Spin density wave order, topological order, and Fermi surface reconstruction

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1103/PhysRevB.94.115147</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33973828">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33973828</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP</a></td>
</tr>
</tbody>
</table>
Spin density wave order, topological order,
and Fermi surface reconstruction

Subir Sachdev,1,2 Erez Berg,3 Shubhayu Chatterjee,1 and Yoni Schattner3

1Department of Physics, Harvard University, Cambridge MA 02138, USA
2Perimeter Institute for Theoretical Physics,
Waterloo, Ontario, Canada N2L 2Y5
3Department of Condensed Matter Physics,
The Weizmann Institute of Science, Rehovot, 76100, Israel
(Dated: September 22, 2016)

Abstract

In the conventional theory of density wave ordering in metals, the onset of spin density wave (SDW) order co-incides with the reconstruction of the Fermi surfaces into small ‘pockets’. We present models which display this transition, while also displaying an alternative route between these phases via an intermediate phase with topological order, no broken symmetry, and pocket Fermi surfaces. The models involve coupling emergent gauge fields to a fractionalized SDW order, but retain the canonical electron operator in the underlying Hamiltonian. We establish an intimate connection between the suppression of certain defects in the SDW order, and the presence of Fermi surface sizes distinct from the Luttinger value in Fermi liquids. We discuss the relevance of such models to the physics of the hole-doped cuprates near optimal doping.
I. INTRODUCTION

A number of recent experiments [1–4] have highlighted a remarkable transformation in the electronic state of the hole-doped cuprates at a hole density around \( p = p_c \approx 0.19 \): many electronic properties change from those characteristic of a Fermi gas of charge \(+e\) carriers of density \( p \) for \( p < p_c \), to those of a Fermi gas of charge \(+e\) carriers of density \( 1 + p \) for \( p > p_c \). As the density of holes is conventionally measured relative to those of the insulator at unit density, a conventional Fermi liquid is required by the Luttinger theorem to have a Fermi surface of size \( 1 + p \), as found for \( p > p_c \).

Starting from the Fermi liquid with a Fermi surface of size \( 1 + p \), there are two reasonable routes to a Fermi surface reconstruction of size \( p \) that could apply to the cuprates:

(i) The conventional route involves the onset of spin density wave (SDW) order (other density wave orders have also been suggested [5]), which reconstructs the “large” Fermi surface to pocket Fermi surfaces. This route appears appropriate for the electron-doped cuprates, where antiferromagnetic order is observed [6] not too far from the critical electron doping.

(ii) The more ‘exotic’ route relies on the development of topological quantum order in the metallic state, which has been linked to changes in the Fermi surface size [7–9]. This is an attractive and exciting possibility for the hole-doped cuprates, given the absence in observations so far of significant correlations in any order parameter which breaks translational symmetry near \( p = p_c \).

The purpose of this paper is to present the simplest models in which the existence of the three metallic phases mentioned above (the Fermi liquid, the Fermi liquid with SDW order, and the metal with small Fermi surfaces and topological order) can be reliably established. We wish to describe models which can serve as convenient starting points for analyzing the quantum phase transitions between these metals, and which do not have extraneous exotic phases which are ultimately unstable to confinement.

The models described below are closely connected to previous work [10–13] using a SU(2) gauge theory and a Higgs field to represent local antiferromagnetic correlations in a metal. These previous works, along with related works using a Schwinger boson formulation [14, 15] or a quantum dimer model [16], show that the phases we obtain below are allowed ground states of a single-band Hubbard model. However, in the interests of simplicity and of keeping this paper self-contained, we will not introduce the models using these prior connections. Instead, we will emphasize the relationship of our models to those using a conventional Landau-Ginzburg-Wilson (LGW) order parameter framework for the onset of spin density wave order in metals; as emphasized by Hertz [17], the general rules of LGW theories apply to such quantum phase transitions in metals, and the main role of the Fermi surface is to damp with dynamic order parameter fluctuations. Developing our model by deforming the LGW theory will clearly expose the intimate link between topological
defects in the SDW order, and the possibility of metallic states which have Fermi surface sizes distinct from the Luttinger value.

As the LGW-Hertz theory is an expansion around weak coupling, our analysis below can be viewed as providing the minimal ingredients necessary to put strong coupling Mott-Hubbard physics back into the LGW-Hertz model. Indeed in the limit of $p = 0$, our model will have a Mott insulator with $\mathbb{Z}_2$ topological order [18, 19] (and other topological orders in related models), in addition to the ‘Slater’ insulator with Néel order present in the LGW-Hertz model.

A. Easy-plane model

The most transparent introduction to our models is obtained by focusing on the case in which the spin density wave order parameter is restricted to lie in the $x$-$y$ plane in spin space. Such a restriction can only arise from spin-orbit couplings, which are known to be rather weak in the cuprates. Nevertheless, we will describe this case first because of its simplicity.

1. LGW-Hertz theory

The LGW-Hertz theory for the onset of SDW order can be described by the following Hamiltonian

$$H_{\text{sdw}} = H_c + H_\theta + H_Y,$$

where $H_c$ describes electrons (of density $(1-p)$) hopping on the sites of a square lattice

$$H_c = -\sum_{i,j} (t_{ij} + \mu \delta_{ij}) c^\dagger_i c_j$$

with $c_{\alpha}$ the electron annihilation operator on site $i$ with spin $\alpha = \uparrow, \downarrow$. We represent the SDW order by a lattice XY rotor model, described by an angle $\theta_i$, and its canonically conjugate number operator $N_i$, obeying

$$H_\theta = -\sum_{i<j} J_{ij} \cos(\theta_i - \theta_j) + 4\Delta \sum_i N_i^2 \ ; \ \ [\theta_i, N_j] = i\delta_{ij},$$

where $J_{ij}$ positive exchange constants, and $\Delta$ is proportional to the bare spin-wave gap (the 4 is for future convenience). A term linear in $N_i$ is also allowed in $H_\theta$, but we ignore it for simplicity; such a linear term will not be allowed when we consider models with SU(2) symmetry in Section IV. Finally, there is a ‘Yukawa’ coupling between the XY order parameter, $e^{i\theta}$, and the fermions

$$H_Y = -\lambda \sum_i \eta_i \left[ e^{-i\theta} c^\dagger_{i\uparrow} c_{i\downarrow} + e^{i\theta} c^\dagger_{i\downarrow} c_{i\uparrow} \right],$$

(1.4)
where
\[ \eta_i \equiv (-1)^{x_i+y_i} \quad (1.5) \]
is the staggering factor representing the opposite spin orientations on the two sublattices. Note that the Yukawa coupling, and the remaining Hamiltonian, commute with the total spin along the \( z \) direction
\[ S_z = \sum_i \left( N_i + \frac{1}{2} c_i^\dagger c_i^\dagger - \frac{1}{2} c_i^\dagger c_i \right). \quad (1.6) \]

The Hamiltonian \( H_{\text{sdw}} \) displays the two conventional metallic phases noted above. These phases can be conveniently accessed by tuning the value of \( \Delta/J \), where \( J \) is the nearest-neighbor exchange. For large \( \Delta/J \), the correlations of \( e^{i\theta} \) are short-ranged, and we obtain the Fermi liquid with a large Fermi surface controlled mainly by \( H_c \); we can account for \( H_Y \) perturbatively in \( \lambda \), and the large Fermi surface leads to damping in the order parameter correlation functions. On the other hand, for small \( \Delta/J \), we expect long-range XY order with \( \langle e^{i\theta} \rangle \neq 0 \); now \( H_Y \) has a stronger effect, and in the presence of the XY condensate the fermion dispersion is modified, leading to a reconstruction of the Fermi surface into small pockets. These phases of \( H_{\text{sdw}} \) are illustrated by the small \( K \) regime of Fig. 1, where they are labelled A and B respectively (the parameter \( K \) will be introduced below). The phase transition between these two phases has been extensively studied \([20–27]\) since the original work by Hertz \([17]\), including by recent sign-problem-free quantum Monte Carlo simulations \([28–30]\). Note that in such a phase transition, two important physical changes happen at the same point in the phase diagram: the appearance of long-range XY order, and the reconstruction of the Fermi surface.

It is also interesting to consider the \( p = 0 \) limit of phases A and B. The large Fermi surface has size \( 1 + p \) and so phase A is not sensitive to \( p \) approaching 0: it remains a Fermi liquid. On the other hand, in phase B, the hole pockets disappear at \( p = 0 \), so phase B is an insulator. This insulating behavior is a direct consequence of the presence of strong long-range XY order, and so B should be considered a Slater insulator at \( p = 0 \). We note that in between the Slater insulator and the large Fermi surface Fermi liquid, there is a metal with hole and electron pockets, and this is not shown in Fig. 1.

2. Fractionalizing the order parameter

We now ask if it is possible, at non-zero \( p \), to realize a situation in which XY long-range order and Fermi surface reconstructions happen at distinct points of the phase diagram. If so, we will obtain an intermediate phase with small Fermi pockets but no long-range XY order. (At \( p = 0 \), such a phase would be an insulator without long-range XY order, and so would be a Mott insulator.) To obtain such a phase, we use the idea of transforming to a ‘rotating reference frame’
FIG. 1. Schematic, minimal, phase diagram of the easy-plane Hamiltonian $H_1$ in Eq. (1.9). The vortices are the usual defects in the XY SDW order $e^{i\theta}$. The Fermi surfaces are shown in the first Brillouin zone: those in A and B are of electrons, while those in C can be either of electrons or chargons. In phase C, the single vortices in the SDW order are gapped excitations, identified as the visons of the $\mathbb{Z}_2$ topological order. The sketched Fermi surfaces are for hole-doping with the cuprate band structure: in phases B and C only hole pockets are shown, but electron pockets will appear near the boundaries to phase A. We propose that the SU(2) spin rotation invariant analogs of phase C (discussed in Section IV) describe the pseudogap state in the hole-doped cuprates. Other less-correlated superconductors (such as the pnictides) are proposed to bypass phase C and evolve directly from phase B to A.

determined by the local orientation of the XY order [10, 31, 32]. In particular, by a rotation about the $z$ axis in spin space, let us define the canonical fermion operators

$$
\psi_+ = e^{i\theta/2} c^\dagger_+ , \quad \psi_- = e^{-i\theta/2} c^\dagger_- .
$$

(1.7)

Then the Yukawa coupling, $H_Y$, takes a simple form independent of the orientation of the XY order [10]:

$$
H_Y = -\lambda \sum_i \eta_i \left[ \psi^\dagger_{i+} \psi_{i-} + \psi^\dagger_{i-} \psi_{i+} \right] .
$$

(1.8)
In other words, the $\psi_{\pm}$ fermions move in the presence of a spacetime-independent XY order, even though the actual orientation of the XY order rotates from point to point. Moreover, from the electron hopping term in $H_c$, we can obtain an effective hopping $Z_{ij}t_{ij}(\psi_{i+}^\dagger\psi_{j+} + \psi_{i-}^\dagger\psi_{j-})$ where $Z_{ij} = \langle e^{\pm i(\theta_i - \theta_j)/2} \rangle$ is a renormalization factor of order unity (computed later in Eq. (2.21)). So it appears we can realize a situation in which the $\psi_{\pm}$ fermions are approximately free, and their observation of constant XY order implies that they will form small pocket Fermi surfaces (or be fully gapped at $p = 0$). From (1.6), it can be verified that the $\psi_{\pm}$ fermions have $S_z = 0$, and so these are spinless fermions which carry only the charge of the electron: we will refer to them as 'chargons' in the remaining discussion. A metallic phase with chargon Fermi surfaces was called an 'algebraic charge liquid' (ACL) in Ref. 15.

