splitting the exponential,

$$Z = Tr[(e^{-\beta H})^M],$$  

(42)

with $M$ ‘‘time slices’’ and $\epsilon = \beta M$. Here, we have inserted projection operators into each time slice. Working with fermion coherent states and eigenstates of the chargon phase $\phi$, a path integral representation can be readily derived, detailed in Appendix A, giving

$$Z = \int \prod_{i\alpha} df_{ia} df_{ia}^\dagger d\phi_i \sum_{N_i = \text{even}} \sum_{\sigma_r = \pm 1} e^{-S},$$  

(43)

where the integration is over Grassman numbers $f$ and $\tilde{f}$ and a $c$-number phase $\phi$ in the interval zero to $2\pi$. Here, $i = (r, \tau)$ runs over the $2 + 1$-dimensional space time lattice with $\tau = 1, 2, \ldots, M$ time slices. The Euclidian action takes the form,

$$S = S_\tau + S_\phi + \epsilon \sum_{r = 1}^M H(N_r, \phi_r, \tilde{f}_r f_r),$$  

(44)

with
Here, we have suppressed the explicit $r$ and $\alpha$ subscripts on the fields, displaying only the time-slice dependencies. As usual, the bosonic phase field and the Ising field both have the expected periodic boundary conditions, whereas the fermions are antiperiodic:

$$\phi_{r=M+1} = \phi_{r=1}, \quad \sigma_{M+1} = \sigma_{1}, \quad f_{M+1} = -f_{1}. \quad (47)$$

Notice that the Ising variables live on the links connecting adjacent time slices, and can thus be correctly interpreted as a gauge field. In fact, the Ising field $\sigma$ is minimally coupled to both spinons and chargons as the time component of a gauge field. Moreover, the local $Z_2$ symmetry of the Hamiltonian in Eq. (33), is manifest in the path integral as a full fledged Ising $Z_2$ gauge symmetry:

$$f_{i\alpha} \rightarrow \epsilon f_{i\alpha}, \quad \bar{f}_{i\alpha} \rightarrow \epsilon \bar{f}_{i\alpha}, \quad \phi_{i} \rightarrow \phi_{i} + \frac{\pi}{2} (1 - \epsilon_{i}), \quad (48)$$

together with a transformation of the gauge field,

$$\sigma_{ij} \rightarrow \epsilon_{ij} \sigma_{ij} \epsilon_{ij}. \quad (49)$$

Here, $\epsilon_{i}=\pm1$, and $\sigma_{ij}$ lives on the link connecting two “nearest neighbor” space-time lattice points, differing by one time slice.

Our final goal is to beat the model into a form which also includes $Z_2$ gauge fields on the spatial links, so that space and time end up on more equal footing. Our approach follows closely the standard methods employed in slave fermion or slave boson treatments of Heisenberg magnets. First, we perform a Hubbard–Stratanovich decoupling of the spin interaction terms in the Euclidian action:

$$e^{-eH_j} = \int \prod_{(rr')} \prod_{\tau} d\chi_{rr} (\tau) d\chi_{rr}^* (\tau) e^{-S_{hs}}. \quad (50)$$

The terms in $S_{hs}$ correspond to the hopping of spinons and chargons in the presence of a common fluctuating gauge field, $\chi$, on the near neighbor links.

Up to this stage, all of the formal manipulations that we have performed have been exact, so that the full Euclidian action gives a faithful representation of the original microscopic electron Hamiltonian. But now, following standard slave fermion/boson techniques, we perform an approximation, treating the functional integral over the Hubbard–Stratanovich field, $\chi$, within a saddlepoint approximation. (While it might be possible to find an appropriate “large-$N$” generalization of the model for which this approximation becomes exact, we do not pursue this tack here.) The simplest saddlepoint corresponds to setting all of the link fields equal to a single real constant: $\chi_{rr} = \chi_{0}$. The saddlepoint value for $\chi_{0}$ can in principle be obtained by integrating out the spinons (which are Gaussian) and the chargons (which are not). This saddlepoint respects two important discrete symmetries of the model, translational and time-reversal invariance. But the saddlepoint does not respect the $Z_2$ gauge symmetry in Eqs. (48) and (49). This serious flaw can be easily remedied though by retaining a particular set of fluctuations about the saddlepoint. The simplest choice consistent with the $Z_2$ gauge symmetry corresponds to allowing the sign of $\chi_{rr}$ to change, keeping the magnitude fixed, putting

$$\chi_{rr} = \sigma_{rr} \chi_{0}. \quad (55)$$

Here, $\sigma_{rr} (\tau) = \pm 1$ is a set of Ising fields living on the spatial links of the space-time lattice. Within this restricted manifold the theory consists of chargons and spinons hopping on a space-time lattice, minimally coupled to an $Z_2$ gauge field. Note that the fluctuations in the magnitude $\chi_{0}$ of the saddlepoint value of $\chi$ have been ignored in Eq. (55)—these “massive” fluctuations are expected to be unimportant for the issues we address in this paper.

Hereafter we work under this fixed-magnitude approximation. Within this approximation the full partition function can be expressed as a functional integral,

$$\tilde{Z} = \int \prod_{i=0}^{\infty} d\bar{f}_{ia} df_{ia} d\phi_{i} \sum_{N_{i}=0}^{\infty} \prod_{(ij)} \sum_{\sigma_{ij}=\pm1} \sum_{s_{i}} e^{-S_{s}}, \quad (56)$$

with $Z_2$ gauge fields $\sigma_{ij}$ living on the near neighbor links of the space-time lattice, and

$$S = S_{r} + S_{\phi} + S_{0} + S_{s} + S_{\Delta}. \quad (57)$$

with

$$S_{r} = \sum_{i,j=0}^{\infty} [\bar{f}_{ia}(\sigma_{ij} f_{ja} - f_{ij})], \quad (58)$$

$$S_{\phi} = -\sum_{i,j=0}^{\infty} N_{i}(\phi_{i} - \phi_{j} + \frac{\pi}{2} (1 - \sigma_{ij})), \quad (59)$$

$$S_{s} = \epsilon u \sum_{i} (N_{i} - N_{0})^{2}. \quad (60)$$

$$S_{\Delta} = \epsilon \sum_{i} \sum_{\tau} (H_{u} + H_{\Delta}) + S_{x}, \quad (53)$$

with

$$S_{x} = \epsilon \sum_{i} \sum_{\tau} (H_{u} + H_{\Delta}) + S_{x}. \quad (53)$$
\[
S_\Delta = \epsilon \Delta \sum f_{i} f_{i+\hat{t}} + \bar{f}_{i} \bar{f}_{i+\hat{t}},
\]
(61)

\[
S_0 = -\epsilon \sum_{i,j} \sigma_{ij} (t_{ij} b_{i}^{*} b_{j} + J_{0} f_{i} f_{j} + \text{c.c.}),
\]
(62)
where we have defined \( t_{0} = 2t \chi_{0} \) and \( \delta_{0} = J \chi_{0} \).

Notice that the full action is local in the integers \( N \), so the summation can be performed independently at each space-time point. A straightforward Poisson resummation gives

\[
\sum_{N} e^{-\left( S_{\Delta} + S_{\phi}^{h} \right)} = \exp \left( \sum_{i,j} V(\Phi_{ij}) \right),
\]
(63)
where \( \Phi_{ij} = \phi_{i} - \phi_{j} + (\pi/2)(1 - \sigma_{ij}) \) is the gauge invariant phase difference along a temporal link. Here, the periodic potential \( V(\Phi) \) is given by

\[
e^{V(\Phi)} = \sum_{l = -\infty}^{\infty} e^{-\left(1/4\epsilon u\right)(\Phi - 2\pi l)^{2} + i N_{0}(2\pi l - \Phi)},
\]
(64)
and we have dropped an overall multiplicative constant. In the limit of small \( \epsilon u \), the sum over \( l \) will be dominated by precisely one term which minimizes \( |\Phi - 2\pi l| \). This occurs for integer \( l \) satisfying \( |\Phi - 2\pi l| < \pi \) or, equivalently,

\[
l = \text{int} \left( \frac{\Phi}{2\pi} + \frac{1}{2} \right).
\]
(65)
Moreover, for small \( \epsilon u \) we may approximate

\[
e^{-\left(1/4\epsilon u\right)(\Phi - 2\pi l)^{2} - \epsilon^{-1/2}\epsilon u\left[1 - \cos(2\pi l)\right]},
\]
(66)
\[
e^{-\epsilon^{-1/2}\epsilon u\left[1 - \cos(\Phi)\right]},
\]
(67)
Within this approximation the sum over \( l \) becomes simply

\[
e^{V(\Phi)} \approx e^{-(1/4\epsilon u)(\Phi - 2\pi l)^{2} + i N_{0}(2\pi l - \Phi)},
\]
(68)
with \( l \) given by Eq. (65). We have again dropped an overall multiplicative constant.

The full \( N \) sum in the action then leads to

\[
\sum_{N} e^{-\left( S_{\Delta} + S_{\phi}^{h} \right)} = e^{S_{\phi} + S_{\psi} + i \epsilon N_{0}(2\pi l - \Phi) - S_{B}},
\]
(69)
with the Berry phase term \( S_{B} \) given by

\[
S_{B} = -i N_{0} \sum_{i,j} \left( 2\pi l_{ij} - \Phi_{ij} \right),
\]
(70)

\[
= -i N_{0} \sum_{i,j} \left[ 2\pi l_{ij} - \frac{\pi}{2} (1 - \sigma_{ij}) \right].
\]
(71)
In obtaining the last line, we have re-expressed \( \Phi_{ij} \) in terms of \( \phi \) and \( \sigma \), and used the \( \beta \)-periodic boundary conditions on \( \phi \) to drop the term involving \( \phi_{i} - \phi_{j} \). The Berry phase term is the only term in the action which depends on the (average) occupation number per unit cell, \( N_{0} \). It simplifies considerably for integer \( N_{0} \). For even integer \( N_{0} \), we simply have \( e^{-S_{B}} = 1 \), while for odd integer \( N_{0} \),

\[
e^{-S_{B}} = \prod_{i,j = i-\hat{t}} \sigma_{ij}, \quad N_{0} \text{ odd.}
\]
(72)
As we shall see, the Berry’s phase term will lead to subtle yet important differences between Mott insulators with odd integer \( N_{0} \) and band insulators with even \( N_{0} \).

The Euclidian path integral is only identical to the Hamiltonian formulation in the strict \( \epsilon \rightarrow 0 \) limit. But since the original lattice Hamiltonian is already an effective low energy theory, the time continuum limit which involves arbitrarily high energies is not actually of interest. For these reasons, hereafter we keep \( \epsilon \) finite, viewing it as an inverse “high energy” cutoff in the theory. Since the kinetic (\( t \)) and interaction (\( u \)) energy scales are the largest in the theory, it is convenient to choose the value of \( \epsilon \) so that the charge sector of the theory is isotropic on the \( 2 + 1 \)-dimensional space-time lattice. To this end, we require that the spatial chargon hopping strength equals the temporal one: \( 1/2\epsilon u = 2\epsilon t_{0} \), which implies

\[
\frac{1}{\epsilon} = 2\sqrt{\epsilon t_{0}} u.
\]
(73)
Note that the choice of the value of \( \epsilon \) only modifies slightly the physics at the highest energy scales, set by \( t \) and \( u \). The details of the model at these high energy scales will not significantly affect the low energy physics. \(^{27}\)

With this choice of \( \epsilon \) the full Euclidian action reduces to a much simpler and more compact form,

\[
S = S_{\phi} + S_{\psi} + S_{B},
\]
(74)
with

\[
S_{\phi} = -t \sum_{\langle ij \rangle} \sigma_{ij} (b_{i}^{*} b_{j} + \text{H.c.)},
\]
(75)

\[
S_{\psi} = \sum_{\langle ij \rangle} - (t'_{ij} \sigma_{ij} f_{i} f_{j} + \text{c.c.)} + \delta_{ij} (t'^{2} f_{i} f_{j} + \text{c.c.)} - \bar{f}_{i} f_{i}),
\]
(76)
and \( S_{B} \) as defined above. Here, the dimensionless chargon coupling strength is given in terms of the microscopic parameters \( t \), \( u \), and \( \chi_{0} \) to be

\[
t_{c} = \epsilon t_{0} = \frac{\sqrt{\chi_{0}}}{2u}.
\]
(77)
The dimensionless spinon coupling along the nearest neighbor spatial links is

\[
t'_{ij} = \epsilon J_{0} = J \frac{\sqrt{\chi_{0}}}{8u t_{c}}
\]
(78)
whereas \( t'_{ij} = -1 \) along the neighboring temporal links. Similarly, the coupling constant for the spinon pairing is

\[
t'_{xy} = \frac{\Delta}{\sqrt{8t_{c} \chi_{0} u}}.
\]
(79)
As will be shown in Sec. IV, for the physically interesting case of \( d \)-wave pairing near half filling, the parameter \( \chi_{0} \) may be roughly estimated to be
This can be used to obtain rough estimates of the three dimensionless coupling constants, \( t_c \), \( t' \), and \( t^A \). For the most part, however, we will treat these couplings as phenomenological parameters.

The partition function involves an integration over the on-site chargon phase \( \phi_i \) and spinon Grassmann fields \( \overline{f}_i, f_j \), as well as a summation over the \( Z_2 \) gauge fields \( \sigma_{ij} = \pm 1 \) which live on the nearest neighbor links of the Euclidian space time lattice. This “final” form for the theory is exceedingly simple, consisting of chargons and spinons hopping around, minimally coupled to a dynamical \( Z_2 \) gauge field. This form could have essentially been guessed just using knowledge of the field content (chargons and spinons) and the required symmetries: \( U(1) \) charge conservation, \( SU(2) \) spin conservation and the local \( Z_2 \) gauge symmetry. Perhaps the only subtlety is the presence of the term \( S_g \) in the action when the filling factor \( N_0 \) is not an even integer. Among the additional terms which are allowed by these symmetries is a field strength term for the \( Z_2 \) gauge field:

\[
S_\sigma = -K \sum \prod \sigma_{ij}.
\]

(81)

Here, the product denotes the gauge invariant product of the Ising fields around an elementary plaquette. This Ising field strength is then summed over all space-time plaquettes. Clearly, \( S_\sigma \) is the direct Ising analog of the \( F^2_{\mu\nu} \) term which enters the Lagrangian of ordinary \( U(1) \) electromagnetism. Even though not present in the derivation presented here, this field strength term will be generated upon integrating out the chargon or spinon matter fields, as discussed below.

In Appendix B we show how the above analysis can be generalized to the case in which local \( d \)-wave pairing correlations are incorporated from the outset as in the Hamiltonian Eq. (4), rather than \( s \)-wave as assumed above. The derivation of the effective \( Z_2 \) gauge theory proceeds in much the same fashion, and one arrives at the same model except with the spinon action given instead by

\[
S_s = \sum_{\langle ij \rangle} \sigma_{ij}(t_{ij}^* \overline{f}_{ia} f_{ja} + t^A_{ij} \overline{f}_{ij} f_{ij} + c.c.) - \sum_i \overline{f}_{ia} f_{ia}.
\]

(82)

Here, \( t_{ij}^A \) denotes a \( d \)-wave pairing amplitude living on the nearest neighbor spatial bonds, with amplitude \( + t^A \) on the \( x \)-axis bonds and \( - t^A \) along the \( y \)-axis bonds. Notice that the \( Z_2 \) gauge field \( \sigma_{ij} \) enters here because the \( d \)-wave pair field lives on the \( s \) link. This form exhibits the required Ising \( Z_2 \) gauge symmetry, being invariant under the transformation in Eq. (48). As shown in Sec. IV, a rough estimate of the various coupling constants in this case is

\[
t_c \sim \left( \frac{\sqrt{\mu u}}{J} \right)^{1/3} \sqrt{\frac{t}{u}}, \quad t' \sim \left( \frac{J}{t} \right) t_c, \quad t^A \sim \frac{\Delta}{t} t_c.
\]

(83)

Here \( t' \) and \( t^A \) refer only to the spatial couplings. But we will once again regard these as phenomenological parameters.

### III. Fractionalization and Confinement

Here in Sec. III we will analyze some of the phases which are described by the \( Z_2 \) gauge theory model derived in Sec. II. While the \( Z_2 \) gauge formulation is valid in general dimension, for concreteness and simplicity we specialize to two dimensions, generalizing briefly to other dimensions in Sec. VIII A. Moreover, for illustrative purposes we focus first on the simplest case with an even number of electrons per site (unit cell), and presume the presence of local \( s \)-wave pairing correlations. As we shall see, in this case the model can exhibit a conventional band insulator. More interesting, for certain parameter regimes, fractionalized insulating phases also become possible. Note that the microscopic Hamiltonian in Eq. (22) allows for charge fluctuations even when the total charge per site is even. In models of interacting electrons in the "idealized" limit of a single band, an even charge per site implies no charge fluctuations, and is trivial. However, away from this idealized limit, even occupation does not imply no charge fluctuations, and may be nontrivial.

In Sec. IV we will turn to the more physically interesting situation with an odd number of electrons per site. At that stage we will focus on local \( d \)-wave pairing correlations, which are more tenable in the presence of a large positive on-site Hubbard \( u \) as well as being of direct relevance to the cuprates. Doping away from half filling will be discussed in Sec. VI.

