QED_3 theory of underdoped high-temperature superconductors

Igor F. Herbut
Department of Physics, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6
(Received 25 February 2002; revised manuscript received 4 June 2002; published 4 September 2002)

The low-energy theory of d-wave quasiparticles coupled to fluctuating vortex loops that describes the loss of phase coherence in a two-dimensional d-wave superconductor at $T=0$ is derived from first principles. The theory has the form of $(2+1)$ dimensional quantum electrodynamics (QED_3), and is proposed as an effective description of the $T=0$ superconductor-insulator transition and of the pseudogap phase in underdoped cuprates. The coupling constant ("charge") in this theory is proportional to the dual order parameter of the XY model, which is assumed to describe fluctuations of the phase of the superconducting order parameter. Finiteness of the charge is then tantamount to the appearance of infinitely large vortex loops, i.e., to the loss of phase coherence in the system. The principal result is that the destruction of the superconducting phase coherence in the d-wave superconductors typically, and immediately, leads to the appearance of antiferromagnetism. This transition can be understood in terms of the spontaneous breaking of an approximate "chiral" $SU_c(2)$ symmetry, which may be discerned at low enough energies in the standard d-wave superconductor. The mechanism of this spontaneous symmetry breaking is formally analogous to the dynamical mass generation in QED_3, with the "mass" here being proportional to staggered magnetization. Other phases with broken chiral symmetry include the translationally invariant "$d+ip$" and "$d+is$" insulators, and the one-dimensional charge-density and spin-density waves, which are all insulating descendants of the d-wave superconductor. All the insulating states have neutral spin-1/2 excitations that one can identify in the superconductor confined by the logarithmic potential. Electron repulsion is in this formalism represented by a particular quartic perturbation to the QED_3 action, which breaks the chiral symmetry and selects the antiferromagnet as the preferred broken symmetry state. I formulate the mean-field theory of the antiferromagnetic instability in presence of a short-range repulsive interaction, and find the staggered magnetization to be significantly enhanced deeper inside the insulating state. The theory offers an explanation for the rounded d-wave-like dispersion seen in angle-resolved photoemission spectroscopy experiments on the insulating $Ca_2CuO_2Cl_2$ [F. Ronning *et al.*, Science **282**, 2067 (1998)]. Relations to other theoretical approaches to the high-T_c problem are discussed.

DOI: 10.1103/PhysRevB.66.094504 PACS number(s): 74.20.Mn, 74.25.Jb, 74.40.+k

I. INTRODUCTION

Soon after the original discovery, it became well appreciated that the high-temperature (high-T_c) superconductors are all quasi-two-dimensional insulating antiferromagnets that become superconducting with the introduction of holes. The nature of the relationship between antiferromagnetism and high-temperature superconductivity has been the central issue in the field. Following the time honored strategy of understanding first the non-superconducting state, most of the approaches to the high-T_c problem focused on finding the mechanism by which doping an antiferromagnet would produce a superconductor.[1] The essential difficulty in pursuing this strategy seems to be that the Mott insulator is itself a nontrivial strongly correlated state, harder to describe in simple terms than the metallic Fermi liquid, which played its role in the BCS theory of the low-temperature superconductivity.[2] The situation becomes only worse away from half-filling, where the ground state of even the simplest models becomes more ambiguous. Experimentally, the cuprates seem to loose their antiferromagnetic ordering with doping before they become superconducting, and many candidates for the intermediate "pseudogap phase" have been discussed in literature. The nature of the nonsuperconducting state that is supposed to be unstable to superconductivity with doping is at this point, however, far from clear and may prove to be nonuniversal. Arguably, the physics of the underdoped regime may be the main mystery of high-temperature superconductivity.

In a remarkable contrast to the uncertainties inherent to the insulating phase, the superconducting phase of most high-T_c materials is well established to have d-wave symmetry of the order parameter,[3,4] typically with well-defined, long-lived quasiparticle excitations.[5,6] This simplicity suggests that an *inverted* approach to the high-T_c problem may be more natural:[7] if there exists a d-wave state in the phase diagram, what *other* states can in principle be inferred from it? The purpose of this paper is to establish the theoretical framework for answering this question, answer it, and show how this may help explain some salient features of the cuprate phase diagram and the angle-resolved photoemission spectroscopy (ARPES) experiments in the insulating state.[8,9]

Loosely speaking, there are two ways to destroy a superconducting state: (1) by driving the amplitude of the order parameter to zero, which is what is well described by the weak-coupling BCS theory at finite temperature,[2] for example. For d-wave superconductors this process presumably is relevant at large dopings, where weak-coupling treatments of the Hubbard and related models can be trusted, and disorder should eventually force T_c to vanish.[10] (2) Even if the amplitude of the order parameter is large and finite, superconductivity will be lost with the destruction of phase order.[11,12] There is evidence that this is what actually occurs in underdoped cuprates, where the superconducting transition temperature (T_c) is much lower that the pseudogap temperature T^*. Since underdoped cuprates are strongly two dimensional, at finite temperatures the loss of phase order may be expected to proceed via the Kosterlitz-Thouless transition, and indeed, there are distinct experimental signatures of the fluctuating vortices above T_c.[13,14] The following question then naturally arises: What is the nature of the $T=0$ phase that derives from a two-dimensional (2D) d-wave superconductor when the phase coherence is lost, but the order parameter amplitude is still finite? The central thesis of this work is that the phase incoherent d-wave superconductor (dSC) is nothing but the insulating (typically incommensurate) spin-density-wave (SDW), i.e., weak antiferromagnet. Short account of this result appeared earlier in Ref. 15.