However, further thought based upon the structure of (1.7) shows that there is a crucial obstacle to realizing nearly-free $\psi_{\pm}$ fermions in a regime where $\langle e^{i\theta} \rangle = 0$. The phase with no XY order described by $H_{sdw}$ has proliferating $2\pi$ vortices in the SDW order, and the half-angle transformation in Eq. (1.7) shows that $\psi_{\pm}$ are not single-valued around such vortices. So the $\psi_{\pm}$ fermions must be confined at the same point where the XY order disappears. In other words, we are back to the conventional scenario in which Fermi surface reconstruction and XY ordering co-incide.

But the above argument also suggests a route around such an obstacle: the $\psi_{\pm}$ fermions are single-valued around doubled $4\pi$ vortices, and so we need the disappearance of XY order to be associated with the proliferation of doubled vortices.

There is a simple route to the loss of XY order by doubled vortices that has been much studied in the literature [33–36]: it involves coupling the square-root of the XY order, the ‘spinon’ field $e^{i\theta/2}$, to a $\mathbb{Z}_2$ gauge field. A microscopic justification for fractionalizing the order parameter in this manner can be obtained from the Schwinger boson theory of frustrated antiferromagnets [10, 18, 37, 38]: $e^{i\theta/2}$ is essentially the staggered Schwinger boson operator. The model we wish to study is obtained by replacing $H_\theta$ in $H_{sdw}$ by the model studied in Refs. 33 and 34. In this manner, we obtain the Hamiltonian (written out completely because of our focus on it in this paper)

\[
H_1 = H_c + H_{\theta,\mathbb{Z}_2} + H_Y
\]

\[
H_c = -\sum_{i,j} (t_{ij} + \mu\delta_{ij}) c_{i\alpha}^\dagger c_{j\alpha}
\]

\[
H_Y = -\lambda \sum_i \eta_i \left[ e^{-i\theta_i} c_{i\uparrow}^\dagger c_{i\downarrow} + e^{i\theta_i} c_{i\downarrow}^\dagger c_{i\uparrow} \right]
\]

\[
H_{\theta,\mathbb{Z}_2} = -\sum_{i<j} J_{ij}\mu_{ij}^x \cos ((\theta_i - \theta_j)/2) + 4\Delta \sum_i N_i^2 - g \sum_{\langle ij \rangle} \mu_{ij}^x - K \sum_i \left[ \prod_{\mu_{ij}} \mu_{ij}^\mu \right],
\]  

(1.9)

where $\mu^{x,z}$ are Pauli matrices on the links of the square lattice representing the $\mathbb{Z}_2$ gauge field. The forms of $H_c$ and $H_Y$ are the same as those in the LGW-Hertz theory, and only the action for
the spin density wave order has been modified by terms that are effectively multi-spin exchange interactions. (It will become clear from our discussion later that at $p = 0$ and small $\Delta$, $\mathcal{H}_1$ reduces to the model studied in Ref. 36.) The Hamiltonian $\mathcal{H}_1$ is invariant under the $\mathbb{Z}_2$ gauge transformation

$$e^{i\theta_i/2} \rightarrow s_i e^{i\theta_i/2}, \quad \mu_{ij}^z \rightarrow s_i \mu_{ij}^z s_j,$$

(1.10)

where $s_i = \pm 1$ is an arbitrary function of $i$, and the other operators remain invariant. Associated with this gauge invariance is the existence of an extensive number of conserved charges, $\hat{G}_i$, which commute with $\mathcal{H}_1$ and obey $\hat{G}_i^2 = 1$; we restrict our attention to the gauge-invariant sector of the Hilbert space in which all the $\hat{G}_i = 1$:

$$\hat{G}_i \equiv e^{2i\pi N_i} \prod_{j \in \text{n.n.}(i)} \mu_{ij}^x = 1,$$

(1.11)

where $j$ extends over the nearest-neighbors of $i$.

The main term driving the appearance of exotic phases in $\mathcal{H}_1$ is the $K$ term, which penalizes configurations with non-zero $\mathbb{Z}_2$ gauge flux. For small $K$, we can trace over the $\mathbb{Z}_2$ gauge field in powers of $K$, and then $\mathcal{H}_1$ only has terms with the same structure as those in $H_{\text{sdw}}$. However, at large $K$, the suppression of $\mathbb{Z}_2$ gauge flux implies that single vortices (but not double vortices) in $e^{i\theta}$ become very expensive: the coupling $J_{ij}$ ties $\mathbb{Z}_2$ gauge flux to a $2\pi$ vortex in $e^{i\theta}$ because of the branch cut in $e^{i\theta/2}$ around such a vortex. Hence, upon increasing $\Delta$ at large $K$, we obtain the needed transition to a phase without XY order by the proliferation of double vortices. The resulting $\mathbb{Z}_2$ topologically ordered phase supports gapped deconfined “spinon” excitations that carry a half integer value of $N_i$, and gapped vison excitations which are the remnants of the single vortices in the SDW ordered phase.

This discussion therefore leads to the schematic phase diagram of $\mathcal{H}_1$ shown in Fig. 1. Further details on the structure of the new phase with $\mathbb{Z}_2$ topological order appear in Sections II and III. The model (1.9) can support additional phases not shown in Fig. 1, which we comment on in Appendix A.

One important distinction between $\mathcal{H}_1$ and previous studies of fractionalization in doped Mott insulators [10, 33, 39, 40] is worth noting here. In all previous works, the Hamiltonian is presented in terms of emergent, fractionalized spinon and chargon degrees of freedom. The electron operator does not appear explicitly in the Hamiltonian, but is described as a composite operator. In contrast, in our model $\mathcal{H}_1$ we have fractionalized the order parameter only into the spinons, while retaining the bare electron operator in the Hamiltonian. In our approach, it is the chargon, rather than the electron, which appears as a composite particle, as a bound state of the electron and the spinon in (1.7). We believe this difference in perspective is important, and that it leads to an efficient and controlled description of the metallic states observed in the cuprates.
We also comment here on an important subtlety in the structure of the metallic phase with \( \mathbb{Z}_2 \) topological order. We have given arguments above on the appearance of reconstructed pocket Fermi surfaces of the \( \psi_\pm \) chargons in this phase. However, as we will see in our computations below, there remains a strong residual attraction \([14-16, 41]\) between the chargons and the spinons due to the hopping terms, \( t_{ij} \) in \( H_c \). Because of this attraction, it is possible that some or all of the chargons form bound states with the spinons, leading to a pocket Fermi surface of electron-like quasiparticles with charge \( e \) and spin \( S_z = \pm 1/2 \). If all of the chargons undergo this bound-state formation, then we obtain a FL* metal \([7, 16, 42-46]\). Microscopic details of the Hamiltonian will determined whether we obtain an ACL, or FL*, or an intermediate phase with co-existing chargon and electron Fermi surfaces \([15]\)—the charge transport properties of all these phases are expected to be very similar, and so we will not focus much on the distinction here. Note that, in the discussion above, even if all chargons bind with spinons to form electrons, we do not obtain back the large Fermi surface Fermi liquid, but obtain FL*: this is because the Fermi surface size of the chargons is \( p \) and not \( 1 + p \). Thus, an important feature of our analysis is that the operations of binding fermions to spinons, and of Fermi surface reconstruction, do not commute.

We note that a model closely related to \( H_1 \) was studied by Grover and Senthil \([47]\) in the context of a two-band Kondo-Heisenberg model. However, their interest was limited to the regime accessible perturbatively in an expansion in \( \lambda \) (in our notation), and they did not obtain the analog of the topological phase C in Fig. 1 with reconstructed Fermi surfaces. In doped Mott insulators, the coupling \( \lambda \sim U \), the on-site Mott-Hubbard repulsion, and so \( \lambda \) is the largest energy scale in the Hamiltonian. We will work throughout in the large \( \lambda \) limit, and will see below that the topological phase appears in a regime inaccessible in a small \( \lambda \) expansion.

The analysis of this paper will neglect superconductivity. But it should be noted that all the metallic phases of Fig. 1 are expected to superconduct at low \( T \) \([29, 30]\).

The outline of the remainder of the paper is as follows. Section II will analyze the properties of phase C of the easy-plane model \( H_1 \) in a strong coupling expansion. Section III will discuss the topological order and dynamics of visons in \( H_1 \). We will generalize the results to various cases in models with full SU(2) spin rotation invariance in Section IV. Section V will summarize our results and discuss the nature of the phase transitions in Fig. 1.

II. STRONG COUPLING EXPANSION WITH TOPOLOGICAL ORDER

This section will present a strong coupling analysis of the topological state C of the easy-plane model \( H_1 \) in Fig. 1. The conventional metals A and B have the same properties as those of the phases of the SDW theory \( H_{\text{sdw}} \), and so don’t need further discussion here. Our analysis will establish the stability of a metallic phase with topological order, and also determine its excitation
spectrum in a limiting regime.

As the conventional phases appear in the limit of small $K$, we will study here the complementary $K \to \infty$ regime. At $K = \infty$, we can work in the gauge $\mu_{ij}^z = 1$ everywhere. So the model of interest in Eq. (1.9) reduces to the gauge-fixed Hamiltonian

$$
\mathcal{H}_1' = -\sum_{i,j} (t_{ij} + \mu \delta_{ij} \alpha_i \alpha_j - \lambda \sum_i \eta_i \left[ e^{-i\theta_i} \gamma_i \gamma_i + e^{i\theta_i} \gamma_i \gamma_i \right] - \sum_{i<j} J_{ij} \cos ((\theta_i - \theta_j)/2) + 4\Delta \sum_i N_i^2, (2.1)
$$

describing fermions coupled to a XY rotor model. Note the crucial and only difference from the conventional SDW theory $H_{sdw}$ in Eq. (1.4): the $J_{ij}$ terms now involve couplings between the spinon field $e^{i\theta/2}$, rather than the XY order parameter $e^{i\theta}$. Consequently, the rotor states on each site have $N_i$ quantized in steps of 1/2, and the spin $S_z$ can be half-integer or integer.