With even integer \( N_0 \) and local \( s \)-wave pairing correlations the full action consists of two contributions, \( S = S_c + S_s \), corresponding to the charge and spin sectors, respectively.

\[
S_c = -t_c \sum_{\langle ij \rangle} \sigma_{ij}(b_i^* b_j + c.c.),
\]

(84)

\[
S_s = -\sum_{\langle ij \rangle} t_{ij}^A \sigma_{ij}(\overline{f}_{ij} f_{ij} + c.c.) - \sum_i \overline{f}_{ia} f_{ia}.
\]

(85)

\[
+ t^A \sum_i (f_{i1}^* f_{i1} + c.c.).
\]

(86)

The first term, which describes the dynamics of the chargons, \( b^* = e^{i\phi} \), minimally coupled to an \( Z_2 \) gauge field, exhibits the global \( U(1) \) charge conservation symmetry. The spinons also carry the \( Z_2 \) Ising “charge.” Due to the \( s \)-wave form of the anomalous “pairing” term, the spinons, which are paired into singlets, should be gapped out.

### A. Correlated “band” insulators

We first consider electrically insulating states. When the dimensionless chargon coupling \( t_c \) is much smaller than unity, the chargons cannot propagate at low energies and a charge gap results. In this case, with both spinons and chargons gapped out, it is possible to integrate them out from the theory, leaving the \( Z_2 \) gauge field \( \sigma \) as the only remaining field. This integration will generate additional terms in the Lagrangian, depending on \( \sigma \), which will be local in space time and must also be gauge invariant. The most important such term \( S_{\sigma} \) is simply,
which describes a pure $Z_2$ gauge theory.

Remarkably, this simple gauge theory exhibits a phase transition as the coupling $K$ is varied. Indeed, as shown originally by Wegner,\textsuperscript{12,11} the pure $Z_2$ gauge theory in 3D is dual to the familiar three-dimensional Ising model:

\begin{equation}
S_{\text{dual}} = -K \sum_{\langle ij \rangle} v_i v_j ,
\end{equation}

with Ising spins, $v_i = \pm 1$, living on the sites of the dual lattice. The dimensionless Ising model coupling, $K_d$, is simply related to $K$: $\tanh(K_d) = e^{-2K}$. This form shows that the high and low "temperature" phases are exchanged under the duality transformation. The details of this duality transformation are given in Appendix C.

As emphasized originally by Wilson,\textsuperscript{29} a direct characterization of the two phases of the pure gauge theory is given in terms of the correlator,

\begin{equation}
G_C = \langle \prod_i \sigma_{ij} \rangle ,
\end{equation}

where the average is for the pure gauge theory and the product is taken around a closed loop in space time, denoted $C$. For $K < K_c$, the Wilson loop satisfies an "area law," with $G_C \sim \exp(-c A)$, with loop area $A$, and $c$ a $K$-dependent constant. When $K > K_c$, $G_C$ decays more slowly, only exponentially with the perimeter of the loop.

What do these two phases correspond to in physical terms? Consider first the large $K$ limit, which is the high temperature phase of the dual Ising model. As $K \to \infty$ all of the gauge field plaquette sums will be equal to $+1$. In this case it is possible to choose a gauge in which all of the Ising link variables are also unity, $\sigma_{ij} = 1$. In this phase the chargons and spinons can propagate at energies above their respective gaps. Apparently, the Hamiltonian contains gapped excitations which carry the quantum numbers of spinons and chargons. The electron has effectively been fractionalized! We denote this exotic insulating state with deconfined chargons and spinons as $T^\sigma$. It is exceedingly important to emphasize that the splintering of the electron into spin and charge carrying constituents is conceptually unrelated to the presence or absence of spin order. Indeed, electron fractionalization can occur even in the presence of strong spin–orbit interactions which destroys spin-rotational invariance—in that case the states of the fermionic $f$ particles cannot be labeled by spin.

As the coupling $K$ is reduced, so long as the gauge theory is in its perimeter phase, the energy to separate particles carrying the $Z_2$ charge remains finite, even for infinite separation. The chargons and spinons are deconfined. Further, with $K < \infty$, configurations of the $Z_2$ gauge theory with plaquette products equal to $-1$ will become possible. One can think of such plaquettes as being "pierced" by nonzero $Z_2$ flux or $Z_2$ vorticity. Because the number of such plaquettes on any given elementary space-time cube is even, the fluxes form "tubes," analogous to Abrikosov vortices in a type II superconductor, which propagate in space time as particles. These particles can scatter and can annihilate in pairs, but since their number is conserved modulo 2 they carry a conserved $Z_2$ "charge." We will refer to these particle-like $Z_2$ vortices as "visons." One can define a vison "three current," $j_v$, a field which lives on the links of the dual lattice and takes one of two values, zero or one, which satisfies,

\begin{equation}
(-1)^{j_v} = \prod_i \sigma_{ij} ,
\end{equation}

with the plaquette pierced by the dual link. In the deconfined phase, $T^\sigma$, these vison particles exist as gapped excitations, in addition to the spinons and chargons. In terms of the dual Ising model, $S_{\text{dual}}$, the Ising spins $v_i$ are essentially vison creation operators. With the Ising model being disordered for large $K$, the visons (Ising spins) are gapped. Thus, the distinct gapped excitations in $T^\sigma$ are (i) the chargons, (ii) the spinons and (iii) the visons. An important property of these excitations is the existence of long-ranged "statistical" interactions between them. Specifically, when a chargon (or a spinon) is adiabatically transported around a vison, it acquires a geometrical phase factor of $\pi$ (because the chargon is minimally coupled to the $Z_2$ gauge field). Similarly, a vison picks up a $\pi$ phase factor upon encircling either a chargon or a spinon. Evidently, visons and chargons (or spinons) are "relative semions."

As $K$ is reduced further the gauge theory undergoes a phase transition at $K_c$ into its "area-law" phase. This implies that the energy to separate two spinons or chargons, inserted as "test" charges at spatial separation, $R$, grows linearly with $R$. In this "confined" insulating phase, denoted $T$, free chargons and spinons do not exist in the spectrum. The only allowed particle excitations are those that are "charge neutral," that is, invariant under the $Z_2$ gauge transformation. Any bound state with an even number of chargons plus spinons is neutral. In addition to the electron, this includes any composite built from electrons, such as a Cooper pair or a magnon. In the phase $T$ these electron-like excitations will be gapped. This phase is the familiar "band insulator" with an even number of electrons per unit cell.

Note that with $K < K_c$, the dual Ising model orders $\langle v_i \rangle \neq 0$. This corresponds to a "condensation" of the visons. Remarkably, $Z_2$ vortex condensation leads directly to a "confinement" for the chargons and spinons. To understand confinement directly in terms of the dual Ising model, consider the effect of inserting two static "test" chargons, separated by a distance $R$. Each chargon lives on a (spatial) plaquette of the dual Ising model. Due to the geometrical phase factor between visons and chargons, the presence of a chargon corresponds to a second ("frustrated plaquette") in the dual Ising model, that is, a plaquette with an odd number of negative Ising couplings. To frustrate two plaquettes, it suffices to introduce an interconnecting string of negative Ising bonds. In the ordered phase of the dual Ising model, the energy of this string will clearly be linear in its length, thereby confining the two chargons.

It is worth drawing a very important distinction between the Ising gauge theory considered here, and the gauge theories introduced by Baskaran and Anderson\textsuperscript{7} and generalized and extensively studied by several authors.\textsuperscript{8} In the simplest
version of these theories, the spin itself is effectively fractionalized, decomposed into a bilinear of spinful (complex) fermion operators, rather than splitting the Cooper pair into two chargons as discussed above. These spinful fermion operators, the spinons, are minimally coupled to a compact $U(1)$ gauge field. But in contrast to the $Z_2$ gauge field which exhibits both a confined and deconfined phase, the $U(1)$ theory has only a single phase.\textsuperscript{13} In this phase, point-like monopole excitations in 2+1-dimensional space time always proliferate, and drive spinon confinement.\textsuperscript{26} The electron is, then, ultimately not expected to be fractionalized in these theories.

B. Superconducting phases

We now turn to a description of superconductivity within the $Z_2$ gauge theory. Since the spinons will be gapped into singlets within the superconducting phase, it is legitimate to integrate them out, generating once again a field strength term for the gauge field as in Eq. (87). When the dimensionless chargon “hopping” amplitude, $t_c$, increases and becomes much larger than unity, one expects the chargons to condense $\langle e^{i\phi}\rangle \neq 0$. For large $K$ so that the gauge field is effectively frozen, this chargon condensation transition is simply a 3D classical XY transition. Since the chargon carries electric charge $e$, in this phase the charge $U(1)$ symmetry is broken, and a Meissner effect results. But the chargon also carries $Z_2$ charge, so that the $Z_2$ gauge symmetry is also spontaneously broken. Within a conventional BCS description of superconductivity, the order parameter (the Cooper pair) carries charge $2e$, so one might be tempted to conclude that this “chargon condensate” is perhaps some sort of exotic unconventional superconducting phase. In particular, it is not \textit{a priori} clear that the chargon condensate can support a conventional $hc/2e$ BCS vortex.

To highlight the confusion, it is instructive to focus on the regime with large $K$, where a good description of the ground state can be obtained by setting $\sigma_{ij} = 1$ on every link, and taking the chargon phase $\phi_i$ as a space-time independent constant. Consider placing an $hc/2e$ vortex at the (spatial) origin. Upon encircling this $U(1)$ vortex at a large distance, the phase of the chargon wave function must wind by $\pi$. This is of course not possible with a smoothly varying phase field, but requires the introduction of a “cut” running from the origin. Upon encircling this $U$ gauge field, the order parameter (the Cooper pair) carries charge $2e$, so one might be tempted to conclude that this “chargon condensate” is perhaps some sort of exotic unconventional superconducting phase. In particular, it is not \textit{a priori} clear that the chargon condensate can support a conventional $hc/2e$ BCS vortex. But imagine changing the sign of all the $Z_2$ gauge fields, $\sigma_{ij}$, which “cross” the cut. This corresponds to placing a $Z_2$ vortex at the origin. These sign changes “unfrustrate” the XY couplings across the cut, so that the line tension vanishes. It is thus apparent that a bound state of a $Z_2$ vortex and the $hc/2e$ $U(1)$ vortex (in the phase of the chargon) can exist within the chargon condensate. It is this bound state which corresponds to the elementary BCS vortex in the conventional description of a superconductor.

It is worth emphasizing that both the “naked” $hc/2e$ $U(1)$ vortex and the $Z_2$ vortex, the vison, are confined in the superconducting phase. For example, the energy cost to pull apart two $Z_2$ vortices also grows linearly with separation. To see this, introduce two visons by changing the sign of the $Z_2$ gauge field along an interconnecting “line.” Due to the chargon condensate which breaks the $Z_2$ gauge symmetry making the gauge field “massive,” each negative bond costs an energy $4t_c$, implying linear confinement.

Thus the distinct massive excitations [apart from the Anderson–Higgs plasma mode necessitated by the $U(1)$ symmetry breaking] in the chargon condensate are the spinons and the BCS $hc/2e$ vortices. This is exactly that is required in a conventional superconducting phase. Further, since the spinons are minimally coupled to the $Z_2$ gauge field, there is a long range statistical interaction between the spinons and the BCS vortices. In effect, a spinon “sees” the $Z_2$ vortex, which is bound to the $hc/2e$ vortex, as a source of “Ising flux.” This too is what is required in a conventional superconductor. Thus, the chargon condensate does in fact describe a conventional superconducting phase, denoted hereafter as $\text{SC}$.

A schematic phase diagram is shown in the $K$–$t_c$ plane in Fig. 2. The transition from the fractionalized insulator $\mathcal{I}^\mathbb{Z}$ into $\text{SC}$ is essentially a superconductor–insulator transition for the charge $e$ chargons. These exist as finite energy excitations in $\mathcal{I}^\mathbb{Z}$, superconducting order is obtained if they condense. On the other hand, the transition from the conventional insulator $\mathcal{I}$ into $\text{SC}$ can be viewed as a superconductor–insulator transition for charge $2e$ Cooper pairs. This can be seen by considering the $K=0$ limit, where it is possible to integrate out the $Z_2$ gauge field and arrive at an effective theory of Cooper pair hopping:

$$S_{\text{pair}} = -2t_c \sum_{(ij)} \cos[2(\phi_i - \phi_j)].$$  \hspace{1cm} (91)

IV. ODD NUMBER OF ELECTRONS PER UNIT CELL WITH $d$-WAVE PAIRING

Having explored the physics of electron fractionalization which follows from the $Z_2$ gauge theory in the simplest of cases with an even number of particles per site in the presence of $s$-wave pairing correlations, we turn now to a much more interesting and challenging situation: correlated Mott insulators with one electron per unit cell in the presence of local $d$-wave pairing correlations. As we shall see, in this case the $Z_2$ gauge theory has two simple limiting regimes— one describing a $d$-wave superconductor and the other a conventional antiferromagnetic insulator. But in the interesting crossover regime between these two limits, a number of other phases can be readily described within the $Z_2$ gauge...
theory. Besides a spin-Peierls ordered phase, the theory gives a simple description of the nodal liquid, an exotic fractionalized insulator with gapless fermionic quasiparticles. With one electron per unit cell, confinement transitions out of the $d$-wave superconductor or nodal liquid are inextricably linked to breaking of translational symmetry.

The full theory of interest can be written as

$$ S = S_c + S_s + S_B, $$

$$ S_c = -2t_c \sum_{(ij)} \sigma_{ij} \cos(\phi_i - \phi_j), $$

$$ S_s = -\sum_{(ij)} (t_{ij}^s \hat{f}_i \hat{f}_j + t_{ij}^\Delta \hat{f}_i \hat{f}_{j+} + c.c.) \sum_i \hat{f}_i \hat{f}_j. $$

As shown in Eqs. (70) and (72), with odd integer $N_0$ there is an extra Berry’s phase term in the action.

$$ S_B = -i \frac{\pi}{2} \sum_{i,j} (1 - \sigma_{ij}). $$

It is instructive to consider various limiting cases described by the above action. First consider the limit $t_c = 0$. Then $S_s = 0$, and the $\phi$ fields may be trivially integrated out. Surprisingly, the partition function for the remaining spin sector of the theory is formally equivalent to the Heisenberg antiferromagnetic spin model. To demonstrate this we first trace over the two allowed values of the $Z_2$ gauge field $\sigma_{ij}$ on each link. Consider first the spatial links, which enter the action in the form.

$$ S'_s = \sum_{(rr')} \sum_{\tau} \sigma_{rr'} A^\tau_{rr'}, $$

$$ A^\tau_{rr'} = -t^\tau_{rr'}(\hat{f}_r \hat{f}_{r'}, + c.c.) - t^\Delta_{rr'}(f_r f_{r'} - (\uparrow \rightarrow \downarrow) + c.c.). $$

For notational simplicity we have suppressed the index $\tau$ on the fermion fields. Tracing over the $\sigma_{rr'}$ fields for each (independent) spatial link and exponentiating the result generates a term in the action of the form.

$$ S_s = -\sum_{(rr')} \ln \cosh(A^\tau_{rr'}). $$

Since $A$ is bilinear in the fermion fields, upon expanding in powers of $A$ one generates a series of terms that involve multiples of four spinons.

Now consider the trace of $\sigma_{ij}$ along the temporal links. Recall that the effect of the gauge field $\sigma_{i,j-1;}$ along the temporal links is precisely to impose the constraint Eq. (37) on the Hilbert space in a Hamiltonian formulation. With the $\phi$ fields integrated out, at $t_c = 0$, this constraint reduces to requiring

$$ (-1)^n = -1 $$

at each site of the spatial lattice. Due to Pauli exclusion this is equivalent to the constraint that $n_t = 1$ at each site. Thus, after tracing out the $\sigma$ field, the Hamiltonian obtained from $S_s$ is constrained to operate on a Hilbert space with exactly one spinon per site. This Hamiltonian consists of a sum of terms for each nearest neighbor spatial link. With the additional requirement of spin rotation symmetry, the Hamiltonian must take the form of the Heisenberg spin Hamiltonian.

$$ H = J \sum_{\langle rr' \rangle} S_r, S_{r'}. $$

This can be verified directly from $S_r$ by expanding out the ln cosh term, and re-expressing the spin operators in terms of the spin operators, $S_r = f_r^*, f_r^s$. This leads to an explicit expression for the exchange interaction:

$$ J = \frac{1}{\epsilon} (t^*)^2 + (t^\Delta)^2, $$

where $\epsilon$ is the discrete time slice defined in Eq. (73).

A few comments are in order on this result. It is certainly obvious from the Hamiltonian in Eq. (4) that killing the superconductor at half filling by letting $u \rightarrow \infty$ will lead to antiferromagnetism. Our point here is, however, different. It is interesting (and reassuring) to see this emerge directly from the $Z_2$ gauge theory action itself, especially as some approximation has gone into deriving the $Z_2$ action from the microscopic Hamiltonian [i.e., ignoring the amplitude fluctuations about the saddlepoint, as discussed below Eq. (55)]. Also, this gives us an alternate way of motivating the $Z_2$ gauge theory starting directly with the Heisenberg magnet.

Recovering the Heisenberg antiferromagnet in the limit $t_c \rightarrow 0$ also provides a way to obtain a rough estimate for the saddlepoint parameter $\chi_0$. First, we note that $t^\prime$ and $t^\Delta$ can be re-expressed in terms of the parameters $t$, $u$, $J$, $\Delta$, and $\chi_0$ using Eqs. (78) and (79). Although these relations are strictly valid for $s$-wave pairing, they suffice to give rough estimates even for the $d$-wave case. It is, however, necessary to modify the equation for $t^\Delta$ due to the slightly different decoupling in the $d$-wave case (see Appendix B). Assuming that the saddle point value $\eta_0 \sim \chi_0$, we get

$$ t^\Delta \sim \frac{\Delta}{J^\tau}. $$

Combining Eqs. (78) and (102) with Eq. (101) and assuming $\Delta \ll J$, leads to an estimate for $\chi_0$.

$$ \chi_0 \sim \left( \frac{tu}{J^2} \right)^{1/3}, $$

which is appropriate in the limit of large $u/t$. Having estimated $\chi_0$, one can use Eqs. (77), (78) and (102) to obtain estimates for the three dimensionless coupling constants, $t_c$, $t^\prime$, and $t^\Delta$, respectively. The resulting estimates are given in Eq. (19).