I show that the minimal continuum theory of the low-energy quasiparticle excitations near the four nodes of the d-wave order parameter coupled to fluctuating vortex loops at $T=0$ is provided by $(2+1)$-dimensional quantum electrodynamics (QED_3):

$$S=\int d^2\vec{r}\, d\tau \left[\bar{\Psi}_i \gamma_\nu(\partial_\nu + i a_\nu)\Psi_i + \frac{1}{2|\langle\Phi\rangle|^2}(\vec{\nabla}\times\vec{a})^2 \right], \tag{1}$$

where $\nu=0$ (imaginary time) 1, 2 (space), and the sum over repeated indices is assumed. The *four-component* Dirac fields Ψ_i, $i=1,2$ represent the sharp, electrically neutral spin-1/2 excitations one can define in the superconducting state (and hence may call "spinons"), which are minimally coupled to a massless gauge field $\vec{a}$. The gauge field derives from the fluctuating topological defects (vortex loops) in the phase of the superconducting order parameter, which have been integrated out in deriving the theory (1). The complex number $\langle\Phi\rangle$ is proportional to the the disorder (dual order) parameter[16] and represents the state of vortex loops: $\langle\Phi\rangle \neq 0$ signals the appearance of infinitely large loops in the system and the loss of phase coherence, which is the $T=0$ analog of the Kosterlitz-Thouless transition.[17] In the superconducting state, on the other hand, all loops are of finite size, $\langle\Phi\rangle=0$, and the gauge field, in the simplest approximation, may be considered effectively decoupled from the spinons: quasiparticle excitations are then sharp, since all the short-range interactions that have not been explicitly written in Eq. (1), if weak enough, are strongly irrelevant. When $\langle\Phi\rangle\neq 0$ the situation becomes radically different, as the gauge field mediates a long-range interaction between spinons. In reality the theory is also strongly anisotropic, but for simplicity this possibly important feature has been neglected in writing Eq. (1). QED_3 has also been recently considered by Franz and Tešanović[18] as an effective description of the pseudogap state. They argued that the presence of the massless gauge field may explain the broad features seen in ARPES measurements in the normal state.[6,19] Here I show that at $T=0$ as soon as $\langle\Phi\rangle$ becomes finite there is a dynamical generation of the mass term $\sim m\bar{\Psi}_i\Psi_i$ in Eq. (1), which can be identified as the staggered potential "felt" by the original electrons, i.e., with the SDW order parameter. Quantum fluctuating dSC is thus at $T=0$ inherently unstable towards SDW ordering once the phase coherence is lost.

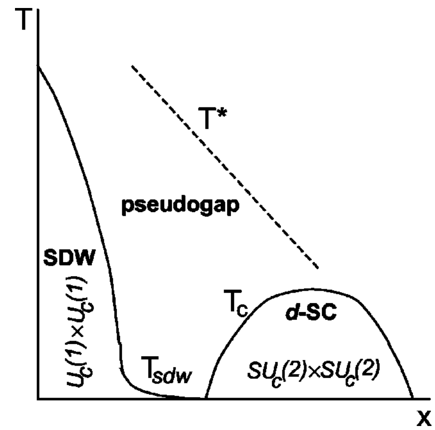

FIG. 1. A schematic phase diagram of cuprate superconductors in terms of the low-energy chiral symmetries $SU_c(2)$ (full symmetry) and $U_c(1)$ (broken symmetry). Besides chiral symmetries, the d-wave superconductor (dSC) also has full spin rotational symmetry, and the spin-density wave (SDW) has superconducting U(1) and spin rotational symmetry around one of the axes. Near and left of the underdoped transition point the system is proposed to be an extremely weak SDW, which becomes reinforced by the electron repulsion, and which continuously evolves into a stronger antiferromagnet near half-filling.

The dSC→SDW quantum phase transition is an example of spontaneous breaking of continuous global symmetry in Eq. (1), which for a lack of better name I will call "chiral" throughout the paper. Chiral symmetry breaking is a well-studied field-theoretic phenomenon, believed to be inextricably linked to confinement in QED_3.[20] Massless QED_3 for single species of Dirac fermions has the continuous U(2) $=U(1)\times SU_c(2)$ symmetry, with the generators I,γ_3,γ_5, and $\gamma_{35}=i\gamma_3\gamma_5$, respectively. In the action in Eq. (1), the U(1) factor represents the residual spin rotational symmetry left by the choice of representation, as will be explained in detail later. It is the additional $SU_c(2)$ symmetry per Dirac component in QED_3 that will be of central interest here. The fermion mass term $m\bar{\Psi}_i\bar{\Psi}$ breaks the $SU_c(2)$ symmetry for each Dirac field to $U_c(1)$, and the two broken generators rotate between different insulating states. Chiral $SU_c(2)$ symmetry arises as an approximate symmetry of the dSC only at low energies, and will be manifestly broken, for example, by higher-order derivatives omitted in Eq. (1). It should not be confused with the spin rotational symmetry, which is, of course, also and exactly present in the dSC. Higher-order derivatives and the electron interaction terms reduce the $SU_c(2)$ to its $U_c(1)$ subgroup, which is related to the spatial translations of the original electrons. The identification of the approximate chiral symmetry in the dSC is essential for establishing the connection between the antiferromagnetic and the superconducting phases advocated in this paper, and represents one of the central results. The idealized cuprate phase diagram may be understood in terms of the chiral symmetries of different states as depicted in Fig. 1.