This section will describe a strong-coupling analysis of $\mathcal{H}_1'$ in which the on-site terms are much larger than the off-site terms i.e.

$$
\lambda, \Delta \gg |t_{ij}|, |J_{ij}|. (2.2)
$$

We will show that in this limit at $p = 0$, $\mathcal{H}_1'$ realizes a Mott insulator with a spin liquid ground state. Furthermore, the spin liquid has odd $\mathbb{Z}_2$ topological order with gapped bosonic spinon and fermionic chargon excitations: ‘odd’ refers to the presence of unit $\mathbb{Z}_2$ background charge on each site of the Mott insulator [33, 48–50]; the vison excitations have infinite energy at $K = \infty$, and will be considered further in Section III. Moving towards $p > 0$, we occupy the lowest energy fermionic holon states and obtain a ACL metal with odd $\mathbb{Z}_2$ topological order.

We will also describe the spectrum of states with one chargon and one spinon excitation above the Mott insulator. Note that these states have total charge $e$ and spin $S_z = 1/2$, and so have the same quantum numbers as the electron. We will find that along with the scattering states in which the holon and spinon are well separated from each other, there is an electron-like bound state below the scattering continuum whose dispersion we shall compute. Because of the large spinon gap in the present strong coupling limit, this holon-spinon bound state is well above the band of fermionic holon states. However, away from the strong-coupling limit it is clearly possible that the bound state becomes the lowest energy charged fermionic state [14, 15]. Then, at $p > 0$, these states will be occupied, leading to Fermi surfaces with electron-like quasiparticles. Such Fermi surfaces can co-exist with a holon Fermi surfaces (leading to the ‘holon-hole’ metal of Ref. 15), or they can exist by themselves in a $\mathbb{Z}_2$-FL* state.
A. Single-site eigenstates

We begin the study of $H'_{1}$ in the limit (2.2) by defining the on-site Hamiltonian

$$H_{0} = \sum_{i} \left(-\mu c_{ia}^{\dagger} c_{ia} - \lambda \eta_{i} \left( e^{-i\theta} c_{i\uparrow}^{\dagger} c_{i\downarrow} + e^{i\theta} c_{i\downarrow}^{\dagger} c_{i\uparrow} \right) + 4\Delta \sum_{i} N_{i}^{2} \right) \tag{2.3}$$

It is easy to determine all the eigenstates of $H_{0}$. In this subsection, we drop the site index, $i$. We denote the state with $N = 0$ (no spinons) and no electrons as $|0\rangle$. Then the eigenstate with $n$ spinons is

$$\hat{N} e^{in\theta/2} |0\rangle = \frac{n}{2} e^{in\theta/2} |0\rangle. \tag{2.4}$$

The empty electron state is implicit in $|0\rangle$, and all electrons will be indicated below by creation operators acting on $|0\rangle$. Notice that the $H_{0}$ conserves $n$ modulo 2, and states with $n$ odd carry the $\mathbb{Z}_{2}$ gauge charge. The state $e^{in\theta/2} |0\rangle$ carries spin $S_{z} = n/2$. Important low-lying states are:

(i) **Mott insulator**

This is the state

$$|G\rangle = \frac{1}{\sqrt{2}} \left( c_{\uparrow}^{\dagger} e^{-i\theta/2} + \eta c_{\downarrow}^{\dagger} e^{i\theta/2} \right) |0\rangle \quad , \quad E_{G} = -\lambda + \Delta - \mu. \tag{2.5}$$

This state carries no total spin, $S_{z} = 0$. It also carries $\mathbb{Z}_{2}$ gauge charge, and so the $\mathbb{Z}_{2}$ gauge theory is ‘odd’ [33, 48–50], provided $E_{G}$ is the lowest energy state at half filling.

(ii) **Spinons**

These are doubly-degenerate states with $S_{z} = \pm 1/2$. The $S_{z} = +1/2$ state is

$$|\uparrow\rangle = \left( a c_{\uparrow}^{\dagger} + \eta b c_{\downarrow}^{\dagger} e^{2i\theta/2} \right) |0\rangle \quad ; \quad E_{s} = -\mu + 2\Delta - \sqrt{\lambda^{2} + 4\Delta^{2}}, \tag{2.6}$$

where $(a, b)$ is an eigenvector of the matrix

$$\begin{pmatrix} 0 & -\lambda \\ -\lambda & 4\Delta \end{pmatrix} \tag{2.7}$$

with eigenvalue $2\Delta - \sqrt{\lambda^{2} + 4\Delta^{2}}$, and similarly

$$|\downarrow\rangle = \left( a c_{\downarrow}^{\dagger} + \eta b c_{\uparrow}^{\dagger} e^{-2i\theta/2} \right) |0\rangle \quad ; \quad E_{s} = -\mu + 2\Delta - \sqrt{\lambda^{2} + 4\Delta^{2}}. \tag{2.8}$$

Relative to the Mott insulator, the spinon carries no electromagnetic charge, and a $\mathbb{Z}_{2}$ gauge charge. We want to be in a regime where the Mott insulator has a lower energy than the spinon, and so we require

$$E_{G} < E_{s} \quad \Rightarrow \quad \lambda > 3\Delta/2. \tag{2.9}$$
The spin gap, $\Delta_s$, of the Mott insulator is
\[
\Delta_s = E_s - E_G = \Delta + \lambda - \sqrt{\lambda^2 + 4\Delta^2}. \tag{2.10}
\]

(iii) **Holon**
This is simply the empty state $|0\rangle$, with energy $E_{hn} = 0$. Relative to the Mott insulator, this state has $S_z = 0$, $+e$ electromagnetic charge, and a non-zero $\mathbb{Z}_2$ gauge charge.

(iv) **Doublon**
This is the state $c_+^\dagger c_+^\dagger |0\rangle$, with energy $E_{dn} = -2\mu$. Relative to the Mott insulator, this state has $S_z = 0$, $-e$ electromagnetic charge, and a non-zero $\mathbb{Z}_2$ gauge charge.

(v) **Holes**
These are the doubly degenerate states $e^{i\theta/2} |0\rangle$ with energy $E_h = \Delta$. Relative to the Mott insulator, they carry electromagnetic charge $+e$ and zero $\mathbb{Z}_2$ gauge charge. They have spin $S_z = \pm 1/2$. We can now examine the energy difference between a pair of sites with hole+Mott insulator and a pair with holon+spinon
\[
E_h + E_G - E_{hn} - E_s = -\lambda + \sqrt{\lambda^2 + 4\Delta^2} > 0. \tag{2.11}
\]
So a hole is unstable to decay into a holon and a spinon in the strong-coupling expansion of Eq. (2.2). Note that, at $\lambda \gg \Delta$, this energy difference can become small.

(vi) **Electrons**
These are the doubly degenerate states $e^{\pm i\theta/2} |0\rangle$ with energy $E_e = -2\mu + \Delta$. Relative to the Mott insulator, they carry electromagnetic charge $-e$ and zero $\mathbb{Z}_2$ gauge charge. They have spin $S_z = \pm 1/2$. The condition for the instability of an electron state is
\[
E_e + E_G - E_{dn} - E_s = -\lambda + \sqrt{\lambda^2 + 4\Delta^2} > 0, \tag{2.12}
\]
which is the same as (2.11).

For subsequent analysis, it is useful to introduce the canonical fermion operators of the holon (this is a linear combination of the operators in (1.7))
\[
\psi = \frac{1}{\sqrt{2}} \left( e^{i\theta/2} c_+^\dagger + \eta e^{-i\theta/2} c_+ \right) \tag{2.13}
\]
Note that $\psi$ is the holon creation operator, i.e., a holon is an empty state in a filled $\psi$ band,
\[
|G\rangle = \prod_i \psi_i^\dagger |0\rangle, \tag{2.14}
\]
which is the Mott insulator.

We also introduce the fermions
\[
\Phi^\dagger_+ = a c_+^\dagger + \eta b c_+^\dagger e^{i\theta}
\]
\[
\Phi^\dagger_- = a c_-^\dagger + \eta b c_-^\dagger e^{-i\theta} \tag{2.15}
\]
Then the spinon creation operator is the boson

\[ b_\alpha^\dagger = \Phi_\alpha^\dagger \psi \]  

(2.16)

This creates the spinon excitation via \( b_\alpha^\dagger |G\).

**B. Effective holon Hamiltonian**

Now we move beyond the single-site Hamiltonian, and examine the influence of the multisite terms on the single holon excitation above the Mott insulator.

First, we note the on-site Hamiltonian

\[ H_{h0} = \sum_i (\mu + \Delta - \lambda) \psi_i^\dagger \psi_i. \]  

(2.17)

which describes the on-site energy of the holon states.

Next, we include the hopping terms \( t_{ij} \) and \( J_{ij} \). We perform a canonical transformation to eliminate the \( \theta/2 \) excitations to obtain an effective Hamiltonian for the holons. This transformation should be performed around single particle excitations of the band insulator \(|G\). For this transformation, it is convenient go back to the original \( c \) fermion formulation. We make a list of all states among a pair of sites, 1, 2, which are important to second order perturbation theory in \( t, J \) with a total charge of \( e \) and a total \( S_z \) of 0; there turn out to be 12 such states

\[
\begin{align*}
&c_{11}^\dagger e^{-i\theta_1/2} |0\rangle, \quad c_{11}^\dagger e^{i\theta_1/2} |0\rangle, \quad c_{21}^\dagger e^{-i\theta_2/2} |0\rangle, \quad c_{21}^\dagger e^{i\theta_2/2} |0\rangle, \\
&c_{21}^\dagger e^{-i\theta_1/2} |0\rangle, \quad c_{21}^\dagger e^{i\theta_2-i\theta_1/2} |0\rangle, \quad c_{21}^\dagger e^{i\theta_1/2} |0\rangle, \quad c_{21}^\dagger e^{i\theta_2+i\theta_1/2} |0\rangle, \\
&c_{11}^\dagger e^{-i\theta_2/2} |0\rangle, \quad c_{11}^\dagger e^{i\theta_1-i\theta_2/2} |0\rangle, \quad c_{11}^\dagger e^{i\theta_2/2} |0\rangle, \quad c_{11}^\dagger e^{-i\theta_1+i\theta_2/2} |0\rangle
\end{align*}
\]

(2.18)

Each site in all of these states is limited to have a spin of \( S_z = 0, \pm 1/2 \). A conventional computation then eliminates the last 8 of these states to yield the effective holon Hamiltonian.