Having established the equivalence of the action in Eq. (92) to the Heisenberg antiferromagnet in the limit $t_c \rightarrow 0$, we briefly consider the opposite large $t_c$ limit. With sufficiently large $t_c$ the chargons will condense and, as argued in Sec. III, this describes a conventional superconducting phase. But due to the assumed form of the pairing correlations, the pairing symmetry here will be $d^\pm_{x^\pm y^\pm}$. Thus, the $Z_2$ gauge theory in Eq. (92) has the remarkable property that it describes a con-
ventional antiferromagnet for small chargon coupling, and a
conventional $d_{x^2-y^2}$ superconductor in the opposite extreme.
We now turn our attention to the exceedingly interesting
regime between these two limits.

A. Correlated Mott insulators

When the chargon coupling strength $t_c$ is small, the chargons
will be gapped out, and the system in an insulating
phase. In this case, it is appropriate to integrate out the chargon
fields to obtain an effective action for the spinons and
the gauge field $\sigma$. The main result of this integration will be
to generate a plaquette product term of the form

$$S_\sigma = -K \sum_{\Box} \prod_{ij} \sigma_{ij}. \quad (104)$$

The full remaining action which is valid within the insulating
phases is then simply

$$S = S_s + S_\sigma + S_B. \quad (105)$$

The parameter $K$ depends on the coupling $t_c$, vanishing at
t $t_c = 0$ and increasing monotonically with $t_c$. The transition
to superconductivity will occur when $t_c = 1$. Near this limit,
but on the insulating side, the value of $K$ will also be of order 1.
Keeping this in mind, we first find it convenient to analyze
the phase diagram of the above action for arbitrary $K$, incorporating
later the superconducting phase.

The action in Eq. (105) has three dimensionless coupling constants, $t'$, $t^2$, and $K$. Considerable progress can be made
in determining the phase diagram by focusing on three dif-
f erent limits. The first, considered above, is $K = 0$ where the
model reduces to the Heisenberg spin model. The second
tractable limit is large $K$. If $K \rightarrow \infty$ the gauge field is frozen
out and it is possible to choose a gauge with $\sigma_{ij} = 1$ on every
link. Then, the only remaining piece of the action describes
noninteracting spinons with a gapless $d$-wave dispersion at
four points in the Brillouin zone. This is the “nodal liquid”
phase obtained in earlier work by vortex pairing within a
dual vortex formulation. The nodal liquid is a fractionalized
insulator with deconfined, gapless spinons and gapped chargons.
For large but finite $K$ and in the absence of $S_B$, the $Z_2$
gauge theory is in its perimeter law phase. As we show below,
this continues to hold even in the presence of $S_B$. In fact,
the region of stability of the perimeter phase is en-
\textit{hanced} by the $S_B$ term. Thus, the chargons and spinons re-
main deconfined and the nodal liquid phase survives for
large but finite $K$.

As with the fractionalized insulator discussed in Sec. III,
the chargons and the spinons there is another distinct excitation in the nodal liquid phase, the $Z_2$ vortex
configuration in the $\sigma$ field, dubbed the vison. The vison is a
gapped excitation in the nodal liquid. As before, due to the
minimal coupling of the chargons and the spinons to the $Z_2$
gauge field $\sigma$, they each acquire a phase of $\pi$ upon encir-
cling a vison. There is thus a long-ranged statistical interaction
between a chargon (or a spinon) and a vison.

The third tractable limit of the action Eq. (105) is small $t'$
and $t^2$. [Estimates appropriate to the cuprates obtained from
Eq. (19) suggest that these couplings will most likely be
much smaller than 1.] In the extreme limit of $t' = t^2 = 0$, we
are left with a pure $Z_2$ gauge theory described by $S_{\text{eff}} = S_\sigma$
$+ S_B$. To explore the effects of the Berry’s phase term $S_B$ on
the gauge theory, it is useful to pass to the dual representa-
tion. Recall that for $S_B = 0$ the dual theory is simply the
$2+1$-dimensional Ising model, with the Ising spin operators
($\sigma_{ij} = \pm 1$) creating the vison excitations. To implement the
duality transformation with the Berry’s phase term present, it is
convenient to first rewrite it as

$$S_B = i \frac{\pi}{4} \sum_{\langle ij \rangle} (1 - \sigma_{ij}) \left(1 - \prod_{\Box} \mu_{ij}^{\text{ext}} \right). \quad (106)$$

Here $\mu_{ij}^{\text{ext}}$ can be viewed as an “external” $Z_2$ gauge field
living on the links of the dual lattice, which satisfies $\prod_{\Box} \mu_{ij}^{\text{ext}} = -1$ through every spatial plaquette. In this form
one can readily generalize the duality transformation in Appen-
dix C to give

$$S_{\text{dual}} = -K_d \sum_{\langle ij \rangle} v_i \mu_{ij}^{\text{ext}} v_j, \quad (107)$$

with dual coupling satisfying: tanh$(K_d) = e^{-2K}$. Due to the
Berry’s phase term, every spatial plaquette (with normals
along the time direction) in the dual Ising model is frus-
trated. In the time continuum limit this becomes a 2D quantum
transverse-field Ising model which is fully frustrated.

The quantum Ising model on a fully frustrated square lat-
tice has been studied extensively by several authors. In partic-
ular, Jalabert and Sachdev studied the model numeri-
}cally (not coincidentally) in the context of frustrated quan-
tum Heisenberg spin models. For small $K_d$ the Ising model
exhibits the usual paramagnetic phase, in which the visons
are gapped (uncondensed) with $\langle v_i \rangle = 0$. This corresponds to the “low temperature” phase of the gauge theory. Deep
within this phase one can set $\sigma_{ij} = 1$ on all the links, which
implies (for $t'$, $t^2 \neq 0$) that the chargons and spinons are de-
confined. This is the nodal liquid phase discussed earlier. It is
noteworthy that the frustration in the Ising model, which is a
direct consequence of being in a Mott insulator with one elec-
tron per site, enhances the stability of the fractionalized
nodal liquid phase (the paramagnetic phase of the Ising
model).

As $K_d$ is increased, it has been found that the Ising
model orders, breaking the global $Z_2$ spin flip symmetry. But
due to the frustration, this ordering is accompanied by a
spontaneous breaking of translational symmetry. It is conve-
nient to characterize this symmetry breaking in terms of the
gauge-invariant energy densities of the near-neighbor bonds:
$E_{ij} = -\langle v_i \mu_{ij}^{\text{ext}} v_j \rangle$. It is found that some of the bonds are
“frustrated” with positive $E_{ij}$, while the remaining are
“happy” with negative bond energies. In the spatially bro-
ken ordered phases, it is found that these frustrated bonds
form lines (see Fig. 3), which run along the principal axis of
the square lattice (columns or rows). There are four favored
configurations, corresponding to frustrated bonds along ev-
ery other column or along every other row. Within each of
these phases, a particular gauge choice can be made with
$\mu_{ij}^{\text{ext}} = -1$ on each frustrated bond. With this choice of
gauge, the Ising spins, $v_i$, exhibit global ferromagnetic or-
K d

the Landau theory of the first reference in Ref. 30. These treated Ising model are possible; some of these are explored in symmetries.

\[ \text{expectation value} \]

various other physical quantities. In particular, the quantum of the energy density will, in general, induce a modulation in Upon including the coupling to the spinons, this modulation \[ \text{s} \]

\[ \text{station invariance remains unbroken, this state corresponds to} \]

expectation value from other bonds. Presuming the spin ro-

\[ \text{s} \]

spatially modulated—bonds which "cross" the frustrated dual Ising model as depicted in Fig. 3.

\[ \text{appropriate gauge choice for} \]

other possibile phases in the present paper. We will not consider these other possible phases in the present paper.

What are the effects of a small nonzero \( t^s \) and \( t^d \) which couple the spinons to the \( Z_2 \) spin flip, translational and rotational symmetries.

In general, several other ordered phases of the fully frustrated Ising model are possible; some of these are explored in the Landau theory of the first reference in Ref. 30. These phases may perhaps be stabilized by very large \( K_d \), and/or longer ranged interactions beyond the simplest nearest neighbor model studied in Ref. 16. We will not consider these other possible phases in the present paper.

What are the effects of a small nonzero \( t^s \) and \( t^d \) which couple the spinons to the \( Z_2 \) gauge field? In the context of quantum antiferromagnets, Sachdev and co-workers have suggested that the spatial ordering of the Ising model corresponds to a spin-Peierls ordering. This interpretation appears to be consistent within our present framework. Specifically, associated with each frustrated bond in the Ising model, is a corresponding frustrated plaquette on the dual lattice "pierced" by that bond. The expectation value of the plaquette product in the gauge theory will therefore be modulated in these ordered phases, with \( \langle \Pi \sigma \rangle \approx -\xi_{ij} \). Upon including the coupling to the spinons, this modulation of the energy density will, in general, induce a modulation in various other physical quantities. In particular, the quantum expectation value \( \langle S_i S_j \rangle \) evaluated for each bond will be spatially modulated—bonds which "cross" the frustrated lines of the dual lattice will have a different value for this expectation value from other bonds. Presuming the spin rotation invariance remains unbroken, this state corresponds to a spin-Peierls phase which we denote as SP. The "singlet bonds" in this phase are arranged in a columnar fashion, running perpendicular to the lines of frustrated bonds in the dual Ising model as depicted in Fig. 3.

Since the Ising spins in the fully frustrated Ising model order ferromagnetically in these modulated phases (with an appropriate gauge choice for \( \mu^{st} \)) implying a vison condensation, \( \langle v_i \rangle \neq 0 \), confinement is expected. To see this, consider evaluating the Wilson loop correlator defined in Eq.

\[ \text{FIG. 4. Schematic zero temperature phase diagram showing one possible scenario for the evolution from the antiferromagnet (AF) to the } \]

\[ d \text{-wave superconductor } dSC. \text{ In this scenario, all the insulating phases are conventional. The thick lines indicate confinement of the chargons and spinons. For concreteness, we have chosen to display a particular sequence of confined phases, namely, a transition from AF to a spin-Peierls (SP) insulator, and a further transition to } dSC. \]

\[ (89). \text{ In the dual frustrated Ising model, this corresponds to changing the sign of all the Ising couplings on bonds which pierce through the loop. Being ferromagnetically ordered, this will cost an energy (action) proportional to the area of the loop, the signature of confinement. Thus, as expected, the spin-Peierls state is a conventional insulator, with confined spinons and chargons. The gapped spin-one excitations made by breaking the singlet bonds can then be thought of as a (confined) pair of spinons.} \]

The three limiting cases discussed above suggest the phase diagram shown in Fig. 1 for the action in Eq. (105). Consider first the regime with small \( t^s \) and \( t^d \). At very small \( K \) a conventional antiferromagnetic insulator is expected. With increasing \( K \) there is presumably a phase transition into a conventional spin-Peierls insulator with confined chargons and spinons. Upon further increasing \( K \), the spin-Peierls phase undergoes a deconfinement transition into the fractionalized nodal liquid phase. For large \( t^s \) and \( t^d \), the antiferromagnet and nodal liquid phases will still be present in the limit of very small and large \( K \), respectively. But it is not clear which phases will be present when all three of the coupling constants are of order 1. In particular, it is unclear whether it is possible to have a direct second order phase transition from the antiferromagnet into the nodal liquid or whether there will always be an intervening (spin-Peierls) phase.

We now discuss the implications of these results for the phase diagram of the full \( Z_2 \) gauge theory in which the charge degrees of freedom are present and superconductivity is possible. Of primary interest is the evolution from the antiferromagnet to the \( d \)-wave superconductor upon increasing the chargon coupling, \( t_c \). A transition into the superconductor is expected to occur at some critical chargon coupling, \( t^c_s \), of order one. For smaller \( t_c \) in the insulating regime, the dimensionless coupling \( K \) will at most be of order 1. One can imagine two qualitatively distinct possibilities upon tuning towards the superconductor from the insulating phases. First, it may be that even when \( t_c \) increases to \( t^c_s \), the value of \( K \) will remain smaller than the critical value needed for deconfinement, \( K_c \). In this case, all the intermediate phases between the antiferromagnet and the superconductor will be conventional confined phases. This is illustrated in Fig. 4. Alternately, it may be that \( K \) exceeds \( K_c \) before the onset of superconductivity. This would imply the existence of the deconfined nodal liquid phase intervening between the \( d \)-wave superconductor and a conventional insulator. This is illustrated in Fig. 5.

Which one of these two possibilities is realized will presumably depend sensitively on microscopic details. Indeed, since \( K \) is of order 1 when \( t_c \) approaches \( t^c_s \), it seems likely that the onset of superconductivity will occur close to the
boundary between the confined and deconfined insulating phases. But in any event, our analysis has firmly established the possibility of the deconfined nodal liquid phase. It remains a challenge to determine whether this exotic fractionalized insulator is realized in the cuprates.

In Sec. V, we will describe much of the physics discussed here in a dual formalism in terms of vortices rather than the chargons. This will provide considerable further insight, and make connections with earlier approaches.

V. DUAL VORTEX REPRESENTATION

For a system of interacting bosons in two spatial dimensions, it is well known that the insulating phases can be described as a condensate of vortices in the many particle wave function. More formally, it is possible to set up a dual description where the vortices, rather than the particles, are the fundamental degrees of freedom such that the insulating phase is a vortex condensate while the superfluid phase is the vortex vacuum. For the electronic systems considered in this paper, it is natural to attempt to do the same, and work with the vortex vacuum. For the electronic systems considered in this paper, it is natural to attempt to do the same, and work with the vortex vacuum. For the electronic systems considered in this paper, it is natural to attempt to do the same, and work with the vortex vacuum.

We will start with the full chargon–spinon action $S = S_v + S_f + S_{CS}$ discussed in Sec. IV, and perform a duality transformation to trade the chargon fields for the $hc/2e$ vortices. This differs somewhat from the conventional duality transformation\textsuperscript{14} from bosons to vortices due to the coupling of the chargons to the $Z_2$ gauge field.

To understand how to deal with the chargon coupling to the $\sigma$ field, it is useful to first review the well-known self-duality of the $Z_2$ gauge theory with Ising matter fields in 2 + 1 dimensions. This is done in detail in Appendix C. The duality proceeds by first rewriting the partition function in terms of a $Z_2$ current for the Ising matter fields and the $Z_2$ gauge field, $\sigma_{ij}$. The $Z_2$ current lives on the links of the lattice and can take one of two values, 0 or 1. It is conserved modulo 2 at each site of the lattice. This conservation law can be implemented by writing the $Z_2$ current as the flux of a dual $Z_2$ gauge field, denoted as $\mu_{ij}$. (This is completely analogous to the duality of the three-dimensional classical $XY$ model.) Eliminating the $Z_2$ current in favor of the dual gauge field gives an action written entirely in terms of two $Z_2$ gauge fields ($\sigma_{ij}$ and $\mu_{ij}$) which are duals of each other. The original $Z_2$ gauge field, $\sigma_{ij}$ may be eliminated by expressing its flux as the current of a dual Ising matter field, the vison $v_i$. The resulting theory has the same form as the original $Z_2$ gauge theory with matter fields, but is dual to it.

To obtain a dual representation of the system of chargons and spinons coupled to the $Z_2$ gauge field $\sigma_{ij}$, we need to combine the dual representation of the $Z_2$ gauge theory with the standard duality transformation of the $XY$ model. As shown in detail in Appendix D, this is readily done. For the time being, we will only consider the situation with local $d$-wave pairing and an odd number of electrons per unit cell. The result is a lattice action in terms of $hc/2e$ vortices, which are minimally coupled to a fluctuating $U(1)$ gauge field $a$ whose circulation is the total electrical current. In addition, the $hc/2e$ vortices are minimally coupled to a $Z_2$ gauge field $\mu_{ij}$. The full action is given by

$$S = S_v + S_f + S_{CS} + S_B,$$

$$S_v = - t_v \sum_{\langle ij \rangle} \mu_{ij} \cos \left( \theta_i - \theta_j + \frac{a_{ij}}{2} \right),$$

$$S_a = \frac{\kappa}{8\pi^2} \sum \left( \Delta \times a_{ij} \right)^2,$$

$$S_f = - \sum_{\langle ij \rangle} \sigma_{ij} [t_{ij}^A \bar{f}_j + t_{ij}^B f_i \bar{f}_j] - \sum_i \bar{f}_i f_i,$$

$$S_{CS} = - i \pi \frac{\theta_i}{4} \left( 1 - \prod_j \sigma_j \right),$$

Here $e^{i\theta_i}$ creates the $hc/2e$ vortex, and $f_i$ is the spinon as before. The first term represents single vortex hopping, while the second is a kinetic term for the $U(1)$ gauge field $a_{ij}$. The flux of $a$ is the total electrical current, in particular a flux of $2\pi$ through a spatial plaquette adds an electric charge of 1 to a chargon. Together these two terms comprise the usual dual vortex representation of a set of charge $2e$ Cooper pairs, except that here the vortices are minimally coupled to an additional $Z_2$ gauge field $\mu_{ij}$. This leads to a vortex–spinon coupling mediated by $S_{CS}$. This term has a structure very similar to a Chern-Simons term (although it is for the group $Z_2$) and, as discussed below, plays a similar role. The Berry’s phase term $S_B$ is the same as before.