Assuming the scale for the SDW transition $T_{SDW}(x)$ in an anisotropic quasi-two-dimensional high-temperature superconductor to be set by the magnitude of the staggered magnetization at $T=0$,[21] the present work suggests that near and left of the superconductor-insulator transition one should expect T_{SDW} to be considerably lower than the superconducting $T_c(x)$ near and right of the critical point: $T_{\mathrm{SDW}}(x_u-\delta) \ll T_c(x_u+\delta)$, where x_u is the critical doping for the dSC-SDW transition, and $\delta \ll 1$ (see Fig. 1). This is because generalized QED$_3$ with N fermion species has a critical point at $N=N_c\approx 3$, above which there is no dynamical mass generation.[20] QED$_3$ in Eq. (1) has $N=2$ components, which together with some numerical factors gives very weak SDW order near the superconducting phase. The pseudogap phase in cuprates at $T=0$ is therefore proposed here to be actually an extremely fragile SDW, likely to be easily destroyed by disorder, for example. As half-filling is approached and the vortex loop condensate $\langle\Phi\rangle$ increases, the repulsion between electrons also becomes important. Short-range repulsion is represented in QED$_3$ by a particular quartic term, which if weak is irrelevant in the superconducting state, but which also manifestly breaks the chiral symmetry of the low-energy theory. I show that the effect of such a term is first to break the degeneracy among states with broken chiral symmetry in favor of the SDW, and then to dramatically increase the SDW order parameter farther from the dSC. The picture implied by the QED$_3$ is qualitatively in accord with the generic phase diagram for the underdoped cuprates, where the antiferromagnetic transition near half-filling raises to $\sim$300 K, but is typically unobservably low very near the superconducting state.

Neutral spinons, which are well-defined quasiparticles in the superconducting state, in the insulator become broad excitations with the lifetime proportional to the antiferromagnetic order parameter. At $T=0$ and at large distances they become confined by a logarithmic potential provided by the gauge field in the presence of the chiral symmetry breaking. Due to the weakness of the SDW order very near the superconducting transition, however, spinon confinement is effective only at very large distances, or equivalently, at very low temperatures. The weak SDW phase therefore appears effectively deconfined at intermediate length scales. The finite-T pseudogap phase has gapless spinons strongly scattered by the massless gauge field, in qualitative agreement with the broad spectral features of the electrons seen in ARPES.[18] Near half-filling the SDW order increases and the bound state of spinons rapidly shrinks, leaving only magnons in the excitation spectrum.

The confined nature of the standard antiferromagnet close to half-filling, if postulated, by itself already points to the QED$_3$ as a viable candidate for the effective theory of underdoped cuprates. If one views the superconducting state as being spin-charge separated,[7] one needs a mechanism by which spinons would eventually become confined in the antiferromagnetic phase. QED$_3$ provides such a mechanism automatically, since the massless gauge field mediates a long-range logarithmic interaction between the spinons that binds them at all energies. Were the gauge field massive, on the other hand, the physics would be equivalent to Z_2 gauge theory, and the antiferromagnetic state would be deconfined and quite different from the usual antiferromagnet.[22–24] The very existence of an ordinary antiferromagnet at, and presumably near, half-filling[25] may therefore be taken as evidence in favor of the type of theory presented in this paper.

The physical picture of the antiferromagnetic (SDW) insulator as a phase-disordered d-wave superconductor is further supported by the ARPES data on insulating $Ca_2CuO_2Cl_2$ and $Sr_2CuO_2Cl_2$.[8,9] These experiments show two unexpected features of the insulating state: (1) although the ARPES spectral function is broad, one can nevertheless identify a remnant of the Fermi surface; (2) the dispersion at such an approximate ''Fermi surface'' has a d-wave form, except that it becomes rounded and without the characteristic cusp at low energies. The ''relativistic'' dispersion for broad quasiparticle excitations that QED$_3$ implies in the insulating state, when measured from the lowest energy given by the dynamically generated chiral mass, provides a very good fit to the data (see Fig. 5). The present theory implies that the rounding of the dispersion is controlled by the size of the sublattice magnetization, and therefore should decrease with doping, as one approaches the superconducting state. It would be desirable to test this prediction in future experiments.

In the body of the paper I develop the above picture in detail. In the next section, I derive the Dirac representation of the Hamiltonian for low-energy nodal quasiparticles, and discuss the coupling to quantum fluctuating vortex loops in Sec. III. A derivation of the dynamics of the gauge field starting from the XY model on a lattice is presented in Sec. IV. This section is somewhat technical and may be skipped at first reading. Instead, the reader may consult Appendix B, where a simpler derivation for finite temperatures is presented. Dynamical breaking of chiral symmetry and the formation of the SDW state is discussed in Sec. V. More general discussion of chiral symmetry and other ordered states on the chiral manifold is provided in Sec. VI. The reduction of chiral symmetry by the irrelevant terms is discussed in Sec. VII, and the mean-field theory of the antiferromagnetic instability of QED$_3$ in the presence of electron repulsion is proposed in Sec. VIII. Confinement of spinons in the insulator is discussed in Sec. IX. The discussion of the ARPES measurements is given in Sec. X. A summary of the main results and discussion of the relations to other theoretical approaches are given in the concluding section. I finish with a list of open problems. Technical details are presented in five appendices.