To leading order in \( J/\lambda \), only hopping within the same sublattice contributes, and the effective holon Hamiltonian turns out to be

\[ H_h = H_{h0} + H_{h1} \]  

(2.19)

with

\[ H_{h1} = - \sum_{i<j, \text{n.n.n.n.}} t_{2} J_{2} Z \left( \psi_i^\dagger \psi_j + \psi_j^\dagger \psi_i \right) - \sum_{i<j, \text{n.n.n.n.}} t_{3} J_{3} Z \left( \psi_i^\dagger \psi_j + \psi_j^\dagger \psi_i \right). \]  

(2.20)

where

\[ Z = \frac{(\lambda + 2\Delta)}{2\Delta \lambda}. \]  

(2.21)
Here, $n.n.n.$ and $n.n.n.n.$ stand for second- and third-neighbor sites, respectively, and the renormalization factor $Z$ is related to the $Z_{ij}$ mentioned in Section IA2. One feature of $H_{h1}$ is that the holons on the two sublattices don’t mix with each other; i.e., hopping between same-sublattice sites on the square lattice is forbidden. This is an exact property of this model due to a symmetry of the XY model: the holon operator in (2.13) is odd or even under the spin inversion $S_z \rightarrow -S_z$, $\theta \rightarrow -\theta$ on the two sublattices.

C. Effective spinon Hamiltonian

Next we examine the hopping of the single spinon excitations above the Mott insulator. In this case, the computation is simpler than the holon case, and the the spinon Hamiltonian is easily obtained by first-order perturbation theory:

$$H_s = \sum_i \left( \Delta + \lambda - \sqrt{\lambda^2 + 4\Delta^2} \right) b_{\alpha}^\dagger b_{\alpha} - \frac{(a + b)^2}{2} \sum_{i<j} J_{ij} \left( b_{\alpha}^\dagger b_{\alpha} + b_{\beta}^\dagger b_{\beta} \right). \quad (2.22)$$

D. Holon and spinon bound state

We are now ready to consider the states with both a holon and a spinon present. The most important coupling between them appears already at first order in $t_{ij}$ when the holon and the spinon exchange positions. The matrix element for this is easily computed and leads to the Hamiltonian

$$H_{hs1} = -\sum_{i<j} a^2 t_{ij} \left( \phi_{\alpha}^\dagger \phi_{\alpha} + \phi_{\beta}^\dagger \phi_{\beta} \right)$$

$$= -\sum_{i<j} a^2 t_{ij} \left( \psi_{\alpha}^\dagger \psi_{\beta} b_{\alpha}^\dagger b_{\beta} + \psi_{\beta}^\dagger \psi_{\alpha} b_{\alpha}^\dagger b_{\beta} \right). \quad (2.23)$$

Introducing the holon operators $h = \psi^\dagger$, and collecting all terms, the Hamiltonian acting on the Hilbert space of one holon and one spinon is

$$H_{hs} = \sum_i \left[ (\mu - \Delta + \lambda) h_{i}^\dagger h_{i} + \left( \Delta + \lambda - \sqrt{\lambda^2 + 4\Delta^2} \right) b_{\alpha}^\dagger b_{\alpha} \right] + \sum_{i<j, n.n.n.} \frac{t_2 J_2 (\lambda + 2\Delta)}{2\Delta \lambda} \left( h_{i}^\dagger h_{j} + h_{j}^\dagger h_{i} \right) + \sum_{i<j, n.n.n.n.} \frac{t_3 J_3 (\lambda + 2\Delta)}{2\Delta \lambda} \left( h_{i}^\dagger h_{j} + h_{j}^\dagger h_{i} \right)$$

$$- \frac{(a + b)^2}{2} \sum_{i<j} J_{ij} \left( b_{\alpha}^\dagger b_{\alpha} + b_{\beta}^\dagger b_{\beta} \right) + \sum_{i<j} a^2 t_{ij} \left( h_{i}^\dagger h_{i} b_{\alpha}^\dagger b_{\beta} + h_{i}^\dagger h_{i} b_{\beta}^\dagger b_{\alpha} \right) \quad (2.24)$$

along with a hard-core constraint which prevents the holon and spinon from residing on the same site.
A number of interesting features of $H_{hs}$ deserve notice. In the strong-coupling limit of (2.2), the holon+spinon states are at larger energy than the energy of a single holon. However, there is an attractive interaction between the holon and spinon ($\sim t$) which is parametrically larger than the bandwidth of the holon ($\sim tJ/\lambda$). This implies that there will be a clear separation between the energy of the holon+spinon bound state and the bottom of the holon-spinon continuum, and this will be evident from our numerical results below. Although such an electron-like bound state does form in the strong coupling limit, its energy remains higher than the energy of the single holon band. This implies that doping the Mott insulator will lead to a Fermi surface of holons only, realizing a $Z_2$-ACL metal in the limit (2.2). However, as we move away from (2.2), the conditions are favorable for the holon+spinon bound state to become the lowest energy charged fermion, and doping will then lead to a $Z_2$-FL* metal. In particular, FL* is favored in the limit

$|t_{ij}| \ll |J_{ij}|$, when the on-site energy cost of the hole state in Eq. (2.11), which is $\sim \Delta^2/\lambda$, can be compensated by an energy gain $\sim -|t_{ij}|$ from the kinetic energy of the hole.

We now establish the above assertions by an exact diagonalization study of $H_{hs}$ in (2.24) in the sector with one holon and one spinon. This is most conveniently carried out in the momentum space Hamiltonian

$$H_{hs} = \sum_k E_h(k) h_k^+ h_k + \sum_k E_b(k) b_{k\alpha}^+ b_{k\alpha} + \frac{1}{L^2} \sum_{k,k',q} V(k+p+q) h_{k+q}^+ b_{-k,\alpha}^+ b_{p,\alpha} h_{p+q}$$

$$E_h(k) = (\mu - \Delta + \lambda) + \frac{t_2 J_2(\lambda + 2\Delta)}{2\Delta \lambda} (4 \cos(k_x) \cos(k_y)) + \frac{t_3 J_3(\lambda + 2\Delta)}{2\Delta \lambda} (2 \cos(2k_x) + 2 \cos(2k_y))$$

$$E_b(k) = \left( \Delta + \lambda - \sqrt{\lambda^2 + 4\Delta^2} \right) - \frac{(a+b)^2}{2} \left[ 2J_1(\cos(k_x) + \cos(k_y)) + 4J_2 \cos(k_x) \cos(k_y) \right]$$

$$+ 2J_3(\cos(2k_x) + \cos(2k_y)) \right]$$

$$V(k) = a^2 W + a^2 \left[ 2t_1(\cos(k_x) + \cos(k_y)) + 4t_2 \cos(k_x) \cos(k_y) + 2t_3(\cos(2k_x) + \cos(2k_y)) \right]$$

where $W \to \infty$ is a large repulsive energy inserted to prevent the holon and spinon from occupying the same site. We diagonalized Eq. (2.25) on a $L \times L$ lattice: after accounting for total momentum conservation, the matrix in the holon+spinon subspace is of size $L^2 \times L^2$. Results for a convenient choice of parameters are shown in Figs. 2 and 3.

### III. DYNAMICS OF VISONS

We now discuss aspects of the topological order of the Mott insulator described so far; the metallic phase C in Fig. 1 inherits the same topological order. These issues require us to return to the full gauge-invariant Hamiltonian in Eq. (1.9), and to no longer work with the large $K$ gauge-fixed version in Eq. (2.1).
FIG. 2. Lowest energy eigenvalues of $H_{hs}$ with one holon and one spinon with total momentum $\mathbf{q}$ on a lattice of size $48 \times 48$. There are $48^2 = 2304$ eigenvalues at each $\mathbf{q}$, and eigenvalues above $E_{hs} = 24.0$ are not shown. Note the bound states (which have charge $e$ and spin $S_z = \pm 1/2$) below the two particle continuum. The parameter values are $\lambda = 30.0$, $\Delta = 8.0$, $t_1 = 3.0$, $t_2 = 2.0$, $t_3 = 2.0$, $J_1 = 0.6$, $J_2 = 0.1$, $J_3 = 0.1$, and $\mu = 0$. The energy levels shift uniformly with changes in $\mu$, and the bound state will form a pocket with electron-like quasiparticles for large enough $\mu$. There is also [15] a Fermi surface of chargons associated with the single holon states, which are not shown above.

Here, a very useful fact is that the Mott insulator in Eq. (2.5) and the $\mathbb{Z}_2$ gauge theory of its bosonic spinon excitations are essentially identical to the theory of bosonic chargons presented by Paramekanti and Vishwanath in Section V.A of Ref. 9; and the change from the spinon to chargon character of the bosons makes essentially no difference to the topological analysis. The main observation is that the Mott insulator in Eq. (2.5) only has states in which the number $2N_i$ equals $\pm 1$ on every site. (We can view the deviation of $2N_i$ from $\pm 1$ as a measure of the number of spinon excitations. Or by a slight abuse of language, we identify $2N_i$ as the number of spinons ‘in the ground state’.) This observation motivates a return to the Hamiltonian in Eq. (1.9) at $p = 0$, from which we integrate out the gapped $c_{i\alpha}$ electronic excitations at large $\lambda$, when there is a large energy cost to deviations of the number, $2N_i$, from $\pm 1$. So, we obtain an effective Hamiltonian of
FIG. 3. Colour density plot of the energy of the lowest holon-spinon bound state in Fig. 2. Parameter values are the same. Note that there is no particular symmetry of the bound state dispersion associated with antiferromagnetic Brillouin zone: the minimum of the dispersion is not exactly at \((\pi/2, \pi/2)\). In contrast, the holon dispersion given by Eq. (2.19) does have a minimum at \((\pi/2, \pi/2)\).

the form

\[
\tilde{H}_{\theta, Z_2} = -\sum_{i<j} J_{ij} \mu_i^z \cos ((\theta_i - \theta_j)/2) - g \sum_{(ij)} \mu_{ij}^x - K \sum \left[ \prod \mu_{ij}^z \right] \\
+ \tilde{\Delta} \sum_i (4N_i^2 - 1)^2, \tag{3.1}
\]

where the \(\tilde{\Delta}\) term is a phenomenological representation of the energy cost for deviation of the ‘spinon’ number \(2N_i\) from \(\pm 1\). Eq. (3.1) is essentially the Hamiltonian \(H_A(\mathcal{I}^*)\) of Ref. 9. All of their arguments associated with momentum balance in the presence of flux insertion in a torus geometry go through unchanged: so we have established the existence of the needed Mott insulator described by an odd \(Z_2\) gauge theory [33, 48–50], which can then act as a parent for metallic states with Fermi surfaces of size \(p\) [13].