The full dual action is invariant under a local $U(1)$ gauge transformation,

$$\theta_i \rightarrow \theta_i + \Lambda_i, \quad \mu_{ij} \rightarrow \frac{\Lambda_i - \Lambda_j}{2}.$$

This is standard in the dual vortex description of $XY$ models in three dimensions. The corresponding conserved charge is the vorticity. The action has an additional $Z_2$ gauge symmetry under which

$$e^{i\theta_i} \rightarrow e^{i\epsilon_i \theta_i}, \quad \mu_{ij} \rightarrow e^{i\epsilon_i \mu_{ij}} e_j,$$

with $\epsilon_i = \pm 1$. We emphasize that this gauge symmetry is distinct from the local $Z_2$ gauge symmetry of the spinon–chargon action, but in fact is dual to it.
To get some intuition about the term $S_{CS}$, it is instructive to replace the vortex hopping term in the action by a Villain potential,

$$e^{i\phi} \cos \theta_{ij} \rightarrow \sum_{J_{ij}=-\infty}^{\infty} e^{-J_{ij}^2/2 e} e^{iJ_{ij} \theta_{ij}},$$

where $\theta_{ij} = \theta_{ij} + (a_{ij}/2) + (\pi/2)(1 - \mu_{ij})$ is the gauge invariant phase difference. Here the integer field $J_{ij}$ that lives on the links of the lattice represents the three current of the $hc/2e$ vortices. After this replacement it is possible to explicitly perform the summation over the gauge field $\mu_{ij}$. For each link of the lattice this contributes a term to the partition function of the form, $1 + (-1)^J \Pi \sigma$, which vanishes unless

$$(-1)^J = \prod \sigma.$$

Thus, the Chern–Simons term has effectively attached a $Z_2$ flux of the gauge field $\sigma$—a vison—to each $hc/2e$ vortex. As discussed in Sec. III, this composite comprised of an $hc/2e$ vortex bound to the $Z_2$ vison is nothing but the familiar BCS vortex. Due to the attached vison, when a spinon is taken around the BCS vortex it acquires the expected $\pi$ phase factor.

Alternatively, it is possible to perform an ‘integration by parts’ on $S_{CS}$ which effectively exchanges the role of $\sigma$ and $\mu$, and then perform a summation over $\sigma$. This leads to the additional constraint,

$$(-1)^J = \prod \mu,$$

with $J_i$ the spinon three current. A $Z_2$ flux in the gauge field $\mu$ has thereby been attached to each spinon. More precisely, since the spinon number is only conserved modulo 2 due to the anomalous pairing term, the $Z_2$ flux is attached whenever an odd number of spinons propagates. The net effect of this $Z_2$ Chern–Simons term is to implement mathematically the long-ranged statistical interaction between BCS vortices and spinons. This kind of flux attachment may be familiar to many readers for the $U(1)$ group from theories of the quantum Hall effect. But since the spinon number itself is not conserved, implementing this statistical interaction with a $U(1)$ Chern–Simons term is problematic. It is a remarkable aspect of the duality transformation in Appendix D, that this Ising-like Chern–Simons terms emerges so naturally.

A. Phases

We now analyze the phases in this dual vortex description, focusing on the most interesting case of an odd number of electrons per site with local $d$-wave pairing correlations. In the vortex description the superconducting phase corresponds to a vortex vacuum, and the insulating phases are vortex condensates. We consider first two simple limiting cases, first the superconductor with vanishingly small vortex hopping $t_v \rightarrow 0$, and then the insulator with $t_v \rightarrow \infty$.

When $t_v$ is zero the summation over the gauge field $\mu$ can be performed, giving the constraint $\Pi \sigma = 1$. It is then possible to pick a gauge with $\sigma_{ij} = 1$ on every link. The resulting action has two pieces, $S_a$ which describes the gapless sound mode of the superconductor (gapped when long-ranged Coulomb interactions are included) and the spinon piece $S_s$. With $\sigma_{ij} = 1$ the spinons can freely propagate and describe the gapless nodal quasiparticles. A correct description of a conventional $d$-wave superconductor is thereby recovered.

Consider next the opposite limit with $t_v \rightarrow \infty$. In this regime the $hc/2e$ vortices will condense, $\langle e^{i\theta} \rangle \neq 0$. The dual Anderson–Higgs mechanism leads to a mass term for the gauge field $a_{ij}$, indicative of a charge gap. With one electron per unit cell the resulting phase is thus a Mott insulator. In the absence of any gapped charge excitations ($\Delta = \Delta = 0$), it is possible to choose a gauge with $a_{ij} = 0$ on every link. The vortex hopping term becomes $S_v = -h \Sigma_{ij} \mu_{ij}$ with a nonzero ‘field;’ $h = t_v \langle e^{i\theta} \rangle^2$. When this field is large one can set $\mu_{ij} = 1$ on each link, so that the Chern–Simons terms vanishes. The full action then reduces to $S_{sf} = S_s + S_B$. At this stage the summation over the $\sigma$ gauge field can be performed explicitly. As detailed in Sec. IV, the resulting model reduces to a simple 2D near-neighbor Heisenberg antiferromagnet. Thus, we readily recover the simple antiferromagnet from the dual representation by condensing $hc/2e$ vortices.

Finally, here in Sec. V we wish to recover a dual description of the fractionalized nodal liquid. Since the nodal liquid is electrically insulating it requires vortex condensation. But as established in Sec. IV, the nodal liquid supports gapped $Z_2$ vortices, the vison excitations. Since the Chern–Simons term attaches a vison to each $hc/2e$ vortex, it is clear that to obtain the nodal liquid the $hc/2e$ vortices cannot be condensed. But since the square of the vison operator is unity ($v^2 = 1$), a pair of $hc/2e$ BCS vortices does not carry a vison with it. As we now show, the nodal liquid can be obtained from the $d$-wave superconductor by pairing BCS vortices, and then condensing the $hc/e$ vortex composite.

To this end, we add an extra vortex pair hopping term to the action,

$$S_{2v} = -t_{2v} \sum_{(ij)} \cos(\theta_{ij} - \theta_{2i} + a_{ij}).$$

Here, $e^{i\theta_{2i}} = (e^{i\theta})^2$, thus creating a pair of BCS vortices. Notice that the $hc/e$ vortex is also minimally coupled to the $U(1)$ gauge field, as required by the dual $U(1)$ symmetry of the action, but is not coupled to the $Z_2$ gauge field $\mu_{ij}$, because it carries no vison charge. We now consider taking $t_{2v}$ large and condensing the $hc/e$ vortex, $\langle e^{i\theta_{2i}} \rangle \neq 0$, keeping the $hc/2e$ vortex uncondensed. Before doing this it is convenient to re-express the $hc/2e$ vortex as

$$e^{i\theta_i} = v_i e^{i\theta_{2i}/2},$$

with $v_i = \pm 1$ the vison operator. Notice that with this identification the field $\theta_i$ can be treated as an angular variable, since the right side is invariant under the combined transformation, $\theta_i \rightarrow \theta_i + 2\pi$ and $v_i \rightarrow -v_i$. We finally find it convenient to absorb the field $\theta_{2i}$ into the gauge field $a_{ij}$ by the gauge transformation,

$$a_{ij} \rightarrow a_{ij} + \theta_{2i} - \theta_{ij}.$$  (121)

In this gauge, the vortex hopping terms become
\[ S_v = -t_v \sum_{ij} v_i \mu_{ij} v_j \cos \left( \frac{a_{ij}}{2} \right), \]  
\[ S_{2v} = -t_{2v} \sum_{ij} \cos(a_{ij}). \]  

In the insulating phase with large \( t_v \), there will again be a charge gap due to the dual Anderson–Higgs mechanism coming from the \( hc/e \) vortex condensate. Above the gap will be charge \( e \) chargons, corresponding to a \( 2\pi \) flux tube in \( a_{ij} \). In the absence of any charged excitations one can set \( a_{ij} = 0 \), and the single vortex hopping term becomes

\[ S_v = -t_v \sum_{(ij)} v_i \mu_{ij} v_j. \]  

The full effective action is \( S_{\text{eff}} = S_v + S_s + S_C + S_B \). When \( t_v \) is small the visons will be uncondensed \( \langle v_i \rangle = 0 \). In this limit the summation over the \( \mu \) gauge field can be performed, and due to the Chern–Simons term, this leads to the constraint, \( \Pi \sigma = 1 \). One can then choose a gauge with \( \sigma_{ij} = 1 \) on each link, which sets \( S_B = 0 \). The only remaining term in \( S_{\text{eff}} \) describes free propagating spinons. These are the gapless nodons in the insulating nodal liquid.

We thereby recover a decimation of the nodal liquid from the dual vortex formulation. In addition to the gapless nodons, the nodal liquid supports two gapped excitations, the chargon and the vison. As it is clear from the above analysis, the vison is a remnant of the \( hc/2e \) BCS vortex which survives into the nodal liquid upon condensation of the \( hc/e \) vortex pair. Physically, since the vorticity is only conserved \( \bmod 2 \), one of the two \( Z_2 \) remains from the \( hc/2e \) BCS vortex. As before, the vison picks up a \( \pi \) phase change when it is transported around either a spinon or a chargon. To see this, note that a chargon corresponds to a \( \pi \) flux in \( a_{ij}/2 \) and the nodon (spinon) a \( \pi \) flux in \( \mu_{ij} \). As seen in Eq. (122), the vison is minimally coupled to both of these gauge fields, thus acquiring a sign change upon encircling the spinon or chargon.

It is worth emphasizing that a clear mechanism for vortex pairing can be found from the analysis in Sec. IV. Since the chargons and visons (or vortices) have a long-ranged statistical interaction, motion of the charge is greatly impeded by the presence of unpaired visons. On the other hand, once the \( hc/2e \) vortices are paired, the charge can move coherently. Thus, the presence of a large kinetic energy makes vortex pairing energetically favorable.

It is finally worth mentioning that in the limit \( S_v = 0 \) one readily recovers the fully frustrated Ising model considered in Sec. IV. To see this, note first that \( S_B \) can be rewritten in the form of a Chern–Simons term with \( \mu \) replaced by \( \mu^{\text{ext}} \), where \( \Pi \sigma^{\text{ext}} = 1 \) through all spatial plaquettes. With \( S_v = 0 \), one can then perform the summation over the \( \sigma \) gauge field, and this sets \( \mu_{ij} = \mu^{\text{ext}}_{ij} \). The remaining term in \( S_{\text{eff}} \) is the fully frustrated Ising model.

\[ S_v = -t_v \sum_{(ij)} \mu_{ij} v_i v_j. \]  

VI. DOPING

Our analysis has so far focused only on situations with an integer number, \( N_0 \), of electrons per unit cell. Finite doping leading to noninteger \( N_0 \) does not crucially modify our discussion of fractionalization issues. Indeed, both confined and fractionalized insulating phases can exist for nonzero doping. At a qualitative level, in both kinds of insulating phases, the main effect of noninteger \( N_0 \) will be to induce charge order, accompanied by translational symmetry breaking. The precise nature of this charge order presumably depends on the details of the system, and may be sensitive to the presence of long-ranged Coulomb interactions.

Formally, noninteger values of \( N_0 \) can be incorporated into either the particle or vortex representations as follows. In the particle representation, as discussed in Sec. II, the main effect of noninteger \( N_0 \) is to modify the Berry phase term to

\[ S_B = -i \sum_{i,j = -1}^{N} N_0 \left( 2\pi l_{ij} - \frac{\pi}{2} (1 - \sigma_{ij}) \right). \]  

Here, \( l_{ij} \) is an integer defined on each temporal link given by

\[ l_{ij} = \text{Int} \left[ \frac{\Phi_{ij} + 1}{2\pi} \right], \]  

where \( \Phi_{ij} = \phi_i - \phi_j + (\pi/2)(1 - \sigma_{ij}) \) is the gauge-invariant phase difference between two sites. When \( N_0 \) is not an integer, this Berry phase term leads to complex Boltzmann weights in the partition function sum. This is not too surprising: even in the absence of any gauge field coupling, the partition function for simple Boson–Hubbard models at arbitrary chemical potential involves complex weights.

The presence of such complex weights does not pose a problem for the existence of the fractionalized insulator. We recall that the fractionalized phase is obtained when the gauge field \( \sigma_{ij} \) is in its perimeter phase. Deep in this phase, we may set \( \sigma_{ij} \approx 1 \) on each space-time link so that the Berry phase term \( S_B \) becomes independent of \( \sigma_{ij} \). The resulting action then describes a lattice model of bosonic chargons at filling \( N_0 \) and the fermionic spinons, decoupled from one another. Thus the chargons and spinons will still be decoupled. However, the ground state will generally exhibit charge ordering accompanied by broken translational invariance. Confined conventional insulating phases at noninteger \( N_0 \) clearly also exist.

Numerical simulations of the \( Z_2 \) gauge theory at arbitrary \( N_0 \) to determine the precise nature of the charge ordering in these insulating phases will be seriously hampered by the presence of these complex weights in the partition function. Fortunately, in the dual vortex representation, noninteger \( N_0 \) enters in a more innocuous manner. To generalize the duality transformation to arbitrary \( N_0 \) is straightforward, because the Villain representation of the chargon hopping term in Eq. (D4) is simply modified to read

\[ \frac{K}{2} \sum_{(ij)} \left( J_{ij} - 2\pi N_{ij} \right)^2. \]  

Here, \( N_{ij} = N_0 \) for temporal links, and is zero otherwise. Proceeding with the duality transformation gives the action,
\[ S = S_v + S_z + S_{CS} + S_a, \]  

where the first three terms are the same as before in Eq. (108). The last term, which was equal to \( S_a + S_B \) for integer \( N_0 \), becomes instead

\[ \tilde{S}_a = \frac{\kappa}{8\pi^2} \sum_{\langle ij \rangle} (\Delta \times a_{ij} - 2\pi N_{ij})^2. \]

Notice that in this dual representation, \( N_0 \) acts like an external “magnetic field” piercing each spatial plaquette.

For the particular case of odd integer \( N_0 \), it is instructive to see how the term \( S_B \) may be recovered. To that end, we define a new “external” gauge field \( a^{\text{ext}} \) on the links of the dual lattice such that

\[ \Delta \times a^{\text{ext}}_{ij} = 2\pi N_{ij}. \]

We now absorb \( a^{\text{ext}} \) into \( a \) by the shift \( a \rightarrow a - a^{\text{ext}} \). This eliminates \( a^{\text{ext}} \) so that \( \tilde{S}_a = S_a \), but modifies the vortex hopping term which becomes

\[ S_v = -t_v \sum_{\langle ij \rangle} \mu_{ij} \cos \left( \theta_i - \theta_j - \frac{a_{ij} + a^{\text{ext}}_{ij}}{2} \right). \]

For odd integer \( N_0 \) (say, \( N_0 = 1 \)) one may choose

\[ a^{\text{ext}}_{ij} = 2\pi n_{ij}, \]

with integer \( n_{ij} \), which satisfies \( \Delta \times n_{ij} = N_0 = 1 \) for every spatial plaquette and is zero for all other plaquettes. With this choice we may write

\[ S_v = -t_v \sum_{\langle ij \rangle} \mu_{ij} \mu^{\text{ext}}_{ij} \cos \left( \theta_i - \theta_j + \frac{a_{ij}}{2} \right), \]

where \( \mu^{\text{ext}}_{ij} = (-1)^{n_{ij}} \). Notice that the flux \( \Pi_{ij} \mu \) is \(-1\) for every spatial plaquette and zero for other plaquettes. If we now perform the shift, \( \mu \rightarrow \mu \mu^{\text{ext}} \), the field \( \mu^{\text{ext}} \) is eliminated from \( S_v \) but reappears in \( S_{CS}(\mu \mu^{\text{ext}}) \). But upon noting the form of the Berry’s phase term in Eq. (106), one can easily demonstrate that \( S_{CS}(\mu \mu^{\text{ext}}) = S_{CS}(\mu) + S_B \). We thereby recover the earlier Berry’s phase form for the case with odd integer \( N_0 \).

The dual representation for arbitrary \( N_0 \) is simpler looking than the one in the particle formulation, and is probably better suited to discuss issues such as the nature of charge ordering at finite doping. In particular, if we ignore the coupling to the spinons and set \( \Pi_{ij} \mu = 1 \), the remaining partition function sum involves only real weights, and can presumably be evaluated numerically.

**VII. OTHER EXOTIC FRACTIONALIZED PHASES**

Here in Sec. VII we will briefly explore the possibility of obtaining other fractionalized phases different from the ones discussed so far. The most interesting phase that emerges is a novel fractionalized superconductor; we will describe its properties in both the particle and vortex formulations.

![Image](https://via.placeholder.com/150)

**FIG. 6.** Schematic zero temperature phase diagram for the \( Z_2 \) gauge theory coupled to matter fields described by the action of Eq. (137).

**A. Particle description**

We have already argued that when the charge \( e \) chargons condense, the resulting phase is a conventional superconductor. This is perhaps surprising, since in a conventional BCS description the order parameter carries charge \( 2e \). One might ask whether it is possible to have a superconducting phase in which the chargon pairs (i.e., the Cooper pairs) have condensed, while single chargons have not. As we now demonstrate, such a superconducting phase, which we denote \( SC^a \), can exist and has a surprisingly simple description in terms of our \( Z_2 \) gauge theory. For simplicity, we will initially present the discussion for \( s \)-wave pairing with an even number of electrons per unit cell.