II. DIRAC THEORY FOR NODAL EXCITATIONS

I begin by assuming that the superconducting state, except from being a d wave, otherwise exhibits the standard BCS phenomenology. In particular, I assume that the quasiparticles are well-defined, long-lived excitations. Generally, the quasiparticle action at $T\neq 0$ may then be taken to be

$$S=T\sum_{\vec{k},\sigma,\omega_n}\Big[(i\omega_n-\xi_{\vec{k}})c^{\dagger}_{\sigma}(\vec{k},\omega_n)c_{\sigma}(\vec{k},\omega_n) - \frac{\sigma}{2}\Delta(\vec{k})c^{\dagger}_{\sigma}(\vec{k},\omega_n)c^{\dagger}_{-\sigma}(-\vec{k},-\omega_n)+\mathrm{h.c.}+O(c^4)\Big], \tag{2}$$

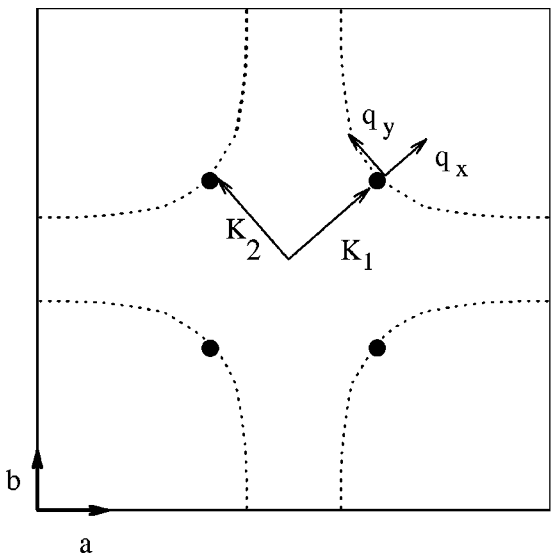

FIG. 2. The wave vectors $\vec{K}_i$, $i=1,2$, and $\vec{q}$. The dashed line stands for the putative Fermi surface. The SDW ordering wave vectors are $\vec{Q}_i=2\vec{K}_i$.

where $\Delta(\vec{k})$ has the usual d-wave symmetry, and two spatial dimensions (2D) are assumed. c and $c^\dagger$ are the electron operators, $\sigma=\pm$ labels the z projection of electron spin, and ω_n are the fermionic Matsubara frequencies. Units are chosen so that $h=c=e=1$. The $O(c^4)$ term stands for all possible short-range interactions between quasiparticles.

We may represent the quasiparticle Hamiltonian in terms of two *four-component* fields,

$$\Psi_i'^\dagger(\vec{q},\omega_n)=(c_+^\dagger(\vec{k},\omega_n),c_-(-\vec{k},-\omega_n), c_+^\dagger(\vec{k}-\vec{Q}_i,\omega_n),c_-(-\vec{k}+\vec{Q}_i,-\omega_n)), \quad (3)$$

where $\vec{Q}_i=2\vec{K}_i$ is the wave vector that connects the nodes within the diagonal pair $i=1,2$, as in Fig. 2. For spinor 1, $\vec{k}=\vec{K}_1+\vec{q}$, with $|\vec{q}|\ll|\vec{K}_1|$, and analogously for the second pair. The construction of the four-component field is not unique. The choice in Eq. (3) differs from the one made in the Ref. 7 for example. I postpone the discussion of the alternative construction used there for Appendix D. Using the construction in the Eq. (3), and by observing that $\xi_{\vec{k}}=-\xi_{\vec{k}-\vec{Q}_i}$, and $\Delta_{\vec{k}}=-\Delta_{\vec{k}-\vec{Q}_i}$, for $\vec{k}\approx\vec{K}_i$, and then by linearizing the spectrum as $\xi_{\vec{k}}=v_f q_x+O(q^2)$ and $\Delta_{\vec{k}}=v_\Delta q_y+O(q^2)$, one arrives at the low-energy action

$$S[\Psi']=\int d^2\vec{r}\int_0^\beta d\tau \Psi_1'^\dagger[\partial_\tau+M_1 v_f\partial_x+M_2 v_\Delta\partial_y]\Psi_1' +(1\to 2,x\leftrightarrow y)+O(\partial\Psi'^\dagger\partial\Psi',\Psi'^4), \quad (4)$$

with $\beta=1/T$. The continuous Dirac field $\Psi_i'(\vec{r},\tau)$ is defined as

$$\Psi_i'(\vec{r},\tau)=T\sum_{\omega_n}\int\frac{d^2\vec{q}}{(2\pi)^2}e^{i\omega_n\tau+i\vec{q}\cdot\vec{r}}\Psi_i'(\vec{q},\omega_n), \quad (5)$$

with the integral over momenta performed over $|\vec{q}|<\Lambda<T^*$. The 4×4 matrices in Eq. (4) are $M_1=i\sigma_3\otimes\sigma_3$ and $M_2=-i\sigma_3\otimes\sigma_1$. $\vec{\sigma}$ are the usual Pauli matrices, and the coordinate system has been rotated as in Fig. 2.