As argued in Ref. 9, we can proceed a step further and also integrate out the gapped spinon excitations from Eq. (3.1). Then, we obtain a pure \(Z_2\) gauge theory

\[
\tilde{H}_{Z_2} = -g \sum_{(ij)} \mu_{ij}^z - K \sum \left[ \prod \mu_{ij}^z \right], \tag{3.2}
\]
FIG. 4. A gauge-fixed state, \(|\Psi_{0v}\rangle\) of two visons marked with the X’s. The dotted line connects links with \(\mu_{ij}^z = -1\). The state \(|\tilde{\Psi}_{0v}\rangle\) is obtained after the top vison encircles the site, \(j\), marked with a circle.

where the ‘spinons’ in the ground state can be accounted for by an ‘odd’ constraint on every site \(i\) derived from Eq. (1.11)

\[
\hat{G}_i \equiv \prod_{j \in \text{n.n.}(i)} \mu_{ij}^x = -1. \tag{3.3}
\]

This is the most convenient form of the theory to investigate the dynamics of visons. Note that if we have obtained an effective theory of visons simply by integrating out the \(c_{i\alpha}\) in a weak-coupling perturbation theory in \(\lambda\), then we would have obtained the effective theory in Eq. (3.2), but with the opposite sign even constraint in Eq. (3.3): this was the procedure used in Ref. 47. So the important constraint in Eq. (3.3) relies crucially on our focus on the large \(\lambda\) physics, and the fact that at large \(\lambda\) each electron binds a spinon into the state \(|G\rangle\) in Eq. (2.14).

In the large \(K\) limit, the ground state of \(\tilde{H}_{zz}\) with the constraint in Eq. (3.3) is the same as an ‘odd’ toric code model [51]. Starting with the gauge-fixed ground state, \(|\Psi_0\rangle\), used in Section II with all \(\mu_{ij}^z = 1\), we can obtain a ground state, \(|\Psi_G\rangle\) obeying Eq. (3.3) by summing over all combinations of gauge transformations applied to this state (normalized on a torus with an even number of sites)

\[
|\Psi_G\rangle = \frac{1}{\sqrt{2}} \prod_i \frac{1}{\sqrt{2}} \left(1 - \hat{G}_i\right) |\Psi_0\rangle. \tag{3.4}
\]

All the terms on the right hand side have \(\prod_{\square} \mu_{ij}^z = 1\) on every plaquette, and so minimize the energy at large \(K\). Similarly, we can also obtain the vison excited states in the large \(K\) limit from a gauge-fixed vison wavefunction. We show in Fig. 4 a two vison state, \(|\Psi_{0v}\rangle\), with two plaquettes (each marked by an X) which have \(\prod_{\square} \mu_{ij}^z = -1\). The two vison state obeying Eq. (3.3) is obtained
by
\[ |\Psi_{Gv}\rangle = \frac{1}{\sqrt{2}} \prod_i \frac{1}{\sqrt{2}} \left( 1 - \hat{G}_i \right) |\Psi_{0v}\rangle. \] 

Unlike the toric code, these visons are not localized, and acquire a non-zero dispersion at linear order in a perturbation theory in \( g/K \). A crucial property of these mobile visons in an odd \( \mathbb{Z}_2 \) gauge theory is that they experience a Berry phase of \( \pi \) upon encircling any site of the square lattice \([33, 48, 49]\). This is apparent from the state \( |\tilde{\Psi}_{0v}\rangle \) in Fig. 4 in which the top vison has encircled the site, \( j \), marked with the circle. Then, it is easy to show that vison states obeying Eq. (3.3) satisfy
\[ |\tilde{\Psi}_{Gv}\rangle = \frac{1}{\sqrt{2}} \prod_i \frac{1}{\sqrt{2}} \left( 1 - \hat{G}_i \right) |\tilde{\Psi}_{0v}\rangle = \hat{G}_j |\Psi_{Gv}\rangle = -|\Psi_{Gv}\rangle. \] 

This Berry phase leads to a double degeneracy in the vison spectrum at all momenta. Alternatively, we can obtain an effective Hamiltonian for the visons by applying a duality transformation to Eq. (3.2): this can be done using the operator methods described by Kogut \([52]\), by the perturbative Berry phase computation of Ref. \([53]\), or by path integral methods \([33, 48, 49]\). The result by any of these methods is a transverse field Ising model on the dual lattice, with the Ising order representing the vison field operator. The odd constraint in Eq. (3.3) leads to \( \pi \) flux per plaquette in the vison hopping matrix elements. Such a model for visons was extended to FL* metals in recent work \([54]\).

IV. GENERALIZATIONS TO SU(2) GLOBAL SPIN SYMMETRY

All of our analysis so far has been restricted to the simplest case with only a U(1) global spin rotation symmetry, so that the SDW order is described by a XY order parameter. Now we consider the generalization to the physically important case of full SU(2) spin rotation symmetry. There are now significant differences determined by the specific configuration of the SDW order. In general, we can characterize a state with long-range SDW order by the expectation value
\[ \langle c_{i\alpha}^\dagger \sigma^{f}_{\alpha\beta} c_{i\beta} \rangle \sim \Phi_{x\ell} e^{iK_x \cdot r_i} + \text{c.c.} + \Phi_{y\ell} e^{iK_y \cdot r_i} + \text{c.c.}. \] 

where \( \sigma^{f} \) are the Pauli matrices, \( K_{x,y} \) are ordering wavevectors along the \( x \) and \( y \) directions, and \( \Phi_{x\ell}, \Phi_{y\ell} \) are 6 complex numbers determining the nature of the spin density order. The quantum fluctuations of the \( \Phi \) are controlled by a Landau free energy of the form \([55]\)
\[ V(\Phi) = s(|\Phi_{x\ell}|^2 + |\Phi_{y\ell}|^2) + \frac{u_1}{2} (|\Phi_{x\ell}|^4 + |\Phi_{y\ell}|^4) + \frac{u_2}{2} (|\Phi_{x\ell}^2|^2 + |\Phi_{y\ell}^2|^2) 
+ w_1 |\Phi_{x\ell}|^2 |\Phi_{y\ell}|^2 + w_2 |\Phi_{x\ell}\Phi_{y\ell}|^2 + w_3 |\Phi_{x\ell}^*\Phi_{y\ell}|^2 + \ldots \] 

Depending upon the relative values of the Landau parameters \( w_{1,2,3} \), the 6 complex numbers \( \Phi_{x\ell}, \Phi_{y\ell} \) can realize physically distinct types of spin density wave order which we consider separately
in the subsections below. Each type of order leads to a different routes towards taking the ‘square root’ of the order parameter into fractionalized spinor variables.

A. Spiral order

The case most similar to the easy-plane order is when then SDW order has a spiral form at a wavevector not equal to \((\pi, \pi)\). For simplicity, we consider the case with circular spiral spin correlations only along the wavevector \(K_x\); it is not difficult to extend the action below to also include the \(K_y\) direction. The case corresponds to (after an overall normalization)

\[
\Phi_{x\ell} = n_{1\ell} + i n_{2\ell}
\]

where \(n_{1,2\ell}\) are a pair of real orthonormal vectors

\[
\sum_{\ell} n_{1\ell}^2 = n_{2\ell}^2 = 1, \quad \sum_{\ell} n_{1\ell} n_{2\ell} = 0.
\]

The order parameter defined by Eqs. (4.3) and (4.4) is a doublet of orthonormal 3-vectors, and this is equivalent to the SO(3) manifold.

To obtain the intermediate metallic state with topological order, we need to fractionalize the above order. For the easy-plane case, we accomplished this by working with the square-root of the order parameter, \(e^{i \theta/2}\). Here, we need to introduce a bosonic complex spinor \(z\), representing the spinon excitation, which we take to be of unit length

\[
|z_\uparrow|^2 + |z_\downarrow|^2 = 1.
\]

Then the parametrizations in Eqs. (4.3,4.4) can be satisfied by the representation [37, 56]

\[
\Phi_{x\ell} = \varepsilon_{\alpha\gamma} z_\gamma \sigma^\ell_{\alpha\beta} z_\beta
\]

Note that Eq. (4.6) is invariant under the \(\mathbb{Z}_2\) gauge transformation \(z_\alpha \rightarrow -z_\alpha\), and so we will again obtain here a \(\mathbb{Z}_2\) gauge theory, similar to the easy-plane case (which corresponds to \(z_\alpha = (e^{-i\theta/2}, 0)\)).

We can now write down our theory for the interplay between spiral SDW order and \(\mathbb{Z}_2\) topological order, which is analogous to the easy-plane Hamiltonian \(\mathcal{H}_1\) in Eq. (1.9),

\[
\mathcal{H}_2 = H_c + H_{z,z_2} + H_Y
\]

\[
H_Y = -\lambda \sum_i \left[ \varepsilon_{\alpha\gamma} z_{i\gamma} \sigma^\ell_{\alpha\beta} z_{i\beta} e^{iK_x \cdot r_i} + \text{c.c.} \right] c^\dagger_{i\alpha} \sigma^\ell_{\alpha\beta} c_{i\beta}
\]

\[
H_{z,z_2} = - \sum_{i<j} J_{ij} \mu_{ij} \hat{z}_i \hat{z}_j + \text{c.c.} + \Delta \sum_i \hat{L}_i^2 - q \sum_{(ij)} \mu_{ij} - K \sum \left[ \prod \mu_{ij} \right]
\]

19
where $H_c$ was defined in Eq. (1.2), and $\vec{L}_i$ are angular momenta of the O(4) rotor defined by (4.5), analogous to $N_i$ for the easy-plane case. The on-site eigenstates of $\mathcal{H}_2$, in the limit of $K \to \infty$ and strong coupling as defined by Eq. (2.2), are described in Appendix B1.