The appropriate action from Eqs. (84) and (87) in Sec. III takes the form \( S = S_c + S_s + S_K \). As discussed there, the kinetic term for the gauge field \( S_c \), although not present in the original action, will in any case be generated upon integrating out high-energy modes. To access the chargon pair condensate phase, it is extremely convenient to add an explicit pair hopping term to the action, \( S_{\text{pair}} \), from Eq. (91). For large pair-hopping amplitude, \( t_2 \), the chargon pairs will condense, leaving the single chargons uncondensed,

\[ \langle e^{i2\phi_i} \rangle \neq 0, \quad \langle e^{i\phi_i} \rangle = 0. \]

This still breaks the global \( U(1) \) charge symmetry, and so describes a superconductor, but one with rather exotic properties. To examine this phase it suffices to take \( t_2 \rightarrow \infty \) which allows one to set \( 2\phi_i = \pi \) times an integer or, equivalently,

\[ \phi_i = \frac{\pi}{2}(1 - s_i), \]

with the value of the Ising spins, \( s_i = \pm 1 \), arbitrary. In this limit, the chargon creation operator equals the Ising spin, \( e^{i\phi_i} = s_i \). After integrating out the massive spinons, this leaves an effective theory of the form

\[ S_{\text{gauge}} = -2t_c \sum_{\langle ij \rangle} s_i \sigma_{ij} s_j - K \sum_{\langle ij \rangle} \prod \sigma_{ij}, \]

with \( t_c \) the chargon “hopping” strength.

This theory, which describes Ising spins “minimally coupled” to a \( Z_2 \) gauge field, has been extensively studied by Fradkin and Shenker as a toy model of confinement. The phase diagram in the \( t_c-K \) plane is shown in Fig. 6. In the \( K \rightarrow \infty \) limit the model reduces to a global Ising model for
the spins. With increasing $t_c$, there is an Ising transition into a phase with $(s_i) \neq 0$ (the Higgs phase), which corresponds to the chargon-condensed SC phase. Along the $t_c = 0$ axis the pure $Z_2$ gauge field exhibits a confinement transition with decreasing $K$. Fradkin and Shenker argued that the Higgs and confined phases could be continuously connected by noting the absence of a phase transition along the $t_c = \infty$ and $K = 0$ lines. Moreover, as detailed in Appendix C, this model is in fact self-dual, and maps into an equivalent model with new parameters reflected across the dashed line.

The phase with large $K$ but small $t_c$ corresponds to the exotic new superconducting phase, $SC^*$. In this phase there are four deconfined massive excitations: (i) the spinon, (ii) an $hc/2e$ $U(1)$ vortex, (iii) the Ising spin $s_i$, and (iv) the $Z_2$ vortex in the gauge field $\sigma$, the vison. In striking contrast to a conventional superconducting phase, in $SC^*$ the $U(1)$ and $Z_2$ vortices can exist as separate excitations, and are not confined to one another. In order to distinguish this $hc/2e$ vortex from the BCS vortex, we will refer to it as an $hc/2e$ vorton. The Ising spin excitation $s$ is a remnant of the chargon. In the paired-chargon condensate $SC^*$ phase, the global $U(1)$ charge symmetry is not broken; there is an unbroken $Z_2$ “charge” symmetry $(s_i \rightarrow -s_i)$ corresponding to an invariance under a sign change of the chargon operator. Although the electrical $U(1)$ charge of the chargon is not conserved, the chargon number is conserved modulo 2, a reflection of this unbroken Ising symmetry. Indeed, one can define a conserved Ising charge as $Q = (1)^N = \pm 1$, where $N$ is the chargon number operator. Since the Ising spin operator changes the sign of $Q$, the massive spin excitation carries the conserved $Z_2$ electrical charge of the chargon. We refer to this excitation as an “ison.”

To gain some physical insight into this strange ison particle, consider what happens when an electron is added to a superconductor. The electron creation operator can be decomposed into the product of a spinon and a chargon,

$$c_i^\dagger = b_i^\dagger f_i^\dagger = s_i f_i^\dagger.$$  \hspace{1cm} (138)

The second equality is valid within the two superconducting phases. In the conventional superconductor $SC$, the ison is also condensed, $(s_i) \neq 0$, so that the electron is essentially equal to the spinon. Thus the spin of the added electron is carried away by the spinon, the conventional BCS quasiparticle, whereas the electrical charge is carried by the condensate. On the other hand, in $SC^*$, adding an electron not only increases the conserved spin by 1/2, but changes the conserved $Z_2$ “electrical charge.” The spin and $Z_2$ charge are carried away by two separate massive excitations—the spinon and ison. Thus, the $SC^*$ phase exhibits an exotic form of spin-charge separation.

It is again important to ask about geometric phase factors acquired when any of the four massive excitations in $SC^*$ encircle another. First, note that both the ison and the spinon are minimally coupled to the gauge field $\sigma$. Consequently, they both acquire a phase factor of $\pi$ on encircling the $Z_2$ vortex, namely, the vison. The ison, being a remnant of a chargon, also acquires a phase of $\pi$ on encircling an $hc/2e$ vorton. Thus the pairs, (spinon, ison), (vortex, ison), and (ison, $hc/2e$ vorton) acquire phase factors of $\pi$ upon encircling one another. Equivalently, there are long-ranged statistical interactions between any two members of a pair. All other pairs of excitations do not acquire any geometrical phase factors. Note in particular that the $hc/2e$ vorton, being unbound from the $Z_2$ vison, does not have a long-range statistical interaction with the spinon in $SC^*$. This distinguishing feature will have several important consequences in the dual vortex description developed in the next section.

The transition from $SC^*$ to $SC$ occurs on condensing the ison so that single chargons are themselves condensed. Note that ison condensation leads to confinement of the excitations; it has long-ranged statistical interactions with the $hc/2e$ vorton and the vison (i.e., the $Z_2$ vortex). The result is the BCS $hc/2e$ vortex, as discussed earlier in Sec. III.

The transition from $I^*$ into $SC^*$ upon increasing $t_c$ can be understood as a superconductor–insulator transition of charge $2e$ chargon (or Cooper) pairs. Note that a direct transition from the conventional insulator $I$ to $SC^*$ is not generically possible.

Figure 7 is a schematic phase diagram exhibiting the four phases, $SC$, $SC^*$, $I$, and $I^*$, as well as the intervening transitions. Of the four, it is only in the band insulator $I$ that spinons are confined. In the other three phases the $Z_2$ vortex is gapped out and uncondensed. These three phases exhibit excitations with “fractionalized” quantum numbers. It is the condensation of the $Z_2$ vortex which leads to confinement, leaving only the electron in the spectrum.

1. Odd number of electrons per unit cell

We now briefly consider the superconducting phases with odd integer filling, but still presuming $s$-wave pairing. Since chargon pairs are condensed in both $SC$ and $SC^*$, it suffices again to consider very large pair hopping amplitude, $t_2$. Moreover, with condensed chargon pairs, the chargon operator can be replaced by the Ising spin, $b_i^\dagger = s_i = \pm 1$, the ison, as discussed above. After integrating out the gapped spinons, the effective theory again reduces to the Ising matter-plus-gauge theory as in Eq. (137), but with the addition of the Berry’s phase term, $S_B$.

$$S_{el} = -2t_s \sum_{ij} s_i \sigma_{ij} s_j - K \sum \prod \sigma_{ij} + S_B[\sigma_{ij}].$$  \hspace{1cm} (139)

Note that the $SC^*$ phase is realized only for large $K$, as discussed above. In this limit, as we have emphasized several times, the effects of the Berry phase term $S_B$ are expected to be innocuous. Thus, $SC^*$ will continue to exist even in the presence of $S_B$. To see this in more detail, it is once again illuminating to pass to a dual representation, which exchanges the isons for the visons.
The discussion above generalizes readily to the case of $d$-wave pairing. In particular, a $d\text{SC}^\ast$ phase where chargon pairs, but not single chargons, have condensed is an allowed phase in the model. Its properties are the same as those for the $s$-wave case above, except that the spinons have a gapless $d$-wave dispersion. Also possible is a $d\text{SC}$ phase coexisting with spin-Peierls order, just as in the $s$-wave case.

2. $d$-wave pairing and doping

The discussion above readily to the case of $d$-wave pairing. In particular, a $d\text{SC}^\ast$ phase where chargon pairs, but not single chargons, have condensed is an allowed phase in the model. Its properties are the same as those for the $s$-wave case above, except that the spinons have a gapless $d$-wave dispersion. Also possible is a $d\text{SC}$ phase coexisting with spin-Peierls order, just as in the $s$-wave case.

In the presence of finite doping with noninteger $N_0$, in either the $s$-wave or the $d$-wave case, the $\text{SC}^\ast$ phase is expected to survive, since the $S_4$ term is innocuous in this phase. The conventional superconducting phases will be more sensitive to the value of $N_0$: several additional superconducting phases with broken lattice symmetries are presumably possible.

B. Vortex description

We now show how the superconductor $\text{SC}^\ast$ may be described in the dual vortex formulation. The discussion in Sec. V was based on the action in Eq. (108) for the spinons and $hc/2e$ vortices. The symmetries of the action allow the addition of “kinetic” terms for both $Z_2$ gauge fields $\sigma$ and $\mu$. Once again, although not present in the original action, these terms will be generated upon integrating out high energy modes,

$$S_\sigma = -K_\sigma \sum_i \sigma_{ij},$$

$$S_\mu = -K_\mu \sum_i \mu_{ij}.$$  

It is of interest to explore the phase diagram for arbitrary positive values of the couplings $K_\sigma$ and $K_\mu$. We will show that the superconductor $\text{SC}^\ast$ emerges quite naturally for large $K_\sigma$ and $K_\mu$. As shown below, an important physical consequence of the addition of these $K_\sigma$ and $K_\mu$ terms is that the Chern–Simons term $S_{CS}$ is no longer effective in attaching flux to the vortices and the spinons. Note that, in the absence of flux attachment, the field $e^{i\phi}$ creates a “naked” $hc/2e$ vortex, i.e., an $hc/2e$ vorton. Attaching flux of the field $\sigma$, i.e., a vison, converts this into a regular $hc/2e$ BCS vortex.

For ease of presentation, we specialize on the case of $s$-wave pairing and an even number of electrons per unit cell. In that case, the term $S_4$ may be dropped from the action. Further, the spinons are gapped and can be integrated out. This will lead to an innocuous renormalization of the value of $K_\sigma$.

In the vortex description, superconducting phases correspond to vortex vacua. To analyze these, it is then appropriate to imagine integrating out the vortices. This will renormalize the value of $K_\mu$ (or generate it if not present originally). The resulting action has only the terms,

$$S = S_\sigma + S_\mu + S_{CS}.$$  

The term $S_\sigma$ leads to a gapless linear dispersing excitation (in the absence of long-ranged Coulomb interactions), and corresponds physically to the sound modes of the superconductor. The remaining three terms only involve the two $Z_2$ gauge fields $\sigma$ and $\mu$. As shown in Appendix C, this action is equivalent to that of the $Z_2$ gauge theory with Ising matter fields. If we choose to integrate out the $\mu$, it is exactly the same as the Ising effective action derived in Sec. VII A to discuss the superconducting phases. Alternatively, we can integrate out the $\sigma$ field to obtain the dual theory as in Eq. (140),

$$S = S_{\mu},$$  

where $\mu_{ij}$ is a dynamical $Z_2$ gauge field and, as before, $\mu_{ij}$ is an “applied” field with $\prod_{\Box} \mu_{ij}^2 = -1$ through all spatial plaquettes. This theory is a direct $Z_2$ analog of a $U(1)$ superconductor in the presence of an applied magnetic field.

Consider briefly the phase diagram in the $t_d$–$K$ plane. A schematic phase diagram is shown in Fig. 8. Progress can be made in various limiting regimes. For $K_d=0$ the theory reduces to a pure $Z_2$ gauge theory with gauge field, $\tilde{\mu}_{ij} = \mu_{ij}^\ast \mu_{ij}$. Since $\mu_{ij}^\ast$ plays no role in this limit, the resulting phases are identical to that with even integer $N_0$ analyzed in Sec. VII A. In particular, for large $t_d$, we have a conventional superconductor $\text{SC}$ with broken $Z_2$ gauge symmetry, while for small $t_d$, we get the exotic superconductor $\text{SC}^\ast$. These phases survive for small $K_d$. It is easy to establish the absence of phase transitions for $t_d = \infty$ and $K=0$. For $t_d = \infty$, on the other hand, one can set $\mu_{ij} = \mu_{ij}^\ast$, and the model reduces to the fully frustrated Ising model. As discussed extensively in Sec. IV, the results of Ref. 16 show the existence of an ordered phase for large $K_d$ where translational symmetry is spontaneously broken. In general, this is expected to lead to spin-Peierls order. In this case, though, the spin-Peierls order coexists with superconductivity. We will denote this phase as $\text{SC} \rightarrow \text{SP}$. Several other ordered phases are presumably also possible although we will not discuss these here.

The $\text{SC} \rightarrow \text{SP}$ phase is the external gauge field “penetrates” with $\mu_{ij} \rightarrow \mu_{ij}^\ast$, and the Ising model is frustrated. But as $t_d$ is reduced, it eventually becomes favorable to “screen” this external field, and enter a Meissner phase with $\prod_{\Box} \mu_{ij} \approx 1$. When this happens the broken translational symmetry disappears, along with the frustration, and one enters into $\text{SC}$. 

In the $\text{SC} \rightarrow \text{SP}$ phase the external gauge field “penetrates” with $\mu_{ij} \rightarrow \mu_{ij}^\ast$, and the Ising model is frustrated. But as $t_d$ is reduced, it eventually becomes favorable to “screen” this external field, and enter a Meissner phase with $\prod_{\Box} \mu_{ij} \approx 1$. When this happens the broken translational symmetry disappears, along with the frustration, and one enters into $\text{SC}$. 

By the time quantum fluctuations become important, the Meissner phase will become unstable, and enter a superconducting phase with fields varying on the scale of the superconducting coherence length.
\[ S_{\text{vis}} = -K^d_{\sigma} \sum_{ij} v_i \mu_{ij} v_j. \]  \hspace{1cm} (145)

Here \( \tanh(K^d_{\sigma}) = e^{-2K^d_{\sigma}} \), so that \( K^d_{\sigma} \) is the coupling dual to \( K_{\sigma} \). Once again, \( v_i \) creates a vison, whose \( Z_\sigma \) current is equal to the flux of the \( \sigma \) field. On the other hand, the vortex configurations of the gauge field \( \mu \) correspond to the ison excitations.

As discussed earlier, the \( Z_\sigma \) gauge theory with matter fields has two phases: a Higgs-confined phase and a deconfined phase. The Higgs-confined phase describes the conventional superconductor \( SC \), and is perhaps easiest to understand in the limit in which both \( K_{\mu} \) and \( K^d_{\mu} \) are small. With small \( K_{\mu} \) the gauge field is in its confining phase, so that test charges coupling to the gauge field \( \mu \) are confined. There are actually two different particles minimally coupled to \( \mu \)—the \( hc/2e \) vorton and the vison, with creation operator \( e^{i\theta} \) and \( v_1 \), respectively. As before, the confined vorton—vison bound state is the conventional \( hc/2e \) BCS vortex.

The deconfined phase describes the exotic superconductor \( SC^e \). In this phase, test charges that couple to \( \mu \) are deconfined. This implies that the \( hc/2e \) vorton and the vison are not bound together, and can propagate as independent gapped excitations, in agreement with the earlier discussion. In effect, within \( SC^e \) the Chern–Simons term has been rendered ineffective and does not attach flux. Also, configurations with \( \pi \) flux in the gauge field \( \mu \), corresponding to the ison, exist as finite energy excitations. Thus, as before, we conclude that there are four gapped excitations in \( SC^e \)—the \( hc/2e \) vorton, the spinon, the vison, and the ison.

Note that a transition from \( SC^e \) to an insulator obtained by condensing the \( hc/2e \) vortons [which are the fundamental \( U(1) \) vortices in this phase] leads naturally to the fractionalized insulator \( \mathcal{I}^e \). This is because the vison is unbound from the \( hc/2e \) vorton in \( SC^e \), so that condensation of the latter leaves the former uncondensed. Indeed, the distinct excitations in the resulting insulator are the chargons, the spinons, and the vortons, as appropriate to \( \mathcal{I}^e \). Thus, the exotic insulator \( \mathcal{I}^e \) may either be reached from \( SC \) by condensing \( hc/e \) vortices or from \( SC^e \) by condensing \( hc/2e \) vortons. In either case, the vison remains uncondensed.

This completes the dual description of \( SC^e \). Complications such as \( d \)-wave pairing or arbitrary filling \( N_0 \) can be handled straightforwardly in this dual formulation as well, although we shall not do so here.

VIII. EXTENSION AND GENERALIZATIONS

A. General spatial dimension

The \( Z_\sigma \) gauge theory formulation (in the particle representation) is readily generalized to arbitrary spatial dimension. The cases of physical interest are 3D and 1D, which we discuss in turn. For simplicity, we will restrict our attention to situations with integer filling per unit cell. The most important effect of spatial dimensionality enters into the properties of the pure \( Z_\sigma \) gauge theory with action,

\[ S = S_\sigma + S_B, \]  \hspace{1cm} (146)

with \( S_B \) included when there is an odd number of electrons per unit cell.

1. \( d = 3 \)

In 3D and in the absence of \( S_B \), the \( Z_2 \) gauge theory again has two phases distinguished by the behavior of the Wilson loop correlator (area law versus perimeter law). As in 2D, the presence of \( S_B \) will enhance the stability of the perimeter phase, but the area law phase will still be present. The presence of the perimeter law phase implies the existence of 3D insulators with electron fractionalization. But in contrast to 2D, the flux tubes in the \( Z_2 \) gauge field, the visons, are not pointlike excitations, but become extended string-like excitations in 3D. The area law phase again describes various confined insulating phases. Whether the presence of \( S_B \) leads to broken translational symmetry as in 2D is an interesting unanswered question. Note, however, that in 3D it is not possible to pass to a dual global Ising model. In fact, the pure \( Z_2 \) gauge theory (in the absence of \( S_B \)) is in fact self-dual\(^\text{12} \) in three spatial dimensions.