To cast the theory in Dirac form we may invoke the matrix $\gamma_0=\sigma_1\otimes I$, where I is the 2×2 unit matrix. Then $\gamma_0^2=I\otimes I$, and $M_i=\gamma_0\gamma_i$, with $\gamma_1=\sigma_2\otimes\sigma_3$, and $\gamma_2=-\sigma_2\otimes\sigma_1$. $\{\gamma_\nu,\gamma_\mu\}=2\delta_{\nu\mu}$, $\nu,\mu=0,1,2$, so the γ matrices indeed satisfy the Clifford algebra. The quasiparticle action (2) at *low energies* becomes equivalent to the field theory

$$S[\Psi']=\int d^2\vec{r}\int_0^\beta d\tau\,\bar{\Psi}_1'[\gamma_0\partial_\tau+\gamma_1 v_f\partial_x+\gamma_2 v_\Delta\partial_y]\Psi_1' +(1\to 2,x\leftrightarrow y)+O(\partial\bar{\Psi}'\partial\Psi',\Psi'^4), \quad (6)$$

where $\bar{\Psi}_i'=\Psi_i'^\dagger\gamma_0$. Weak quartic interactions, as long as they are short-ranged, are irrelevant by simple power counting. This simply reflects the severe phase-space restrictions for scattering of the nodal quasiparticles. I will therefore omit them temporarily, together with the second-order derivative terms, and return to their effects in Sec. VII.

The reader would be correct to note that there is a considerable freedom in selecting the form of the matrix γ_0. In fact, any 4×4 matrix that anticommutes with M_1 and M_2 and squares to the unit matrix would yield an equally valid Dirac representation. It is shown later that this freedom will correspond to different "directions" in the space of ordered states with broken chiral symmetry. The specific choice for γ_0 made here will be analogous to choosing a direction in real space along which to search for a finite magnetization, for example, in the more familiar magnetic phase transitions.

III. COUPLING TO TOPOLOGICAL DEFECTS

The goal in this section will be to find the most economical form of the coupling between nodal excitations in the dSC and the fluctuations of the phase of the superconducting order parameter. The working assumption is that the amplitude fluctuations are frozen well below the pseudogap temperature T^*, so it is only the phase degree of freedom that remains active at low energies. With this in mind I write

$$v_\Delta\to v_\Delta(\vec{r},\tau)=|v_\Delta|e^{i[\phi_s(\vec{r},\tau)+\phi_r(\vec{r},\tau)]}, \quad (7)$$

where ϕ_r represents the regular ("spin-wave") part of the order parameter phase, and ϕ_s is the singular contribution due to topological defects. At $T=0$ these would be the vortex loops[17] or the more familiar vortices and antivortices at $T\neq 0$. At this point it is tempting to transform both spin-up and spin-down fermionic operators by absorbing a half of the total superconducting phase into each. In the presence of topological defects, however, this would lead to multivalued fermionic fields and would not be a local change of variables in the partition function. This problem may be circumvented by allowing only vortices of double vorticity[7] for example, which then leads to the Z_2 gauge theory representation of the problem, and a possibility of spin-charge separation in the pseudogap regime.[24] It is the single vortices, however, that first become relevant at the $T\neq 0$ Kosterlitz-Thouless transition,[26] and they should be included in the description of the $T=0$ transition as well. I will therefore utilize the idea of Franz and Tešanović,[27,28] who suggested dividing a given vortex configuration into two groups *A* and *B*, and transforming the electron operators with spin up and spin down differently. We write

$$\phi_A(\vec{r},\tau)=\frac{\phi_r(\vec{r},\tau)}{2}+\phi_{sA}(\vec{r},\tau), \tag{8}$$

and similarly for *B*. ϕ_{sA} is the piece of the singular part of the phase that comes from the defects grouped in *A*. One may then make a *local* change of variables by introducing a new Dirac field Ψ as

$$\Psi(\vec{r},\tau)=U(\vec{r},\tau)\Psi'(\vec{r},\tau), \tag{9}$$

where $U=\mathrm{diag}\{e^{-i\phi_A},e^{i\phi_B},e^{-i\phi_A},e^{i\phi_B}\}$. Since any given vortex defect is either in group *A* or *B*, and therefore associated with either up or down spin by the transformation (9), circling around it with the transformed fermion would yield either 2π or zero of the accumulated phase change. Components of the new field Ψ are therefore single-valued functions.

The gauge-transformed action for the Dirac field Ψ is then

$$\begin{aligned} S[\Psi']\rightarrow S[\Psi,\vec{a},\vec{v}] \\ =\int d^2\vec{r}\int_0^\beta d\tau\overline{\Psi}_1[\gamma_0(\partial_\tau+ia_0) \\ +\gamma_1 v_f(\partial_x+ia_x)+\gamma_2|v_\Delta|(\partial_y+ia_y)]\Psi_1 \\ +(1\rightarrow 2,x\leftrightarrow y)+iv_\mu J_\mu, \end{aligned} \tag{10}$$

with $a_\nu=\partial_\nu(\phi_A-\phi_B)/2$, $v_\nu=\partial_\nu(\phi_A+\phi_B)/2$, and $J_\nu=(\Psi_i^\dagger(I\otimes\sigma_3)\Psi_i,v_F\Psi_1^\dagger(\sigma_3\otimes I)\Psi_1,v_F\Psi_2^\dagger(\sigma_3\otimes I)\Psi_2)$. Since the vector J_ν is built out only of the products of the creation and the annihilation operators with same spin, it also represents the *physical* charge current carried by the quasiparticles. On the other hand, since the regular part of the phase ϕ_r was in Eq. (8) divided equally between spin up and spin down, the Dirac field Ψ is invariant under a regular gauge transformation. Components of Ψ therefore create electrically *neutral* excitations with spin 1/2,[7] which may therefore be referred to as *spinons*.