The analog of the relationship between the chargon and electron operators in Eq. (1.7) transforming to the rotating reference frame now becomes the SU(2) rotation

$$
\begin{pmatrix}
\psi_+

\psi_-
\end{pmatrix} = 
\begin{pmatrix}
z^*_\uparrow & z^*_\downarrow \\
-z^*_\downarrow & z^*_\uparrow
\end{pmatrix}
\begin{pmatrix}
c^\uparrow \\
-c^\downarrow
\end{pmatrix}.
$$

(4.8)

In terms of the chargon operators, the Yukawa term, $H_Y$ in Eq. (4.7) takes the simple form [11]

$$
H_Y = -2\lambda \sum_i \left[ \psi^\dagger_i \psi_{i-} e^{iK_x \cdot r_i} + \text{c.c.} \right],
$$

(4.9)

which is the analog of Eq. (1.8). So analogous to the easy-plane case, the $\psi$ fermions move in a background of spatially uniform spiral order.

The subsequent discussion is a close parallel to that described above for the easy-plane case. We expect a phase diagram very similar to that in Fig. 1, with the $e^{i\theta}$ order parameter replaced by $\Phi_{x\ell}$. One difference is that the interpretation of the phase transitions in terms of vortex proliferation now needs some modification, as the order parameter is no longer XY-like but takes values in SO(3). Such an order parameter does have $\mathbb{Z}_2$ vortices, associated with the homotopy group $\pi_1(\text{SO}(3)) = \mathbb{Z}_2$. So the topological phase C is now associated with the suppression of such $\mathbb{Z}_2$ vortices, which become gapped excitations identified with visons. In contrast, the Fermi liquid phase A has proliferating $\mathbb{Z}_2$ vortices.

### B. Néel order

Next, we consider the case of two-sublattice collinear antiferromagnetism. For the cuprates, this corresponds in Eq. (4.1) to the wavevectors $K_x = K_y = (\pi, \pi)$ and real $\Phi_{x\ell} = \Phi_{y\ell}$, and is applicable to the electron-doped compounds.

In this case, the order parameters are related to a single, real vector $\Phi_{x\ell} = \Phi_{y\ell} = n_\ell/4$ obeying

$$
\sum_\ell n^2_\ell = 1.
$$

(4.10)

We fractionalize this vector by

$$
n_\ell = z^*_\alpha \sigma^\ell_{\alpha\beta} z_\beta,
$$

(4.11)

which leaves a U(1) gauge invariance under $z_\alpha \to e^{i\ell} z_\alpha$. So now, our generalization of the $\mathbb{Z}_2$ gauge
theory $\mathcal{H}_2$ in Eq. (4.7) is a U(1) gauge theory for SDW order in metals:

$$
\mathcal{H}_3 = H_c + H_{z,U(1)} + H_Y
$$

$$
H_Y = -\lambda \sum_i \eta_i z^*_i \sigma^{\ell}_{\alpha\beta} z_i \sigma^{\ell}_{\alpha\beta} c_i^{\dagger} c_i
$$

$$
H_{z,U(1)} = -\sum_{i<j} J_{ij} (z^*_i e^{iA_{ij}} z_j + \text{c.c.}) + \Delta \sum_i \vec{L}_i^2 + g \sum_{(ij)} E_{ij}^2 - K \sum \cos \left( \sum A_{ij} \right), \quad (4.12)
$$

where $A_{ij}$ is a compact U(1) gauge field on the links of the square lattice, and $E_{ij}$ is the canonically conjugate electric field. The on-site eigenstates of $\mathcal{H}_3$, in the limit of $K \to \infty$ and strong coupling as defined by Eq. (2.2), are described in Appendix B2.

The analysis proceeds as in Section IV A. We transform to the rotation reference frame in terms of chargons as in Eq. (4.8), and the Yukawa coupling for the chargons is [10]

$$
H_Y = -\lambda \sum_i \eta_i \left[ \psi^+_i \psi_i + \psi^+_i \psi_i - \psi^+_i \psi_i \right], \quad (4.13)
$$

which is the analog of Eqs. (1.8) and (4.9). So the phase diagram of $\mathcal{H}_3$ will be analogous to that for $\mathcal{H}_1$ in Fig. 1, with the chargons experiencing uniform Néel order given by Eq. (4.13) in phase C.

One important difference between the cases with $\mathbb{Z}_2$ and U(1) gauge theory is that pure U(1) gauge theory is always confining in two spatial dimensions due to the proliferation of monopoles. At $p = 0$, this mechanism will lead to insulating states with valence bond solid order [57–59]. At non-zero $p$, monopoles can be suppressed by Fermi surfaces of particles carrying U(1) gauge charges [60, 61], and this mechanism can stabilize a U(1)-ACL in the intermediate phase C. But it should be noted that there is a non-BCS pairing instability of such a metallic state [62], and after the Fermi surface has been gapped by pairing, monopole-induced confinement will reappear.

### C. Stripe order

Finally, we consider the case of collinear spin order at wavevectors not equal to $(\pi, \pi)$. By symmetry, such spin order is accompanied by charge density wave order at twice the wavevector [63]. Considering the case of uni-directional stripes with wavevector $\mathbf{K}_x$ for simplicity, the ordering is described by Eq. (4.1) with

$$
\Phi_{x\ell} = e^{i\phi} n_{\ell}, \quad (4.14)
$$

where $n_{\ell}$ is a real vector obeying Eq. (4.10). The order parameter for a charge density wave at wavevector $2\mathbf{K}_x$ is $e^{2i\phi}$.

Following the discussion in the previous subsections, we now have to examine fractionalizations of the stripe order parameter in Eq. (4.14). This question has been considered in a number of
previous works [55, 64–68], all of which used a $\mathbb{Z}_2$ gauge theory to fractionalize $\Phi_{x\ell}$ into its charge and spin components, represented by $e^{i\phi}$ and $n_{\ell}$ respectively; such a fractionalization is invariant under

$$e^{i\phi_i} \rightarrow s_i e^{i\phi_i} \quad n_{\ell} \rightarrow s_i n_{\ell}$$

(4.15)

with $s_i = \pm 1$ the $\mathbb{Z}_2$ gauge transformation. However, this fractionalization is not suitable for our purposes because it does not involve spinor variables, and so cannot yield Fermi surface reconstruction in the phase with topological order. So we examine combining the above fractionalization with the spinor decomposition of $n_{\ell}$ in Section IV B:

$$\Phi_{x\ell} = e^{i\phi} z_\alpha^* \sigma_{\alpha\beta}^\ell z_\beta.$$  

(4.16)

Then, in terms of the chargon variables in Eq. (4.8), the Yukawa term coupling the order parameter to the fermions (analogous to Eqs. (1.8), (4.9) and (4.13)) becomes

$$H_Y = -\lambda \sum_i 2 \cos (\phi + K_x \cdot \mathbf{r}_i) \left[ \psi_{i+}^\dagger \psi_{i+} - \psi_{i-}^\dagger \psi_{i-} \right] .$$

(4.17)

This expression makes it clear that the chargons move in the presence of a non-fluctuating potential only in a state with long-range charge density wave order with $\langle e^{i\phi} \rangle \neq 0$. In such a situation, the theory of SDW order in the stripe model reduces [10] to the U(1) gauge theory already considered in Section IV B. So Fermi surface reconstruction in a state with topological order requires charge density wave order in the stripe model.

V. CONCLUSIONS

This paper has presented an alternative approach to symmetry breaking and topological order in doped Mott insulators. Instead of the conventional focus on electron fractionalization, we set up a formalism based upon order parameter fractionalization. So our main Hamiltonian for the case of a XY SDW order parameter in Eq. (1.9) involved a fractionalized ‘square root’ of the SDW order parameter, but retained the unfractionalized bare electron operator. Starting from Eq. (1.9), we obtained the phase diagram in Fig. 1, containing states that had previously been obtained from the more common electron fractionalization route. The advantage of our formalism is that offers a focus on just the phases observed in experiments, while being very economical in using extraneous degrees of freedom which have to be projected out. Furthermore, generalizations of the models in Eqs. (1.9) and (2.1) are amenable to sign-problem-free quantum Monte Carlo simulation by the methods of Refs. 28–30, which can also study the connection to superconductivity.

We began with the LGW-Hertz theory for the onset of SDW order in metals: this exhibits phases A and B in Fig. 1, the Fermi liquid with large Fermi surface, and the SDW metal with small pocket
Fermi surfaces. Both phases have well-defined electronic quasiparticles and their Fermi surfaces sizes obey the conventional Luttinger theorem. The phase transition between A and B has also been extensively studied [20–27]. (In the limit of zero doping, \( p = 0 \), phase A remains a Fermi liquid, while phase B becomes a Slater insulator with long-range antiferromagnetic order.)

We argued that the LGW-Hertz theory could be modified to Eq. (1.9) by introducing a \( Z_2 \) gauge field, and this allowed a phase transition in which the destruction of SDW order did not coincide with appearance of a large Fermi surface: this led to phase C with pocket Fermi surfaces (of chargons and/or electrons) and no SDW order. (In the limit \( p = 0 \), phase C becomes a Mott insulator with \( Z_2 \) topological order.) For the case of easy-plane SDW order, we showed that the transition from phase B to phase C was associated with the proliferation of doubled vortices. The universality class of the B-C transition has been identified in earlier work [47, 69] as a relativistic 2+1 dimensional \( O(2)^* \) field theory for the easy plane case (\( O(4)^* \) for the Heisenberg case) [70].

Phase C (or more properly, its SU(2) spin rotation analogs in Section IV) is proposed as the pseudogap state of the hole-doped cuprates, present in between the phase B at low \( p \), and phase A above optimal doping. Other less-correlated high temperature superconductors (such as the pnictides) are proposed to go directly from phase B to phase A. We maintain that this simple connection between different families of superconductors supports our model, and the unified phase diagram in Fig. 1.

The main problem left open in our analysis is the nature of the transition from phase C to phase A. This is a candidate transition for the physics near optimally hole-doped cuprate superconductors. The simple model in Eq. (1.9) contains such a transition, but does not easily yield a continuum theory for the quantum criticality. Phase C is in a deconfined phase of a \( Z_2 \) gauge theory, while phase A is in a confined phase. However, the transition between them is not just a \( Z_2 \) confinement transition: the \( Z_2 \) gauge theory also changes from an “odd” gauge theory [33, 48–50] (with \( e^{2\pi Ni} = -1 \)) in phase C to an “even” gauge theory (with \( e^{2\pi Ni} = 1 \)) in phase A. Continuum formulations of confinement transitions in \( Z_2 \) gauge theories in the presence of gauge-charged matter require duality transforms to vison fields via mutual Chern-Simons terms [71], which we have not discussed here. It is possible that such an analysis of the criticality will eventually lead to the deconfined SU(2) gauge theory for the C-A transition proposed in Refs. 10–13.