To discuss the superconducting phases \( SC \) and \( SC^e \), it is necessary to understand the properties of the \( Z_2 \) gauge theory coupled to Ising matter fields. In the absence of \( S_B \), it is known\(^\text{31} \) that in three spatial dimensions, there are again two phases, the Higgs-confined phase and the deconfined phase. These correspond to \( SC \) and \( SC^e \), respectively. Their distinguishing properties will be qualitatively similar to the 2D case. As in 2D, we expect that the main effect of \( S_B \) would only be to make possible the existence of an \( SC \) phase with broken translational symmetry.

In layered quasitwo dimensional systems, fractionalized insulating phases in which each layer is decoupled from the others are possible, and exist as distinct phases from the isotropic ones discussed above. Such phases are currently under further investigation.

Finally, it is worth emphasizing that while the extension to 3D is straightforward in the particle representation, the dual vortex representation necessarily involves string-like vortex degrees of freedom.

2. \( d = 1 \)

In one spatial dimension (1D), the \( Z_2 \) gauge theory is always in its area law phase, with or without the \( S_B \) term. Thus, our formulation is incapable of describing electron fractionalization in one dimension. Evidently, fractionalization in \( d = 1 \) must have different physical origins than for \( d > 1 \). To highlight this point, note that 1D fractionalization can be continuous, as exemplified by the spinless Luttinger liquid which supports charge-carrying excitations with essentially arbitrary (even irrational) charge. For \( d > 1 \), on the other hand, fractionalization is discrete—the fractionally charged excitations carry a definite rational fraction of the electron charge. As in the fractional quantum Hall effect, this discreteness can be traced to the binding (and condensation) of a discrete number of vortices. These physics appear to be qualitatively different from the “solitonic” mechanism responsible for fractionalization in 1D.

B. Finite temperature

In our formulation there is a sharp distinction between fractionalized and confined phases at zero temperature, which is independent of whether or not the phases in ques-
tion have any sort of conventional long-ranged order. It is extremely interesting to ask whether this sharp distinction survives at finite temperature. Consider first the deconfined phases in 2D. In these phases, the point-like vison excitations are gapped at zero temperature. However, since the energy cost to create a vison is finite, at any nonzero temperature there will be a nonvanishing density of thermally excited visons. In the absence of other kinds of order (e.g., magnetic), this low temperature regime will be smoothly connected to the high temperature limit, without an intervening finite temperature transition. Thus, in 2D the sharp distinction between fractionalized and confined insulators does not survive at finite temperature.

But in 3D, the vison excitations in the deconfined phase are string-like extended objects, with an energy cost proportional to their length. Consequently, at low temperatures arbitrarily large vison loops will not be thermally excited—the vison loops will be “bound.” As the temperature increases, there will be a transition at which the vison loops unbind and proliferate. Thus, the fractionalized insulator in three spatial dimensions undergoes a finite temperature phase transition associated with the unbinding of vison loops. A defining characteristic of the low temperature phase is that vison loops will cost a free energy linear in their length. Equivalently, $\text{hc} / 2 \text{e}$ (or $Z_2$) magnetic monopole “test charges” are confined even at finite temperature, with an infinite free energy cost to separate them. A confinement of monopoles is also one of the characteristics of a 3D superconductor, but quite remarkably the confinement here is occurring in a “normal” nonsuperconducting phase. The conventional insulating phases with confinement at zero temperature, on the other hand, will not exhibit finite temperature transitions (other than those associated with the loss of conventional long-ranged order, e.g., magnetic).

To understand the origin of these results, we briefly discuss the properties of the pure $Z_2$ gauge theory (with no matter fields) in 3 + 1 space-time dimensions in more detail. At zero temperature the theory is self-dual:12 the duality transformation interchanges the “electric” and “magnetic” fields of the gauge theory. For $K > K_c$, when the gauge theory is in its deconfining phase, the theory has string-like vison excitations (which are $Z_2$ magnetic flux tubes) with a finite energy cost per unit length. For $K < K_c$, the gauge theory confines with area law Wilson loops, but there are nevertheless string-like excitations in this phase as well. These can be understood via duality, which interchanges the area and perimeter law phases; the string-like excitations in the area law phase are simply flux tubes of the dual $Z_2$ gauge field. Physically, these dual tubes are “electric flux tubes” responsible for the confinement of electric charge in the area law phase. Specifically, when two test $Z_2$ electric charges separated by a distance $R$ are introduced into the system, the resulting electric flux is concentrated in a tube that extends from one test charge to the other with an energy cost proportional to $R$, the linear confinement. Similarly, in the perimeter phase, dual test charges ($Z_2$ monopoles) that act as sources for the visons are confined.

Now consider the properties of the gauge theory at finite temperature. The phase diagram is well known33 and is shown in Fig. 9. There are three finite temperature phases. For $K > K_c$, at small but nonzero temperatures, large (“magnetic”) vison loops are bound as their energy cost is proportional to their length. Similarly, for $K < K_c$ at low temperatures, large electric flux loops are bound. At high temperature, for any $K$, both kinds of loops are unbound. The transition from the low temperature to the high temperature phase is therefore associated with the unbinding of (electric) magnetic vison loops for $K$ (lesser) greater than $K_c$.

In the low temperature phase for $K > K_c$, the free energy of an isolated static test electric charge diverges so that test charges are confined. In the high temperature deconfined phase, the free energy cost is finite. Formally, the pure $Z_2$ gauge theory has a global Ising symmetry at finite temperature which is broken in the high temperature phase. As shown by Polyakov,32 a convenient characterization of this transition is through the following operator:

$$L_r = \prod_{n=0}^{M-1} \sigma_r^{n, \tilde{r}} + \sigma_r^{n+1, \tilde{r} + (n+1) \hat{r}} \tag{147}$$

for each site $\tilde{r}$ of the spatial lattice. Here $\hat{r}$ is a vector along the (imaginary) time direction of length the time slice. The product is over all the temporal links at that site, and $M$ is the number of time slices. This operator $L_r$ is often referred to as the Polyakov loop. The free energy $F(r, r')$ to introduce two test charges at sites $r, r'$ is directly related32 to the correlator of $L_r$ through

$$e^{-F(r, r') / T} = \langle L_r, L_{r'} \rangle. \tag{148}$$

Thus, test charges will be confined if this correlator goes to zero at large distances; on the other hand, if this correlator goes to a constant, the test charges will be deconfined. Furthermore, consider the following transformation on the gauge fields:

$$\sigma_r^{n, \tilde{r}, \tilde{r} + (n+1) \hat{r}} = e^{\epsilon} \sigma_r^{n, \tilde{r}, \tilde{r} + (n+1) \hat{r}}, \tag{149}$$

where $\epsilon = \pm 1$ independent of $r$, and $n_0$ is fixed. The action of the pure gauge theory is invariant under this transformation, implying a global Ising symmetry of the theory. The operator $L_r$, however, transforms as

$$L_r \to \epsilon L_r. \tag{150}$$
Thus $L_r$ is an order parameter for this global Ising symmetry. In the low temperature phase for $K<K_c$, $L_r$ has no expectation value, the global Ising symmetry is unbroken, and test charges are confined. At high temperatures, however, $L_r$ acquires an expectation value breaking the global Ising symmetry, and the test charges are deconfined.

For $K>K_c$, the self-duality of the $Z_2$ gauge theory implies the existence of a dual global Ising symmetry, with an order parameter that is the dual analog of the Polyakov loop. In the low temperature phase, this dual global symmetry is unbroken: in this phase dual test charges (i.e., $Z_2$ monopoles) are confined. At high temperatures this dual global symmetry is broken and the dual test charges are deconfined.

Consider next the effects of coupling matter fields (the chargons and the spinons) to the $Z_2$ gauge field. As these carry $Z_2$ gauge electric charge, it is easy to see that the action is no longer invariant under the transformation in Eq. (149). Indeed, this transformation is equivalent to changing the boundary conditions on the chargon fields from $(\beta)$ periodic to antiperiodic, and vice versa for the spinons. Moreover, if the matter coupling is weak, the matter fields may formally be integrated out\cite{35} to leave behind a “magnetic field” term that couples linearly to the Polyakov loop order parameter of the global Ising symmetry. There is then no longer any transition separating the low and high temperature regimes. Physically, this is exactly as expected: for $K<K_c$, the electronic system is in a conventional confined insulating phase at zero temperature.

On the other hand, since the chargons and spinons do not carry any dual $Z_2$ magnetic charge, the dual global Ising symmetry remains even in their presence. The finite temperature transition for $K>K_c$ should thus remain in fact. Consequently, we arrive at the striking conclusion that the three-dimensional fractionalized insulator undergoes a finite temperature transition associated with the unbinding of vison loops. This conclusion will not be affected by the Berry’s phase term $S_g$, which is quite innocuous in the fractionalized insulator.

C. Spin-rotation noninvariant systems

The $Z_2$ gauge theory formulation (in either the particle or vortex representations) works equally well in the absence of spin rotation invariance. In particular, fractionalized phases continue to exist even when spin is not a good quantum number. (Spinless fermion systems can also be handled with no fundamental modifications.) For these reasons, we have avoided the term spin-charge separation, in favor of the more general term electron fractionalization.

D. Analogies with nematics

Certain aspects of our formulation might be familiar from the classical statistical mechanics of nematics. The order parameter for a nematic is a headless three component vector. Lattice models of nematics are usually formulated in terms of an ordinary three component vector, the headless nature being incorporated through a local $Z_2$ gauge symmetry which inverts the local vector order parameter. Here, we briefly explore the analogies between the classical phases of nematic systems and the quantum phases discussed in this paper.

The analogy is closest if we consider $s$-wave pairing with an even number of electrons per unit cell and, further, integrate out the spinons to work with just the chargons and the $\sigma$ field. The action describing the system is then

$$S = -2t_c \sum_{ij} \sigma_{ij} \cos(\phi_i - \phi_j) - K \sum i \sum j \sigma_{ij}. \quad (151)$$

As formulated, this describes a quantum problem of chargons coupled to a fluctuating $Z_2$ gauge field in two spatial dimensions. But alternately, we may view it as a classical Hamiltonian for a three-dimensional XY nematic. Indeed, an $O(3)$ version of the same model was introduced a few years ago by Lammert, Rokhsar, and Toner\cite{34} to describe nematic ordering in three dimensions. Further, they argued that their lattice gauge nematic model admits three distinct phases, an ordered nematic phase, and two isotropic phases. The nematic phase breaks the rotational symmetry and the $Z_2$ gauge symmetry. For an $XY$ system, this is the direct analog of the superconducting phase. Moreover, the physical $hc/2e$ vortices of the superconductor correspond directly to the “disclinations” in the nematic fluid.

The two isotropic phases in the nematic are distinguished\cite{34} by the free energy cost per unit length to externally impose a disclination line through the system. In particular, in the conventional isotropic phase, the free energy cost per unit length is zero (as the length goes to infinity). The disclinations are condensed. But, in the unconventional isotropic phase\cite{34}, the free energy cost per unit length is a constant (as the length goes to infinity). In the context of this paper, the isotropic phases correspond to insulating phases. As we have elaborated at length, there are two insulating phases $I$ and $I^*$ which are distinguished by whether or not the visons (which are the relics of the $hc/2e$ vortices in the insulating phases) are condensed. Thus, the conventional insulator corresponds, in the nematic analog, to the conventional isotropic phase. Note that the energy cost of a vison (which is the action cost per unit length of the worldline) is zero in this phase. Similarly, the fractionalized insulator $I^*$ corresponds to the unconventional isotropic phase of the $XY$ nematic. In $I^*$ the visons have finite energy cost, again just like the disclination lines in the unconventional isotropic fluid.

The phase transition between $SC$ and either insulating phase is second order. In contrast, for the $O(3)$ nematic system considered in Ref. 34, the transition between the nematically ordered phase and the conventional isotropic phase is first order, while that to the other isotropic phase is second order. This difference is due to the $XY$ symmetry of the superconducting system, as opposed to the $O(3)$ symmetry of the nematic.

For the more general situation, with coupling to the spinons or with an odd number of electrons per unit cell, a direct correspondence with the nematic system no longer holds. Nevertheless, we believe that the discussion in this subsection may help (some) readers get further intuition and insight into our formulation.

IX. RELATION TO PREVIOUS APPROACHES

We now comment on the connection between the $Z_2$ gauge theory and earlier approaches to electron fractionaliza-
tion. We begin by making contact with earlier papers on the nodal liquid. Earlier formulations of the nodal liquid (in Refs. 9 and 10) focused on the importance of “vortex pairing” as a means to describe charge fractionalization in two dimensions. In Ref. 9 a theory was formulated in terms of vortices in a local superconductor pair field, and it shares many features with the approach taken here, particularly the dual formulation detailed in Sec. V. In Ref. 10, Chern–Simons theory was used to convert spinful electrons into bosons, and a dual formulation was developed in terms of vortices in these bosonic fields. The $Z_2$ gauge theory and its dual Ising Chern–Simons vortex theory developed in this paper not only ties together both earlier approaches into a unified framework, but allows for a more direct quantitative analysis of “microscopic” models. We now describe this connection in a bit more detail.

In Ref. 9 a spinon operator was defined as an electron with its charge screened by “one half” of a Cooper pair. The latter coincides precisely with the chargon introduced in Eq. (with charge screened by “one half” of a Cooper pair. The connection in a bit more detail.

In Ref. 9 a spinon operator was defined as an electron with its charge screened by “one half” of a Cooper pair. The latter coincides precisely with the chargon introduced in Eq. (28), showing the equivalence of the spinons as well. The importance of the long-ranged interaction between the spinon and $hc/2e$ vortex was emphasized in Ref. 9. It was suggested that this interaction could be implemented by employing a $U(1)$ Chern–Simons term to attach flux to both species of particles. But since the spinon number is not conserved, it was suggested that the flux could be attached to the (conserved) $z$ component of the spin. Moreover, it was argued in Ref. 9 that due to the statistical interactions, condensation of $hc/2e$ vortices should lead to confinement of spinons. In the dual vortex formulation presented in this paper the statistical interaction between vortex and spinon is described in terms of a novel Ising-like Chern–Simons term. It is important to stress that this does not require the spin of the spinon to be conserved, in contrast to the $U(1)$ approach, since the Ising-flux is attached to the conserved $Z_2$ charge of the spinons. Moreover, the Ising formulation clearly shows that condensation of the $hc/2e$ vortices, or the visors, leads to confinement of spinons and chargons. In the global Ising model for the visors with $\langle \langle \rangle \rangle = 0$, the linear confinement is due to the required line of negative Ising couplings connecting the two spinons. In the $Z_2$ gauge theory formulation, it follows from the area law for the Wilson loop.

In Ref. 10, a theory was developed by converting spinful electrons into spinful bosons, using Chern–Simons to attach flux to the electrons spin, and then passing to a dual representation of vortices in these bosonic fields, denoted $\Phi_a$ with spin label $a = \uparrow, \downarrow$. A lattice version of this theory can be written in terms of the phases, $\theta_a$, of the vortex field operators, $\Phi_a = e^{i\theta_a}$, with effective Euclidian action,

$$ S = -t_v \sum_{\langle ij \rangle} \cos(\theta_{ia} - \theta_{ja} + a_{ij}^\alpha) + S_{\text{cs}}(\theta^\alpha). $$

(152)

Here, $i, j$ are label sites of the $2+1$ space-time lattice, $t_v$ is a dimensionless vortex hopping term and $S_{\text{cs}}$ is a Chern–Simons term involving the field $\theta^\alpha = a^\uparrow - a^\downarrow$. The curl of $a^\alpha$ corresponds to the conserved electrical current of the electrons with spin $a$. In Ref. 9, two different composite “pair” vortex operators were considered:

$$ \Phi_\rho = \Phi_{a} \Phi_{\downarrow} = e^{i\theta_\rho}; \quad \Phi_\sigma = \Phi_{a} \Phi_{\downarrow} = e^{i\theta_\sigma}, $$

(153)

which are minimally coupled to $a^\rho,a^\sigma = a^\uparrow \pm a^\downarrow$, respectively. The action can be re-expressed in terms of these composite phase fields using the relation

$$ \theta_\perp = \frac{1}{2} (\theta^\rho - \theta^\sigma) + \frac{\pi}{2} \nu, $$

(154)

giving

$$ S = -t_v \sum_{\langle ij \rangle} v_i v_j \cos((\theta_i^\rho - \theta_j^\rho + a_{ij}^\rho)/2) \cos((\theta_i^\sigma - \theta_j^\sigma + a_{ij}^\sigma)/2). $$

(155)

Here, the Ising spins $v_i = \pm 1$ are the visons. The primary emphasis of Ref. 10 was an analysis of fractionalized phases, such as the nodal liquid. It was emphasized that fractionalization occurs when $\langle \langle \rangle \rangle = 0$, and breaking the Ising symmetry with $\langle \langle \rangle \rangle \neq 0$ corresponds to confinement. Deep within the deconfined phase it is possible to integrate out the massive visons, which generates local terms such as

$$ S_{\text{hc/e}} = -t_{2e} \cos(\theta_i^\rho - \theta_i^\sigma + a_{ij}^\rho), $$

(156)

which describes the hopping of the $hc/e$ vortex pair, $\Phi_\rho$, and

$$ S_{\text{spinon}} = -t_s \cos(\theta_i^\rho - \theta_i^\rho + a_{ij}^\rho). $$

(157)

Due to the Chern–Simons terms above, this corresponds to the hopping of fermionic spinons which carry $S_z = 1/2$.