The action (10) has two rather different gauge symmetries, and it may be worthwhile pausing a little to reflect on them. First, the physical electromagnetic gauge field A_μ would enter the action (10) by the replacement $v_\mu\rightarrow v_\mu+A_\mu$, and couple to the charge current. Under a regular gauge transformation $A_\mu\rightarrow A_\mu+\partial_\mu\chi$, the Volovik field[29] $v_\mu\rightarrow v_\mu-\partial_\mu\chi$, while the gauge field a_μ and Ψ remain the same. The action (10) is therefore gauge invariant, in the standard sense. But it also has an additional internal gauge symmetry, under the transformation $a_\mu\rightarrow a_\mu+\partial_\mu\chi$, $v_\mu\rightarrow v_\mu$, $\Psi\rightarrow e^{-i\chi}\Psi$. This reflects the freedom of choice in Eq. (8); one could have equally well chosen the regular part of ϕ_A to be $(\phi_r/2)+\chi$, and of ϕ_B to be $(\phi_r/2)-\chi$. One deals with this gauge freedom, as usual, by eventually introducing the gauge-fixing term for a_μ that allows one to freely sum over all regular internal gauges χ. Similarly, the division of the singular part of the superconducting phase into that which comes from the defects in group *A* and the defects in group *B* is equally arbitrary. Just as one effectively sums over all *regular* internal gauges by the introduction of the gauge-fixing term, we will sum over all *singular* internal gauges by averaging over all possible divisions of defects into two groups. This is explained in the next section, and in the Appendix B. As a by-product, the averagings over regular and singular internal gauges will ensure that up and down spinons are treated equally in QED$_3$, preserving the symmetry of the original electronic action (2).

The crucial observation about the action (10) is that the coupling of spinons to phase fluctuations is furnished by *two* U(1) fields that play quite different roles in the problem. The total superconducting phase determines the Volovik field v_ν and couples to the charge current, in the same way as the true electromagnetic field would. v_ν will therefore inevitably become *massive* once the high-energy spinons in Eq. (10) begin to be integrated out. Its fluctuations therefore may provide only a short-range interaction between spinons. The gauge field a_ν, on the other hand, enters Eq. (10) in a gauge-invariant way, and therefore is protected from acquiring a mass from spinons. Both gauge fields, however, depend on the fluctuating positions of the topological defects, and acquire their dynamics not only from the spinons, but from the defects as well. To determine their dynamics one therefore needs to integrate the defect degrees of freedom out. If a_μ would stay massless even after this integration is performed, it would mediate a long-range interaction between the nodal excitations, which, unlike the short-range quartic terms in Eq. (6), would not be made irrelevant by the phase space restrictions. This, however, depends on the precise way a_μ acquires its dynamics from the fluctuating vortex loops, to which I turn next.

IV. DYNAMICS OF THE GAUGE FIELDS

The zero-temperature partition function for the coupled system of *d*-wave quasiparticles and superconducting phase fluctuations is therefore

$$Z=\int D[\Psi,\vec{a},\vec{v}]e^{-(S[\Psi,\vec{a},\vec{v}]+S_{\mathrm{U}(1)}[\vec{a},\vec{v}])}, \tag{11}$$

with $S[\Psi,\vec{a},\vec{v}]$ defined by the Eq. (10), and with $S_{\mathrm{U}(1)}[\vec{a},\vec{v}]$ to be derived by integrating out the phase fluctuations. For simplicity, I will assume that these may be described the $(2+1)$-dimensional *XY* model. The bare stiffness for the phase fluctuations will be assumed to be provided by the high-energy modes that have been integrated out in arriving at the low-energy theory. Our goal will be then to rewrite the partition function for the *XY* model as the functional integral over the fields $\vec{a}$ and $\vec{v}$. In particular, we want to integrate over the topological defects implicit in the *XY* model.

I first discretize the space and the imaginary time in writing the partition function of the *XY* model. This is done to facilitate a more rigorous treatment of the topological defects, and it will prove possible to return to the continuum description we employed until now. On a lattice, in the standard lattice gauge-theory notation[17]

$$Z_{xy}=\int_0^{2\pi}\left(\prod_i d\phi_i\right)\exp\left(K\sum_{i,\hat{\mu}=\hat{x},\hat{y},\hat{\tau}}\cos(\phi_{i+\hat{\mu}}-\phi_i)\right), \tag{12}$$

where the index i labels the sites of a $(2+1)$-dimensional lattice, and $\hat{x}$ is the lattice unit vector in the x direction, for example. For simplicity, full isotropy in the XY model is assumed. Using the Villain approximation[30] and then integrating over the phases leads to

$$Z=\int_{-\infty}^{\infty}d\vec{s}\sum_{\vec{n}}{}'\exp\left(-\frac{1}{2K}\sum_i(\vec{\nabla}\times\vec{s}_i)^2+i2\pi\sum_i\vec{n}_i\cdot\vec{s}_i\right), \tag{13}$$

where $\vec{n}_i=(n_{i,\tau},n_{i,x},n_{i,y})$ is an integer vortex-loop vector variable, satisfying the constraint $\vec{\nabla}\cdot\vec{n}_i=0$ (indicated with the prime on the sum). $\vec{\nabla}$ and $\vec{\nabla}\times$ should be understood as the lattice gradient and the curl, respectively. Summing over $\vec{n}_i$ forces $\vec{s}_i$ to take integer values, and the above expression becomes the standard current representation of the XY model.[17]