Acknowledgements

We thank D. Chowdhury, M. Metlitski, T. Senthil, and A. Vishwanath for useful discussions. The research was supported by the NSF under Grant DMR-1360789, by the ISF under Grant 1291/12, by the BSF under Grant 2014209, and by a Marie Curie CIG grant. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province
Appendix A: Additional phases

In our phase diagram in Fig. 1 for the easy-plane Hamiltonian \( \mathcal{H}_1 \) in Eq. (1.9), we have 2 phases with no broken symmetry, phases A and C. In phase C we are in the deconfined phase of the \( \mathbb{Z}_2 \) gauge theory, and the spinon number obeys \( e^{2\pi N_i} = -1 \) on each site. In contrast, in phase A we are in the confined phase of the \( \mathbb{Z}_2 \) gauge theory, and the spinon number obeys \( e^{2\pi N_i} = 1 \) on each site. There is no fundamental reason for these assignments of spinon number, and we can also imagine additional phases with the opposite assignment; such phases are not shown in Fig. 1.

A \( \mathbb{Z}_2 \) deconfined phase with \( e^{2\pi N_i} = 1 \) would have Fermi surfaces of chargons and/or electrons of total size \( 1 + p \), by the flux piercing arguments in Refs. 8, 9, and 13. However, such a phase is energetically disfavored at large \( \lambda \), and so not suitable for the physics of the Mott-Hubbard systems under consideration here.

More relevant is a \( \mathbb{Z}_2 \) confined phase with \( e^{2\pi N_i} = -1 \). This must have valence bond solid (VBS) order, as established in early work \[33, 48, 49\]. More recent work has shown \[54\] that the VBS order can have a variety of complex spatial configurations, depending upon the nature of frustrating interactions. The Fermi surface is expected to be small by the flux piercing arguments, but with the doubling of the unit cell by the VBS order, there is no fundamental distinction between small and large Fermi surfaces. Such a confining phase with VBS order is a possibility in Mott-Hubbard models with frustrated exchange interactions \[14\], but is not shown in Fig. 1.

Appendix B: Single-site eigenstates with SU(2) symmetry

1. O(4) model: Spiral order

In the limit \( K \to \infty \), the fluctuations of the \( \mathbb{Z}_2 \) gauge field are frozen, and we choose the gauge \( \mu_{ij}^z = 1 \). Then the model in Eq. (4.7) reduces to:

\[
\mathcal{H}'_2 = -\sum_{i,j} (t_{ij} + \mu \delta_{ij}) \, c_i^\dagger c_j - \lambda \sum_i \left[ \bar{z}_i \gamma \sigma_{\alpha \beta} z_i e^{i K \cdot r_i} + c.c. \right] c_i^\dagger \sigma_{\alpha \beta} c_i \\
- \sum_{i<j} J_{ij} \left( \bar{z}_{i\alpha} z_{j\alpha} + c.c. \right) + \Delta \sum_i L_i^2
\]

Analogous to the \( \mathbb{Z}_2 \) case, the \( J_{ij} \) term involves coupling to the fractionalized spinon-fields \( z_\alpha \), which has \( S_z \) quantized in units of 1/2 on every site. We now further specialize to the strong coupling limit as defined by Eq. (2.2), and show that the ground state at \( p = 0 \) is again a Mott
insulator with odd $\mathbb{Z}_2$ topological order. To do this, we define the on-site Hamiltonian $H_o$ by dropping the site index $i$ and letting $\xi_i = e^{iK\cdot r_i}$:

$$H_o = -\mu c_{\alpha}^\dagger c_{\alpha} - \lambda \left[ \varepsilon_{\alpha\gamma} z_{\gamma} \sigma_{\alpha\beta} z_{\beta} \xi_i + \text{c.c.} \right] c_{\alpha}^\dagger \sigma_{\alpha\beta} c_{\beta} + \Delta \sum_{\mu=1}^{6} \tilde{L}_{\mu}^2$$

$$= -\mu c_{\alpha}^\dagger c_{\alpha} - 2\lambda \left[ (\xi_i z^2_{\gamma} - \xi_i^* (z^*_{\gamma})^2) c_{\gamma}^\dagger c_{\gamma} + (-\xi_i z^2_{\gamma} + \xi_i^* (z^*_{\gamma})^2) c_{\gamma}^\dagger c_{\gamma} + (\xi_i z^2_{\beta} - \xi_i^* (z^*_{\beta})^2) c_{\beta}^\dagger c_{\beta} + (\xi_i z^2_{\alpha} - \xi_i^* (z^*_{\alpha})^2) c_{\alpha}^\dagger c_{\alpha} \right]$$

$$+ \Delta \sum_{\mu=1}^{6} \tilde{L}_{\mu}^2$$

We start by looking at the eigenmodes of the $O(4)$ rotor angular momenta $\sum_{\mu=1}^{6} \tilde{L}_{\mu}^2$. These are given by the hyperspherical harmonics $Y_{l,m}^n$, which are a complete set of eigenfunctions on the 3-sphere $S^3$ (generalizations of the spherical harmonics $Y_{m}^l$ on $S^2$).

$$\left( \sum_{\mu} \tilde{L}_{\mu}^2 \right) Y_{l,m}^n = n(n+2)Y_{l,m}^n, \quad n \in \{0, 1, 2, \ldots\}$$

We can conveniently describe these eigenmodes in the toroidal coordinates [72], which we define as follows:

$$(z_{\gamma}, z_{\beta}) = (\cos(\beta)e^{-i\theta}, \sin(\beta)e^{-i\phi}), \quad \text{where } 0 \leq \beta \leq \pi/2, \ 0 \leq \theta, \phi < 2\pi$$

In these coordinates, the hyperspherical harmonics are given by (here we choose a slightly different basis compared to Ref. 72 for later computational convenience)

$$Y_{l,m}^n = N_{l,m}^n e^{il\theta} e^{im\phi} \cos^{|l|}(\beta) \sin^{|m|}(\beta) P_d^{|l|,|m|}(\cos(2\beta)), \quad d = \frac{n - (|l| + |m|)}{2} \in \mathbb{Z}, \ |l| + |m| \leq n$$

where $P_d^{|l|,|m|}(u)$ are the Jacobi polynomials and $N_{l,m}^n$ are appropriate normalization constants, which we provide explicitly below for completeness:

$$P_d^{|l|,|m|}(u) = \frac{1}{2^d} \sum_{i=0}^{d} \binom{|m| + d}{i} \left( \binom{|l| + d}{d - i} \right) (u+1)^i (u-1)^{d-i}$$

$$N_{l,m}^n = \sqrt{ \frac{2(n+1)d!(|l| + |m| + d)!}{(|l| + d)!(|m| + d)!} }$$

Using the above coordinates, we list the important low-lying eigenstates.

(i) Mott insulator

In analogy with the easy-axis case, we look for the lowest energy state with a single electron per site in the $n = 1$ subspace of the hyperspherical harmonics. We choose the following basis, labeling the states as $c_{\gamma}^\dagger Y_{1,m}^1 |0\rangle$:

$$\{ c_{\gamma}^\dagger Y_{1,0}^1 |0\rangle, c_{\gamma}^\dagger Y_{-1,0}^1 |0\rangle, c_{\gamma}^\dagger Y_{0,1}^1 |0\rangle, c_{\gamma}^\dagger Y_{0,-1}^1 |0\rangle, c_{\gamma}^\dagger Y_{1,1}^1 |0\rangle, c_{\gamma}^\dagger Y_{1,-1}^1 |0\rangle, c_{\gamma}^\dagger Y_{0,1}^1 |0\rangle, c_{\gamma}^\dagger Y_{0,-1}^1 |0\rangle \}$$
In this subspace, we have:

\[
H_o = (-\mu + 3\Delta)I_{8\times8} - \frac{2\lambda}{3}
\begin{pmatrix}
0 & 0 & 0 & \xi_i^* & 0 & 0 & 0 & 0 \\
0 & \xi_i & 0 & -2\xi_i & 0 & 0 & 0 & 0 \\
0 & \xi_i^* & 0 & 0 & 0 & 0 & 0 & 2\xi_i^* \\
\xi_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -2\xi_i^* & 0 & 0 & 0 & 0 & -\xi_i & 0 \\
0 & 0 & 0 & 0 & -\xi_i^* & 0 & 0 & 0 \\
0 & 2\xi_i & 0 & -\xi_i & 0 & 0 & 0 & 0
\end{pmatrix}
\]  

(B8)

The lowest-energy Mott insulating state \(|G\rangle\), with energy \(E_G = -\mu + 3\Delta - 2\lambda\), is given by:

\[
|G\rangle = \frac{1}{2} \left( \xi_i^* c_i^\dagger Y_{1,0}^1 + \xi_i c_i^\dagger Y_{1,0}^1 - \xi_i^* c_i^\dagger Y_{0,1}^1 + c_i^\dagger Y_{0,1}^1 \right) |0\rangle
\]

= \frac{1}{2\pi} \left[ \left( c_i^\dagger z_i + c_i^\dagger \xi_i^* \right) \xi_i \left( c_i^\dagger z_i^* + c_i^\dagger \xi_i \right) \right] |0\rangle \sim \left( \psi_i^\dagger - \xi_i^* \psi_i^\dagger \right) |0\rangle

(B9)

where \(\psi_{\pm}\) are the spinless fermionic chargons defined in Eq. (4.8). The last representation makes it evident that the Mott insulating ground state does not carry any spin. However, it carries \(\mathbb{Z}_2\) gauge charge, and therefore the \(\mathbb{Z}_2\) gauge theory is odd.