The relationship between this formulation, in terms of “electron” vortices, and the dual vortex theory of Sec. V constructed in terms of BCS $hc/2e$ vortices is at first not apparent. But consider introducing a vortex operator, $\Phi = e^{i\theta}$, whose square equals the $hc/e$ vortex pair operator, $\Phi^2 = \Phi_{\rho}$. This requires that

$$ \theta = \frac{1}{2} \theta_{\rho} + \frac{\pi}{2} (1 - \nu), $$

(158)

which implies

$$ \Phi = v e^{i\theta_{\rho}/2}. $$

(159)

As defined $\Phi$ carries vorticity $hc/2e$, and can tentatively be identified as the BCS vortex. To complete this identification it is necessary to show that there is a long-ranged statistical interaction between this $hc/2e$ vortex and the spinon. Evidence for this is provided by the following argument. We first imagine explicitly adding the vortex hopping term $S_{\text{spinon}}$ to the action in Eq. (155). We then absorb the field $\theta^\sigma$ into $a^\rho$. We may now re-express the action of Eq. (155) in terms of $\theta_\perp$:

$$ S = -t_v \sum_{\langle ij \rangle} \mu_{ij} \cos(\theta_i - \theta_j + \frac{1}{2} a_{ij}) + S_{\text{spinon}}, $$

(160)

with

$$ \mu_{ij} = \cos \left( \frac{1}{2} a_{ij}^\rho \right). $$

(161)

Here, we have defined $a_{ij} = a_{ij}^\rho$. In the presence of the vortex hopping term $S_{\text{spinon}}$ above, if we specialize to the limit of large $t_v$, it is legitimate to restrict $a_{ij}^\rho$ to be $2 \pi$ times an
integer. With that restriction the gauge field \( \mu_{ij} = \pm 1 \), reducing to an Ising \( Z_2 \) gauge field. Now, imagine putting a stationary spinon on one site of the original spatial lattice. In this dual vortex representation this corresponds to a plaquette with \( \Delta \times a^\pi = 2 \pi \) or, equivalently to a product \( \Pi_\square \mu_{ij} = -1 \) for all plaquettes pierced by the spinon “world line.” Since the \( hc/2e \) vortex is minimally coupled to \( \mu_{ij} \), this establishes that it does indeed acquire a minus sign upon being transported around a spinon. In the dual vortex formulation in Sec. V, a \( \pi \)-flux tube in \( \mu_{ij} \) is attached to each spinon by the Ising-like Chern–Simons term. To complete the mapping between these two formulations requires, finally, re-fermiionization of the spinon creation operator, \( e^{i\phi} \) [fermionic due to the Chern–Simons term \( S_{cs}(a^\pi) \)] effectively replacing it with spinful fermions \( f_{ia} \).

Finally we comment briefly on the relationship with theories based on slave boson/fermion approaches to electron fractionalization. A number of authors have examined insulating Heisenberg antiferromagnetic spin models in the hope of finding phases with deconfined spinon excitations through these approaches. However this program has generally been quite unsuccessful—the \( U(1) \) or \( SU(2) \) gauge symmetry introduced in the slave-boson or fermion representations ultimately leads only to confined phases. A notable exception however is the work of Read and Sachdev on large-N \( Sp(2N) \) frustrated antiferromagnets and related quantum dimer models. Under certain special conditions, these authors demonstrated the existence of quantum disordered phases with deconfined spinons in their theory. It is worth pointing out that fractionalization is achieved when the \( U(1) \) gauge symmetry [introduced by the Schwinger boson representation of the \( Sp(2N) \) spins] is broken down to \( Z_2 \) by condensation of pairs of bosons. The fully frustrated transverse field Ising model appears in that description as well.

Slave boson representations of electron operators have been used extensively to discuss spin-charge separation issues in doped \( t-J \) models. However, the resultant compact \( U(1) \) or \( SU(2) \) gauge theories presumably always lead to confinement, unless the gauge symmetry is broken down to \( Z_2 \). This may be achieved by pairing the spinons. Indeed, the slave-boson mean field treatments of the \( t-J \) model do find pairing of spinons below a finite temperature at low doping. As we have emphasized in this paper though, even in the undoped limit and without frustration, the Heisenberg spin model may be rewritten in terms of fermionic spinon operators coupled to a fluctuating \( Z_2 \) gauge field. Equivalently spinon pairing terms may be added to the Hamiltonian describing the Heisenberg magnet without altering any of the physical symmetries. We have shown that electron fractionalization is definitely possible once charge fluctuations are incorporated into the description.

X. CONCLUSION AND DISCUSSION

A. Summary

The primary focus of this paper was to explore the possibility of electron fractionalization in strongly correlated electron systems in spatial dimension greater than 1, and in the presence of time reversal symmetry. We based our discussion on a particular class of microscopic models designed to capture the physics essential to the cuprates, although our description of fractionalization is more general. Starting from these models, we developed a new gauge theory of strongly correlated systems consisting of charge \( e \), spin-zero bosons (the chargons) and charge zero, spin 1/2 fermions (the spinons), both minimally coupled to a fluctuating \( Z_2 \) gauge field. Remarkably, the spin sector of the theory at half filling and in the absence of charge fluctuations is formally identical to a spin one-half Heisenberg antiferromagnet. In this limit the \( Z_2 \) gauge field enforces the constraint that the spinon number on each site is odd, physically equivalent to the single occupancy constraint, imposed with additional unwanted redundancy in earlier \( U(1) \) gauge theory formulations of the Heisenberg model.

Charge fluctuations, however, are naturally incorporated into our \( Z_2 \) gauge theory, and when they become large the theory describes a \( d_{z^2-r^2} \) superconductor. Analysis of the theory in the intermediate region reveals that there are two qualitatively different routes for the evolution from the antiferromagnet to the superconductor. One route is through conventional insulating phases in which fluctuations of the \( Z_2 \) gauge field confines together the chargon and the spinon, leaving only the electron in the spectrum. But a more interesting possibility takes one through phases in which the electron is fractionalized, and the chargons and spinons exist as deconfined excitations. With \( d_{z^2-r^2} \) pairing, this fractionalized insulator is the nodal liquid, with gapless spinon excitations at four points of the Brillouin zone. It seems likely that the ultimate transition from the insulating phases to the \( d_{z^2-r^2} \) superconductor occurs close to the boundary between the confined and deconfined insulating phases. Thus, which of these two qualitatively different routes is realized in any particular experimental system could depend sensitively on microscopic details.

In addition to the chargons and spinons, the 2D nodal liquid supports Ising-like point excitations, the visons, which correspond to vortices in the \( Z_2 \) gauge field. These gapped vison excitations play a central role in our analysis of fractionalization, as becomes clear upon passing to a dual description in terms of \( hc/2e \) BCS vortices (of a conventional superconductor) and the spinons. In this dual framework, the nodal liquid can be accessed by a pairing and condensation of the \( hc/2e \) vortices, as emphasized in earlier work. This reveals that the vison excitations are simply the remnant of the unpaired \( hc/2e \) vortices which survive in the insulating nodal liquid.

The utility of the vison excitations goes far beyond giving a simple description of the nodal liquid. Indeed, the pure \( Z_2 \) gauge theory in 2 + 1 space-time dimensions is dual to the global 2 + 1 dimensional Ising model, and the Ising spins are simply the vison creation operators. Remarkably, an unusual Berry’s phase term in the gauge theory corresponds simply to frustration in the dual Ising model, with full frustration at half filling. The fully frustrated quantum Ising model arose in earlier work by Sachdev and co-workers in their analysis of frustrated magnets. Ordering the dual Ising model by condensation of the visons generally will break translational symmetry and lead to conventional confined insulating phases such as the spin-Peierls phase. In three spatial dimensions (3D), the visons become loop-like excitations, and are closely related to vortex-line excitations which occur in a
conventional superconductor. Surprisingly, this implies that a 3D fractionalized insulator “survives” at finite temperature, being separated from the high temperature regime by a finite temperature phase transition. As in a conventional superconductor, the 3D fractionalized insulator confines \( h_{c2}e \) monopole excitations even at nonzero temperature.

Within the \( Z_2 \) gauge theory approach, a conventional superconductor is described as a condensate of charge \( e \) chargons. A superconducting phase involving condensation of chargon pairs (i.e., Cooper pairs) without condensation of single chargons was shown to exist, this has several exotic properties distinguishing it from the conventional superconductor.

### B. Experiments

We close with a very brief discussion of some of the experimental signatures of electron fractionalization. As we will see, experimental detection of fractionalization may be quite subtle. Further theoretical understanding of fractionalized phases leading to detailed experimental predictions are clearly called for. Our discussion will necessarily be brief.

1. **Two-dimensional nodal liquid**

   Earlier work on the nodal liquid\(^9,21\) outlined a number of experimental signatures of the two-dimensional nodal liquid, and we have little to add here. As pointed out in the earlier papers, perhaps the most telling indication will be in angle resolved photoemission (ARPES) which directly measures the electron spectral function as a function of the momentum \( k \), and frequency \( \omega \). As the electron is fractionalized into the chargon and the spinon in the nodal liquid, its spectral function will not have a sharp quasiparticle peak even at zero temperature. Note that bound states of the chargon and the spinon (which could lead to sharp spectral features) are not expected here at low energies as the spinons are gapless.

2. **SC*\**

   We have discussed the basic physics of the exotic superconductor \( SC^* \) obtained by condensing chargon pairs in Sec. VII. There are several qualitative experimental distinctions between this phase and the conventional superconductor which we now briefly discuss. The most striking is again in the electron spectral function as measured in ARPES. As discussed in Sec. VII, the electron decays into a spinon and an Ising part of the charge, the ison excitation. Thus, we expect that the electron spectral function does not have a sharp quasiparticle peak in the \( SC^* \) phase. Again, since the isons are massive excitations while the spinons are gapless, bound states of the two are generally not expected at low energies. The presence of gapped ison excitations would also affect the thermodynamics, and contribute to the thermal conductivity at some intermediate temperatures. However, these signatures are likely to be quite subtle. A striking theoretical feature of \( SC^* \) is that the conventional BCS \( h_{c2}e \) vortices are splintered into pieces, the \( U(1) \) “vorton” carrying the circulating electrical currents, and the \( Z_2 \) vison. Since the spinons do not have a long-ranged statistical interaction with the \( h_{c2}e \) vortex, it is tempting to speculate that the structure of the core states in such a vorton would be qualitatively different from that of an \( h_{c2}e \) vortex in a conventional superconductor.

   The experiments on the cuprates mostly do see a sharp quasiparticle peak\(^35\) inside the superconducting state (although there is some recent controversy\(^36\) as to what extent this is true along the nodal directions). Given the current experimental status, we therefore guardedly identify the superconducting phase in the cuprates with \( dSC \) and not \( dSC^* \). However \( dSC^* \) is nevertheless interesting to consider on theoretical and conceptual grounds.

### 3. Three-dimensional effects

In striking contrast to a two-dimensional nodal liquid, a genuinely three-dimensional nodal liquid has a finite temperature phase transition associated with the unbinding of vison loops. This phase transition could lead to observable singularities in the measured properties of the system. But due to the highly anisotropic nature of the cuprates, it is perhaps more natural to speculate that a fractionalized phase would consist of decoupled 2D systems, with a confinement of spinons within each layer. Clarification of such interlayer confinement physics will be necessary in order to disentangle the subtle interlayer behavior of the cuprate materials, both in the normal and superconducting phases.

### ACKNOWLEDGMENTS

The authors are grateful to Phil Anderson, Leon Balents, Tom Banks, Eduardo Fradkin, Doug Scalapino, and R. Shankar for insightful discussions, and Ashvin Vishwanath and Tony Zee for useful comments on the manuscript. They would particularly like to thank Subir Sachdev for many helpful comments and his explanations of \( Z_2 \) gauge theories in the context of frustrated large \( N \) magnets. Thanks are due the Aspen Center of Physics where some of this work was carried out. This research was generously supported by the NSF under Grant Nos. DMR-97-04005, DMR95-28578, and PHY94-07194.

### APPENDIX A: PATH INTEGRAL

We now derive a path integral expression for the partition function of the spinon–chargon Hamiltonian. A crucial role is played by the constraint on the Hilbert space, which naturally introduces an \( Z_2 \) gauge field.

To this end, we work with fermionic coherent states built from the spinon operators, \( f_a \) and \( \bar{f}_a \), which are defined in standard fashion.

\[
|f_a\rangle = e^{-\bar{f}_a f_a}|0\rangle, \tag{A1}
\]

\[
\langle \bar{f}_a| = \langle 0| e^{\bar{f}_a f_a}, \tag{A2}
\]

where the spinon operators are denoted with “hats,” and \( f_a \) and \( \bar{f}_a \) are Grassman numbers. The bra and ket states denoted with a “0,” are fermionic Fock states with no spinons present. Here we have suppressed the dependence of the fermion operators and Grassman fields on the spatial coordi-
In the charge sector of the theory we choose a basis of states diagonal in the phase $\phi$ of the chargon field, denoted $|\phi\rangle$.

The partition function in Eq. (42) can then be expressed as

$$Z = \int d\tilde{f}_{\tau} df_{\tau} \sum_{M=0}^{\infty} \sum_{\sigma_{\alpha}=\pm 1} e^{iN_{\alpha} \phi_{\alpha} - \tilde{f}_{\tau} \phi_{\alpha} + \pi/2(1-\sigma_{\alpha})} e^{\tilde{f}_{\alpha} f_{\alpha}^\dagger f_{\alpha}^\ast} Z_{\tau},$$

(A3)

with $\epsilon = \beta M$ and $P$ the projection operator defined in Eq. (38). Inserting the resolution of the identity between each time slice gives

$$Z = \prod_{\tau=1}^{M} \int d\tilde{f}_{\tau} df_{\tau} \sum_{\sigma_{\alpha}=\pm 1} e^{-\tilde{f}_{\alpha} f_{\alpha}^\dagger f_{\alpha}^\ast} Z_{\tau},$$

(A4)

with matrix elements

$$M_{\tau} = \langle \tilde{f}_{\tau} | \phi_{\tau} \rangle e^{-\epsilon \phi_{\tau} P},$$

(A5)

and appropriate boundary conditions on the fields, $f_{M+1} = -f_1$ and $\phi_0 = \phi_M$.

The matrix elements can be readily evaluated for small $\epsilon$ by inserting a complete set of states diagonal in the chargon number, $N$. Using the definition of the projection operator in Eq. (38) gives

$$M_{\tau} = \frac{1}{Z} \sum_{\sigma_{\alpha}=\pm 1} e^{iN_{\alpha} \phi_{\alpha} - \tilde{f}_{\alpha} \phi_{\alpha} + \pi/2(1-\sigma_{\alpha})} e^{\tilde{f}_{\alpha} f_{\alpha}^\dagger f_{\alpha}^\ast} Z_{\tau},$$

(A6)

with

$$E_{\tau} = e^{-\epsilon \phi_{\tau} P},$$

(A7)

Upon making the change of variables in the Grassman functional integral,

$$\sigma_{\alpha} f_{\alpha} \rightarrow f_{\alpha},$$

(A8)

the full partition function can finally be re-expressed as

$$Z = \prod_{\tau=1}^{M} d\tilde{f}_{\tau} df_{\tau} \sum_{\sigma_{\alpha}=\pm 1} e^{-\tilde{f}_{\alpha} f_{\alpha}^\dagger f_{\alpha}^\ast} Z_{\tau},$$

(A9)

with

$$S = S_F^0 + S_F^\phi + \epsilon \sum_{\tau=1}^{M} H(N_{\tau}, \phi_{\tau}, \tilde{f}_{\tau} f_{\tau}).$$

(A10)

with

$$S_F^0 = \sum_{\tau=1}^{M} \tilde{f}_{\tau} (\sigma_{\tau+1} f_{\tau+1} - f_{\tau}),$$

(A11)

and

$$S_F^\phi = -i \sum_{\tau=1}^{M} N_{\alpha} \phi_{\tau} - \phi_{\tau+1} + \pi/2(1-\sigma_{\tau}).$$

(A12)

Throughout, we have suppressed the explicit $r$ and $\alpha$ subscripts on the fields, displaying only the time-slice dependences.

**APPENDIX B: $Z_2$ GAUGE THEORY WITH $D_{x^2-y^2}$ PAIRING**

We will now provide an outline of a microscopic derivation of the $Z_2$ gauge theory in the presence of $d_{x^2-y^2}$ pairing correlations. We begin with the Hubbard-type Hamiltonian, Eq. (4), discussed in Sec. I:

$$H = H_0 + H_J + H_\Delta + H_u.$$  

(B1)

The crucial difference with the $s$-wave case is in the structure of the ‘pairing’ term $H_\Delta$.