Next, I imagine dividing a given configuration of vortex loops into two arbitrary groups, and write $\vec{n}_i=\vec{n}_{A,i}+\vec{n}_{B,i}$, with $\vec{\nabla}\cdot\vec{n}_{A,i}=\vec{\nabla}\cdot\vec{n}_{B,i}=0$. We will want to sum over all integer $\vec{n}_{A,i}$ and $\vec{n}_{B,i}$, in order to average over all possible divisions of vortices into two groups. Introducing the lattice version of the fields $\vec{a}_i$ and $\vec{v}_i$ as $\vec{B}_i+\vec{b}_i=2\pi\vec{n}_{A,i}$, $\vec{B}_i-\vec{b}_i=2\pi\vec{n}_{B,i}$, where $\vec{b}_i=\vec{\nabla}\times\vec{a}_i$ and $\vec{B}_i=\vec{\nabla}\times\vec{v}_i$, I write[31]

$$\begin{aligned}Z_{xy}=\int_{-\infty}^{\infty}d[\vec{a},\vec{v},\vec{t},\vec{s},\vec{r}]\sum_{\vec{n}_A,\vec{n}_B}{}'\ \exp-\sum_i\Bigg[&\frac{1}{2K}(\vec{\nabla}\times\vec{s}_i)^2\\&+i2\pi\vec{s}_i\cdot(\vec{n}_{A,i}+\vec{n}_{B,i})+i\vec{t}_i\cdot(\vec{B}_i+\vec{b}_i-2\pi\vec{n}_{A,i})\\&+i\vec{r}_i(\vec{B}_i-\vec{b}_i-2\pi\vec{n}_{B_i})\Bigg].\end{aligned} \tag{14}$$

The summations over $\vec{n}_{A,i}$ and $\vec{n}_{B,i}$ then enforce the constraints $\vec{s}_i-\vec{t}_i=\vec{m}_{A,i}$, and $\vec{s}_i-\vec{r}_i=\vec{m}_{B,i}$, where $\vec{m}_{A,i}$ and $\vec{m}_{B,i}$ are new integers. Performing the Gaussian integrals over $\vec{s}_i$, yields

$$\begin{aligned}Z_{xy}=\int_{-\infty}^{\infty}d[\vec{a},\vec{v}]\sum_{\vec{m}_A,\vec{m}_B}{}'\ \exp\Bigg(-\sum_i\{&2K\vec{v}_i^{\,2}\\&+i\vec{v}_i\cdot[\vec{\nabla}\times(\vec{m}_{A,i}+\vec{m}_{B,i})]\\&+i\vec{a}_i\cdot[\vec{\nabla}\times(\vec{m}_{A,i}-\vec{m}_{B,i})]\}\Bigg).\end{aligned} \tag{15}$$

This can be further simplified by noticing that the action is quadratic in the Volovik field $\vec{v}$, which can also be integrated out. In doing so I will neglect the additional coupling of $\vec{v}$ to the charge current $\vec{J}$ in the Eq. (10), which only leads to additional irrelevant interaction between spinons. The integration over $\vec{v}_i$ in the last equation then gives

$$\begin{aligned}Z_{xy}=\int_{-\infty}^{\infty}d\vec{a}\sum_{\vec{m}_A,\vec{m}_B}{}'\ \exp\Bigg[-\sum_i\Bigg(&\frac{1}{8K}[\vec{\nabla}\times(\vec{m}_{A,i}+\vec{m}_{B,i})]^2\\&+i\vec{a}_i\cdot[\vec{\nabla}\times(\vec{m}_{A,i}-\vec{m}_{B,i})]\Bigg)\Bigg].\end{aligned} \tag{16}$$

Integrating over $\vec{a}_i$ in Eq. (16) would give back the current representation of the XY model, Eq. (13). Alternatively, we can introduce the real variables $\vec{\Phi}_{+,i}$ and $\vec{\Phi}_{-,i}$ and write

$$\begin{aligned}Z_{xy}=\int_{-\infty}^{\infty}d[\vec{a},\vec{\Phi}_-,\vec{\Phi}_+]\sum_{\vec{l}_A,\vec{l}_B}{}'\exp\Bigg[-\sum_i\Bigg(&\frac{1}{8K}(\vec{\nabla}\times\vec{\Phi}_{+,i})^2\\&+i\vec{a}_i\cdot(\vec{\nabla}\times\vec{\Phi}_{-,i})+i2\pi(\vec{l}_{A,i}\cdot\vec{\Phi}_{A,i}+\vec{l}_{B,i}\cdot\vec{\Phi}_{B,i})\Bigg)\Bigg],\end{aligned} \tag{17}$$

where $\vec{\Phi}_{+,-,i}=\vec{\Phi}_{A,i}\pm\vec{\Phi}_{B,i}$. The summations over the auxiliary link variables $\vec{l}_{A,B}$ force $\vec{\Phi}_A$ and $\vec{\Phi}_B$ and therefore $\vec{\Phi}_+$ and $\vec{\Phi}_-$ to be integers. To preserve the gauge invariance $(\Phi_{+,i,\mu}\rightarrow\Phi_{+,i,\mu}+\nabla_\mu\chi_i,\ \ \Phi_{-,i,\mu}\rightarrow\Phi_{-,i,\mu}+\nabla_\mu\phi_i)$ of the last expression we must impose $\vec{\nabla}\cdot\vec{l}_{A,i}=\vec{\nabla}\cdot\vec{l}_{B,i}=0$.[32,33] We may next add a small chemical potential for the link variables $\vec{l}_{A,B}$ to the action in Eq. (17) as the term $x\Sigma_i(\vec{l}_{A,i}^{\,2}+\vec{l}_{B,i}^{\,2})$. Up to the Villain approximation, the last expression is then equal to