(ii) Spinons

These are doubly degenerate states which have \(S_z = \pm 1/2\), and a \(\mathbb{Z}_2\) gauge charge relative to the Mott insulator, but no electromagnetic charge. We consider only the |↑\rangle spinon, the calculations for the |↓\rangle spinon are identical. Therefore, we choose the following basis of states which span the subspace with \(S_z = 1/2\):

\[
\{ c_i^\dagger Y_{0,0}^0 |0\rangle, c_i^\dagger Y_{0,0}^2 |0\rangle, c_i^\dagger Y_{2,1}^1 |0\rangle, c_i^\dagger Y_{2,-1}^1 |0\rangle, c_i^\dagger Y_{2,0}^2 |0\rangle, c_i^\dagger Y_{0,2}^2 |0\rangle, c_i^\dagger Y_{0,2}^2 |0\rangle, c_i^\dagger Y_{1,-1}^2 |0\rangle \}
\]

(B10)

In this basis, we have:

\[
H = -\mu I_{7\times7} + \begin{pmatrix}
0 & 0 & -\sqrt{2/3}\lambda\xi_i & -\sqrt{2/3}\lambda\xi_i^* & \frac{2\lambda\xi_i}{\sqrt{3}} & \frac{2\lambda\xi_i^*}{\sqrt{3}} & 0 \\
0 & 8\Delta & 0 & 0 & \lambda\xi_i & \lambda\xi_i^* & 0 \\
-\sqrt{2/3}\lambda\xi_i^* & 0 & 8\Delta & 0 & 0 & -\lambda\xi_i & 0 \\
-\sqrt{2/3}\lambda\xi_i & 0 & 0 & 8\Delta & 0 & 0 & -\lambda\xi_i \\
-\frac{2\lambda\xi_i}{\sqrt{3}} & \lambda\xi_i^* & 0 & 8\Delta & 0 & \lambda\xi_i & \frac{\lambda\xi_i}{\sqrt{2}} \\
-\frac{2\lambda\xi_i^*}{\sqrt{3}} & \lambda\xi_i & 0 & 0 & 8\Delta & \frac{\lambda\xi_i}{\sqrt{2}} & \frac{\lambda\xi_i}{\sqrt{2}} \\
0 & 0 & -\lambda\xi_i & \lambda\xi_i^* & \frac{\lambda\xi_i}{\sqrt{2}} & \frac{\lambda\xi_i}{\sqrt{2}} & 8\Delta
\end{pmatrix}
\]

(B11)

For all positive values of \(\lambda\) and \(\Delta\), we find that the energy of the lowest-lying spinon state is \(E_s = -\mu + 4\Delta - \sqrt{16\Delta^2 + 4\lambda^2}\). For the Mott insulator to have lower energy than the spinon, we
require \( E_G < E_s \), which translates to \( \lambda > 15\Delta/4 \). The spin-gap of the Mott-insulator is given by:

\[
\Delta_s = E_s - E_G = 2\lambda + \Delta - \sqrt{16\Delta^2 + 4\lambda^2}
\]

(iii) Holon
This is the empty state \(|0\rangle\), with energy \( E_{hn} = 0 \). Relative to the Mott insulator, it has \( S_z = 0 \), electromagnetic charge \( +e \) and non-zero \( \mathbb{Z}_2 \) gauge charge.

(iv) Doublon
This is the state \( c^\dagger_l c^\dagger_m |0\rangle \), with energy \( E_d = -2\mu \). Relative to the Mott insulator, it has \( S_z = 0 \), electromagnetic charge \( -e \) and non-zero \( \mathbb{Z}_2 \) gauge charge.

(v) Holes
These are the 4 degenerate states given by the \( n = 1 \) hyperspherical harmonics, which can be represented as \( Y^1_{l,m} |0\rangle \), for \( \{ l = \pm 1, m = 0 \} \) and \( \{ l = 0, m = \pm 1 \} \). Each state has energy given by \( E_h = 3\Delta \). They have electromagnetic charge \( +e \) and \( S_z = \pm 1/2 \) relative to the Mott insulator, but no \( \mathbb{Z}_2 \) gauge charge. The energy difference between a pair of sites with hole+Mott insulator and a pair with holon+spinon is given by:

\[
E_G + E_h - E_{hn} - E_s = 2\Delta - 2\lambda + \sqrt{16\Delta^2 + 4\lambda^2} > 0
\]

which is the same condition as Eq. (B13). Therefore, the hole is unstable to decay to a holon and a spinon in the strong-coupling limit.

(vi) Electrons
These are again 4 degenerate states given by \( c^\dagger_l c^\dagger_m Y^1_{l,m} |0\rangle \) for \( \{ l = \pm 1, m = 0 \} \) and \( \{ l = 0, m = \pm 1 \} \). One can check that \( H_{int} \) acting on any of these states gives zero, so only the diagonal terms matter and therefore, they have energy \( E_e = -2\mu + 3\Delta \). They have electromagnetic charge \( -e \) and \( S_z = \pm 1/2 \) relative to the Mott insulator, but no \( \mathbb{Z}_2 \) gauge charge. The energy difference between a pair of sites with electron+Mott-insulator and a pair with doublon+spinon is given by:

\[
E_G + E_e - E_{dn} - E_s = 2\Delta - 2\lambda + \sqrt{16\Delta^2 + 4\lambda^2} > 0
\]

which is the same condition as Eq. (B13). Therefore, the electron is also unstable to decay to a doublon and a spinon in the strong-coupling limit.

Finally, note that we can define a new spinless fermionic operator as a linear combination of \( \psi_{i,\pm} \), given by:

\[
\psi_i = \frac{1}{\sqrt{2}} (\psi_{i,+} - \xi_i \psi_{i,-})
\]

\( \psi \) is the holon creation operator, and the Mott insulator is a filled band of \( \psi \), given by

\[
|G\rangle = \prod_i \psi_i^\dagger |0\rangle
\]

We can also define bosonic spinon creation operators analogous to Eq. (2.16) which will create \( S_z = \pm 1/2 \) excitations over the Mott insulating ground state using the eigenstates of Eq. (B11).
2. O(4) model: Néel order

For the fractionalization defined by \( n_\ell = z_\alpha^* \sigma^\ell_{\alpha \beta} z_\beta \), the constraint defined by Eq. (4.10) can be re-written as \( |z_\uparrow|^2 + |z_\downarrow|^2 = 1 \). Therefore, the dynamics of the order parameter field are described by an O(4) model which is coupled to the \( c \) fermions. We again consider the limits of \( K \to \infty \) and strong-coupling as defined by Eq. (2.2), resulting in the following on-site Hamiltonian \( H_o \) (with \( \eta_i = (-1)^{x_i + y_i} \)):

\[
H_o = -\mu c_{\alpha}^\dagger c_\alpha - \lambda \eta_i z_\alpha^* \sigma^\ell_{\alpha \beta} z_\beta c_{\alpha}^\dagger \sigma^\ell_{\alpha \beta} c_\beta + \Delta \sum_\mu \vec{L}_\mu^2
\]

\[
= -\mu c_{\alpha}^\dagger c_\alpha - \lambda \eta_i \left[ 2 z_\uparrow^* z_\downarrow c_{\uparrow}^\dagger c_{\downarrow}^\dagger + 2 z_\downarrow^* z_\uparrow c_{\downarrow}^\dagger c_{\uparrow}^\dagger + \left( z_\uparrow^* z_\downarrow - z_\downarrow^* z_\uparrow \right) \left( c_{\uparrow}^\dagger c_{\downarrow}^\dagger - c_{\downarrow}^\dagger c_{\uparrow}^\dagger \right) \right] + \Delta \sum_\mu \vec{L}_\mu^2 \quad (B17)
\]

The rest of the calculation exactly follows Appendix B1. Only the Mott-insulator and the spinons eigenstates are different, so we restrict the following description to these two kinds of eigenstates.

(i) Mott insulator: In the basis described in Eq. (B7) we have:

\[
H_o = (-\mu + 3\Delta) I_{8 \times 8} - \frac{\lambda \eta_i}{3} \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & -1 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} \quad (B18)
\]

The lowest energy state is given by:

\[
|G\rangle = \begin{cases}
\frac{1}{\sqrt{2}} \left( c_{\uparrow}^\dagger Y_{1,0}^1 + c_{\downarrow}^\dagger Y_{0,-1}^1 \right) |0\rangle \quad \text{for } \eta_i = 1, \\
\frac{1}{\sqrt{2}} \left( -c_{\uparrow}^\dagger Y_{0,1}^1 + c_{\downarrow}^\dagger Y_{1,0}^1 \right) |0\rangle \quad \text{for } \eta_i = -1
\end{cases}
\]

Therefore, the Mott insulator can be written conveniently in terms of the spinless fermionic chargons as:

\[
|G\rangle = \psi_{\eta_i}^\dagger |0\rangle, \quad E_G = -\mu + 3\Delta - \lambda \quad (B19)
\]

In this representation, it is evident that \( |G\rangle \) does not carry any spin but carries a non-zero \( \mathbb{Z}_2 \) gauge charge, and and therefore the \( \mathbb{Z}_2 \) gauge theory is odd.

(ii) Spinons: These are doubly degenerate states which have \( S_z = \pm 1/2 \), and a \( \mathbb{Z}_2 \) gauge charge relative to the Mott insulator, but no electromagnetic charge. We can find the \( |\uparrow\rangle \) spinon by using
the $S_z = 1/2$ subspace defined in Eq. (B10).

\[ H_o = -\mu I_{7 \times 7} + \begin{pmatrix} 0 & -\frac{\lambda \eta}{\sqrt{3}} & 0 & 0 & 0 & 0 & -\sqrt{\frac{2}{3}} \lambda \eta_i \\ -\frac{\lambda \eta}{\sqrt{3}} & 8\Delta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8\Delta & 0 & -\frac{\lambda \eta}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 8\Delta & 0 & -\frac{\lambda \eta}{\sqrt{2}} & 0 \\ 0 & 0 & -\frac{\lambda \eta}{\sqrt{2}} & 0 & 8\Delta + \frac{\lambda \eta}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{\lambda \eta}{\sqrt{2}} & 0 & 8\Delta - \frac{\lambda \eta}{2} & 0 \\ -\sqrt{\frac{2}{3}} \lambda \eta_i & 0 & 0 & 0 & 0 & 0 & 8\Delta \end{pmatrix} \]  

(B20)

The energy of the lowest-lying spinon state is \( E_s = -\mu + 4\Delta - \sqrt{16\Delta^2 + \lambda^2} \). For the Mott insulator to have lower energy than the spinon, we require \( E_G < E_s \), which translates to \( \lambda > 15\Delta/2 \). The spin-gap of the Mott-insulator is given by:

\[ \Delta_s = E_s - E_G = \lambda + \Delta - \sqrt{16\Delta^2 + \lambda^2} \]  

(B21)

The descriptions of the (iii) holon, (iv) doublon, (v) holes and (vi) electrons are identical to Appendix B1. The only difference arises from the change in energy eigenvalues \( E_G \) and \( E_s \). This changes the energy gap between a pair of sites with hole+Mott insulator (electron + Mott insulator) and a pair with holon+spinon (doublon+spinon):

\[ E_G + E_h - E_{hn} - E_s = E_G + E_e - E_{dn} - E_s = 2\Delta - \lambda + \sqrt{16\Delta^2 + \lambda^2} > 0 \]  

(B22)

As before, we observe that the hole (electron) is unstable to decay to a holon (doublon) and a spinon in the strong-coupling limit.