We now follow exactly the same strategy as in the $s$-wave case, defining chargon and spinon operators. A path integral representation of the partition function is readily set up with the main difference being in the pairing term which becomes

$$S_\Delta = \epsilon \sum_{\tau} \Delta_{\tau} (b_{\tau}^\dagger b_{\tau} + c.c.) B_{\tau}.$$  

(B2)

$$B_{\tau} = \Delta_{\tau} (\tilde{f}_{\tau} f_{\tau}^\dagger f_{\tau}^\ast -(1 \rightarrow 1) + c.c.).$$  

(B3)

We have suppressed the $\tau$ index on all fields. It will be convenient to use a slightly different decoupling of the $H_J$ term. We write

$$e^{-S_{J}} = \int d\chi_{\tau} d\eta_{\tau}^\dagger d\eta_{\tau} d\eta_{\tau}^\dagger e^{-S_{J}} = S_{hs} + S_{h}[\eta].$$

(B4)

$$S_{hs} = S_{hs}[\chi] + S_{hs}[\eta].$$

(B5)

$$S_{hs}[\chi] = \epsilon \sum_{\tau} J(2|\chi_{\tau}|^2 - (\chi_{\tau} \tilde{f}_{\tau} f_{\tau}^\dagger + c.c.)],$$

(B6)

$$S_{hs}[\eta] = \epsilon \sum_{\tau} J(2|\eta_{\tau}|^2 + (\eta_{\tau} \tilde{f}_{\tau} f_{\tau}^\dagger - f_{\tau} f_{\tau}^\dagger + c.c.)].$$

(B7)

Here $a_{\tau} = +1$ for bonds along the $x$ direction, and equals $-1$ for bonds along the $y$ direction. Note that $S_{hs}[\chi]$ is the same as before. This decoupling of the spin–spin interaction is standardly used in the $SU(2)$ gauge theory formulations of the $t$-$J$ model. We emphasize though that our formulation has, as we will show, only an $Z_2$ gauge symmetry. We now shift the two Hubbard-Stratonovich terms:

$$\chi_{\tau} \rightarrow \chi_{\tau} - \frac{t}{2} b_{\tau}^\dagger b_{\tau},$$

(B9)

$$\eta_{\tau} \rightarrow \eta_{\tau} - \frac{\Delta}{J} (b_{\tau}^\dagger b_{\tau} + c.c.).$$

(B10)

The shift of $\chi$ is as before, and eliminates the spinon–chargon interaction coming from rewriting the electron hopping term. The shift of $\eta$ eliminates the pairing term. The net spatial part of the action is then

$$S = \epsilon \sum_{\tau} (2J(|\chi_{\tau}|^2 + |\eta_{\tau}|^2) + S_F^0 + S_F + S_F^2, \quad \hspace{1cm}$$

(B11)
\[ S_{cr} = -\epsilon \sum_{\langle rr' \rangle} \left\{ [2 \chi_{rr'} + 2 \Delta (\eta_{rr'} + \eta_{rr'}^*)] b_r^\dagger b_{r'} + \text{c.c.} \right\}, \]  
\hspace{6cm} \text{(B12)}

\[ S_{cr}^1 = -\epsilon \sum_{\langle rr' \rangle} J_{cr} f_r f_{r'} + \text{c.c.}, \]  
\hspace{6cm} \text{(B13)}

\[ S_{cr}^2 = \eta_{rr'} \Delta_{rr'} (f_r f_{r'} - f_{r} f_{r'}) + \text{c.c.} \]  
\hspace{6cm} \text{(B14)}

The shift in \( \eta \) also generates a Cooper pair hopping term \( \cos(2 \phi_r - 2 \phi_{r'}) \) with a negative hopping amplitude of order \( \Delta^2/J \). This is not expected to be important for the issues of fractionalization that we primarily wish to discuss. So we will for the most part drop it.

The \( \chi, \eta \) integrals may be done by saddlepoint—a uniform, real saddlepoint solution \( \langle \chi_{rr'} \rangle = \chi_0, \langle \eta_{rr'} \rangle = \eta_0 \) breaks the \( Z_2 \) gauge symmetry. Parametrizing the fluctuations about it by \( \chi_{rr'} = \chi_0 \sigma_{ij}, \eta_{rr'} = \eta_0 \sigma_{ij} \) as before, we arrive at the Ising gauge theory appropriate for the \( d_{x^2-y^2} \) superconductor.

**APPENDIX C: ISING SELF-DUALITY**

We will now review the self-duality of the \( Z_2 \) gauge theory with matter fields in \( 2+1 \) dimensions. As a limiting case, we recover the duality of the pure \( Z_2 \) gauge theory to the global Ising model. The theory is defined by the lattice action

\[ S[s, \sigma] = S_s + S_\sigma, \]  
\hspace{6cm} \text{(C1)}

\[ S_s = -J \sum_{ij} s_i \sigma_{ij} s_j, \]  
\hspace{6cm} \text{(C2)}

\[ S_\sigma = -K \sum \prod \sigma_{ij}. \]  
\hspace{6cm} \text{(C3)}

The constants \( J, K \) are assumed to be positive. The indices \( i,j \) label the sites of a three-dimensional cubic lattice. It is convenient to first rewrite the \( s_i \sigma_{ij} s_j \) term on each bond using the following identity:

\[ e^{J_s/\sigma_{ij}s_j} = A \sum_{n_{ij}=0,1} \exp[-2J n_{ij}] \]  
\hspace{6cm} \text{(C4)}

Here \( \tanh(J_d) = e^{-2J} \), and \( A = \cosh J \). From now on, we will drop the constant \( A \) since it just contributes to an overall multiplicative constant to the partition function. The \( n_{ij} \) take the values 0,1. Upon using this identity for every bond, and doing the sum over \( s_i \), we get

\[ \exp(-S_s) = \text{Tr}_{s_i} \text{Tr}_{n_{ij}} \prod_i \cos \left( \frac{\pi}{2} (\Delta \cdot \hat{n}) \right) \]  
\hspace{6cm} \text{(C6)}

\[ \times \exp \left( -2J_d \sum n_{ij} + \frac{\pi}{2} n_{ij} (1 - \sigma_i) \right). \]  
\hspace{6cm} \text{(C7)}

Here \( \Delta \cdot \hat{n} \) is the lattice divergence of the link variable \( n \). We now notice that the cosine can be written as

\[ \cos \left( \frac{\pi}{2} (\Delta \cdot \hat{n}) \right) = (-1)^{(\Delta \cdot \hat{n})/2} \delta((-1)^{(\Delta \cdot \hat{n})}, 1), \]  
\hspace{6cm} \text{(C8)}

where \( \delta(m,n) \) is the Kronecker delta function for two integers \( m,n \). The term multiplying the delta function is a total derivative that contributes zero on summing over all sites, we will therefore drop it. Note that the delta function imposes conservation modulo 2 of the link variable \( n_{ij} \) at every site. This conservation can be made more explicit by defining a \( Z_2 \) current \( \alpha \):

\[ \alpha_{ij} = (-1)^{n_{ij}}. \]  
\hspace{6cm} \text{(C9)}

We now solve the current conservation condition by writing the \( Z_2 \) current \( \alpha \) on any link as the flux of a dual \( Z_2 \) gauge field \( \mu \) through the plaquette of the dual lattice pierced by this link:

\[ \alpha_{ij} = (-1)^{n_{ij}} = \prod \mu_{ij}. \]  
\hspace{6cm} \text{(C10)}

The \( \mu_{ij} \) are understood to be defined on the links of the dual lattice, and the plaquette product for the \( \mu \) is around the appropriate plaquette of the dual lattice. Note that this is directly analogous to the standard duality transformation of the \( XY \) model.

We next solve for the \( n_{ij} \) in terms of the \( \mu_{ij} \):

\[ n_{ij} = \frac{1}{2} \left( 1 - \prod \mu_{ij} \right). \]  
\hspace{6cm} \text{(C11)}

The \( n_{ij} \) may now be eliminated from the action in favor of the \( \mu_{ij} \). The result (after dropping overall multiplicative constants) is the following identity:

\[ \sum_{\{ij\}} e^{J \sigma_{ij} \sigma_{ij}} = \sum_{\mu} \exp(-S_{cs}). \]  
\hspace{6cm} \text{(C12)}

\[ S_{cs} = -J_d \sum \prod \mu_{ij}, \]  
\hspace{6cm} \text{(C13)}

\[ S_{cs} = \sum_{\{ij\}} \frac{\pi}{4} \left( 1 - \prod \mu \right) (1 - \sigma_{ij}). \]  
\hspace{6cm} \text{(C14)}

The last term has a structure similar to a Chern–Simons term, but for the group \( Z_2 \). Its exponential is actually invariant under \( \sigma \leftrightarrow \mu \). This can be seen as follows. Write

\[ e^{-S_{cs}} = \prod_{\{ij\}} \left( \prod \mu \right)^{(1-\sigma_{ij}/2)}, \]  
\hspace{6cm} \text{(C15)}

\[ = \prod_{\{ij\}} e^{i(\pi/4) \Sigma_{(ij)} (1-\mu)(1-\sigma_{ij})}. \]  
\hspace{6cm} \text{(C16)}

In the last equation, \( \Delta \times \mu \) is the lattice curl of \( \mu \) on the plaquette of the dual lattice pierced by \( \langle ij \rangle \). If we now perform a lattice integration by parts, we get
\[
\exp \sum_{(ij)} -i \frac{\pi}{4} (1 - \mu_{ij}) [\Delta \times (1 - \sigma)]
\]
\[
= \exp \left[ - \sum_{(ij)} i \frac{\pi}{4} \left( 1 - \prod_{\square} \sigma \right) (1 - \mu_{ij}) \right],
\]
where now the sum is over links \((ij)\) of the dual lattice.

The full partition function can then be written as
\[
Z = Tr_{\sigma,\mu} \exp(-S_\sigma - S_\mu - S_{CS}).
\]

The duality of the full action is now apparent. In particular, the action is invariant under the exchange \(\sigma \rightarrow \mu, J_d \rightarrow k\).

To make the duality even more explicit, we again use the identity, Eq. (C12), to write
\[
\sum_{\sigma} \exp(-S_{\sigma} - S_{CS}) = \sum_{v_i} \exp \left( K_d \sum_j v_i \mu_{ij} v_j \right),
\]
where \(v_i = \pm 1\) and \(\tan \phi v_i = e^{-2k}\). The partition function now becomes
\[
Z = Tr_{\tau,\mu} e^{K_d \sum_i \mu_{ij} v_i + J_d \sum \square} (1 - \mu_{ij}),
\]
which is exactly of the same form as in terms of the original variables \((s, \sigma)\), but with the dual couplings \((J_d, K_d)\), thus establishing the self-duality of the theory.

As a special case, consider the limit when \(J = 0\). Then the action in Eq. (C1) is that of the pure \(Z_2\) gauge theory. Under the duality transformation, we now get the form of Eq. (C21) but with the dual coupling \(J_d = \infty\). This means that the fluctuations of the dual gauge field \(\mu\) are frozen; we may choose a gauge in which \(\mu_{ij} = 1\) on every link. The dual action then simply reduces to that of a global Ising model for the \(v_i\) with the dual coupling \(K_d\).

**APPENDIX D: DUALITY OF THE MODEL WITH COMBINED \(U(1)\) AND \(Z_2\) INVARIANCES**

We will now perform a duality transformation on the chargon–spinon action \(S = S_c + S_\mu + S_{\phi}\) derived in Sec. II to work instead with vortex variables instead of the chargons. For simplicity, we will restrict ourselves to situations with an integer number of electrons per unit cell. In this case, the Berry phase term \(S_{\phi}\) is independent of the chargon phase field \(\phi\). In Sec. VI, we provided the generalization necessary to handle a noninteger number of electrons per unit cell. All of our transformations focused entirely on the term in the action involving the chargon variables. This is simply a chargon hopping term:

\[
S_c = - \sum_{(ij)} \tau_{ij} (t_c b_i^* b_j + \text{c.c.}),
\]
\[
= - \sum_{(ij)} 2t_c \cos (\phi_i - \phi_j + \pi/2 (1 - \sigma_{ij})).
\]

Note that in the absence of \(\sigma_{ij}\), this is just the action for the three-dimensional \(XY\) model. The duality transformation for the 3D \(XY\) model is standard; here we will generalize it to include the \(Z_2\) gauge field \(\sigma_{ij}\).

Consider the partition function obtained by integrating over the chargon fields in the above action:
\[
Z_{hol}[\sigma] = \int_{0}^{2\pi} \prod_{i} d\phi_i e^{-S_c}.
\]
As with the duality transformation of the \(XY\) model, it will be convenient to work with the Villain form of the action
\[
S[\phi, J, \sigma] = \sum_{(ij)} \kappa J_{ij}^2/2 + iJ_{ij} \left( \phi_i - \phi_j + \pi/2 (1 - \sigma_{ij}) \right).
\]

where \(J_{ij}\) are integer valued fields that live on the links of the lattice, and are to be summed over in the partition function.

As usual, this is strictly justified only in the limit \(t_c \ll 1\) when \(t_c = \exp(-\kappa/2)\), although we do not expect any modifications to the physics by relaxing this assumption.

The quantity \(\tilde{a}\) lives on the links of the dual lattice, and is constrained to be \(2\pi\) times an integer. The right-hand side is the lattice curl of this variable \(\tilde{a}\) on the plaquette of the dual lattice pierced by the link \((ij)\). The chargon action now takes the form
\[
S[a, \sigma] = \sum_{\square} \frac{\kappa}{8\pi^2} (\Delta \times a)^2 + \frac{i}{4} \sum_{(ij)} (\Delta \times a) (1 - \sigma_{ij}).
\]

Here the first term is a sum over plaquettes of the dual lattice, and the lattice curl in the second term is on the plaquette pierced by the link \((ij)\). Now note that as \(\sigma_{ij} = \pm 1\), the exponential of the second term can be written
\[
\prod_{(ij)} (-1)^{(\Delta \times a/2\pi)(1 - \sigma_{ij}/2)},
\]
It is useful now to separate the integer \(a/2\pi\) into its even and odd part by writing
\[
a = 2\pi (2A + s),
\]
where \(A\) is an integer and \(s = 0, 1\). Then, we have
\[
\prod_{(ij)} (-1)^{(\Delta \times a/2\pi)(1 - \sigma_{ij}/2)} = \prod_{(ij)} \left( \prod_{\square} (-1)^s \right)^{(1 - \sigma_{ij}/2)},
\]
where the product inside the brackets denotes the product over the links of the plaquette of the dual lattice pierced by \((ij)\). We now define
\[
\mu_{ij} = (-1)^s = 1 - 2s.
\]
Note that $\mu_{ij}$ lives on the links of the dual lattice and takes values $\pm 1$. The product above can then be written

$$\exp \left[ i \frac{\pi}{4} \left( 1 - \prod \mu \right) (1 - \sigma_{ij}) \right]. \quad (D11)$$

Note that $\mu$ satisfies

$$\prod \mu = (-1)^{ij}, \quad (D12)$$

where the plaquette product on the left-hand side is on the plaquette of the dual lattice penetrated by the link $(ij)$. Thus, the conserved $Z_2$ charge current determines the flux of $\mu$.

The action now becomes

$$S = \sum \frac{\kappa}{8 \pi^2} \left( \Delta \times \left( 2A + \frac{1 - \mu}{2} \right) \right)^2 + S_{CS}, \quad (D13)$$

$$S_{CS} = i \sum \frac{\pi}{4} \prod \mu (1 - \sigma_{ij}). \quad (D14)$$

At this stage, $A$ is constrained to be integer valued. We impose this integer constraint on $A$ softly by adding a term

$$- t_v \sum \mu_{ij} \cos(2 \pi A_{ij}). \quad (D15)$$

Here the sum is over the links of the dual lattice. The action can now be rewritten in terms of $a = 2 \pi \left[ 2A + (1 - \mu)/2 \right]$: 

$$S = S_v + S_a + S_{CS}, \quad (D16)$$

$$S_v = - t_v \sum \mu_{ij} \cos \left( \frac{a_{ij}}{2} \right), \quad (D17)$$

$$S_a = \sum \frac{\kappa}{8 \pi^2} \left( \Delta \times a \right)^2. \quad (D18)$$

It is convenient to extract a “matter field” from the $a_{ij}$ by letting

$$a_{ij} \rightarrow a_{ij} + 2(\theta_i - \theta_j). \quad (D19)$$

This changes $S_v$ to

$$S_v = - t_v \sum \mu_{ij} \cos \left( \theta_i - \theta_j + \frac{a_{ij}}{2} \right), \quad (D20)$$

but leaves all the other terms unchanged. The field $e^{i\theta_i}$ may be interpreted as an $hc/2e$ vortex creation operator. Several symmetries of the action above are apparent. It is invariant under a local $U(1)$ gauge transformation,

$$\theta_i \rightarrow \theta_i + \Lambda_i, \quad (D21)$$

$$a_{ij} \rightarrow a_{ij} - \frac{\Lambda_i - \Lambda_j}{2}. \quad (D22)$$

This is standard in the dual vortex description of $XY$ models in three dimensions. However the action has an additional $Z_2$ gauge symmetry under which

$$\epsilon_{ij} \rightarrow \epsilon_{ij} e^{i\theta_i}, \quad (D23)$$

$$\mu_{ij} \rightarrow \epsilon_{ij} \mu_{ij}. \quad (D24)$$

with $\epsilon_i = \pm 1$. This $Z_2$ gauge symmetry is actually dual to the one in the chargon-spinon action. Note that the action describes the vortices $e^{i\theta_i}$ minimally coupled to the fluctuating $U(1)$ gauge field $a$, and also to the fluctuating $Z_2$ gauge field $\mu$. The field $\mu$ is in turn coupled to the field $\sigma$ by the term $S_{CS}$.

This completes the duality transformation to the vortex description. Adding together the spinon action and the Berry phase term $S_B$ gives the full dual action of Sec. V.

---


2. See, for example, Perspectives in Quantum Hall Effects, edited by S. DasSarma and A. Pinczuk (Wiley, New York, 1997).


11. For an accessible review of lattice gauge theory, see, for example, J. Kogut, Rev. Mod. Phys. 51, 659 (1979).


17. S. Sachdev, private communication; S. Sachdev and Matthias Vojta, cond-mat/9910231.


23 D.H. Lee, cond-mat/9909111; see also Y.B. Kim and Z. Wang, cond-mat/9901003.


27 We emphasize, in particular, that despite this choice of $e$ the full theory is not isotropic in space-time.

28 Integrating out the chargons will also generate other interactions between gauge fields on links that belong to different plaquettes. However, in the insulating phase, these will be short ranged.


32 For some representative papers see, for example, A. Polyakov, Phys. Lett. B 72, 477 (1978); B. Svetitsky and L. Yaffe, Nucl. Phys. B 210, 423 (1982); for a review, see, for example, B. Svetitsky, Phys. Rep. 132, 1 (1986).


35 For a review, see, for example, M. Randeria and J.-C. Campuzano, Varenna Lectures, cond-mat/9709107.