$$\begin{aligned}Z_{xy}=\lim_{x\rightarrow 0}\int_{-\infty}^{\infty}&d[\vec{a},\vec{\Phi}_A,\vec{\Phi}_B]\int_0^{2\pi}d[\theta_A,\theta_B]\\&\times\exp\Bigg[-\sum_i\Bigg(\frac{1}{8K}(\vec{\nabla}\times\vec{\Phi}_{+,i})^2+i\vec{a}_i\cdot(\vec{\nabla}\times\vec{\Phi}_{-,i})\\&-\frac{1}{2x}\cos(\theta_{A,i}-\theta_{A,i+\hat{\nu}}-2\pi\Phi_{A,i,\hat{\nu}})\\&-\frac{1}{2x}\cos(\theta_{B,i}-\theta_{B,i+\hat{\nu}}-2\pi\Phi_{B,i,\hat{\nu}})\Bigg)\Bigg],\end{aligned} \tag{18}$$

where I introduced two sets of "dual" angles $\theta_{A,i}$ and $\theta_{B,i}$ to ensure the gauge invariance, and imposed the "frozen" limit $x\rightarrow 0$. The integration over $\vec{a}_i$ in the Eq. (18) together with the frozen limit ultimately sets $\theta_{A,i}\equiv\theta_{B,i}$, so the last equation becomes another representation of the frozen lattice superconductor (FLS), which is well known to be dual to the XY model in three dimensions.[34,35]

In principle, one would like to integrate out all the fields other than $\vec{a}$ in the Eq. (18), to be left with the effective action $S_{\mathrm{U}(1)}[\vec{a}]$ for $\vec{a}$ only. The result would be an interacting theory for $\vec{a}$, which can be expanded in powers of $\vec{a}$, for example. Instead of doing this, I will approximate the $S_{\mathrm{U}(1)}[\vec{a}]$ with the effective Gaussian action for $\vec{a}$, that reproduces the gauge-field propagator in the full theory (18). This approximation may be understood as the self-consistent mean field theory for $\vec{a}$, with the effect of integration over all other fields in Eq. (18) lumped into the form of the propagator.

In this approximation the problem of dynamics of the gauge field $\vec{a}$ reduces to the computation of the two-point correlation function for $\vec{a}$ from the representation of the XY model in Eq. (18). I therefore introduce the source term into the last expression by adding $i\Sigma_i \vec{j}_i \cdot (\vec{\nabla}\times\vec{a}_i)$ to the exponent. Then

$$\langle(\vec{\nabla}\times\vec{a})_{i,\nu}(\vec{\nabla}\times\vec{a})_{j,\mu}\rangle = \frac{\partial^2}{\partial j_{i,\nu}\partial j_{j,\mu}} \ln Z_{xy}|_{\vec{j}\equiv\vec{0}}. \tag{19}$$

It is convenient then to integrate over $\vec{a}$ in the Z_{xy} first. One finds

$$\langle(\vec{\nabla}\times\vec{a})_{i,\nu}(\vec{\nabla}\times\vec{a})_{j,\mu}\rangle = \delta_{i,j}\delta_{\nu,\mu}\lim_{x\to 0}\frac{\pi^2}{x}\langle\cos(\theta_i-\theta_{i+\hat{\nu}}-2\pi\Phi_{i,\nu})\rangle_{\mathrm{FLS}}, \tag{20}$$

where the last average is to be taken over the configurations of the FLS:

$$Z_{xy} = \lim_{x\to 0}\int d[\vec{\Phi},\theta]\exp\left[-\sum_{i,\hat{\nu}}\left(\frac{1}{2K}(\vec{\nabla}\times\vec{\Phi})^2 - \frac{1}{x}(\cos(\theta_i-\theta_{i+\hat{\nu}}-2\pi\Phi_{i,\hat{\nu}}))\right)\right]. \tag{21}$$

It is well established that the lattice superconductor at a small but finite "temperature" x has a phase transition as K is varied in the same universality class as in the frozen limit $x=0$.[17,34,36] We may therefore relax the constraint $x\to 0$ with impunity and assume x to be finite. The average that appears on the right-hand side of the Eq. (20) can then be computed, for example, by using the mean-field approximation to the FLS action (21) (see Appendix A). This yields

$$\frac{1}{x}\langle\cos(\theta_i-\theta_{i+\hat{\nu}}-2\pi\Phi_{i,\nu})\rangle_{\mathrm{FLS}} \propto |\langle\exp(i\theta_i)\rangle|^2. \tag{22}$$

This result is quite general, and it simply expresses the fact that in the ordered phase of the theory (21) the dual angles become correlated, while at the same time the gauge field becomes massive via Meissner effect. The gauge-field fluctuations can then be neglected, which makes the requisite average finite when the dual angles θ order, i.e., in the *disordered* phase of the original XY model.

Returning to the continuum notation, and switching to the Fourier space, the gauge-invariant expression for the correlation function (19) at low momenta is therefore

$$\langle(\vec{\nabla}\times\vec{a})_\nu(\vec{\nabla}\times\vec{a})_\mu\rangle \propto [|\langle\Phi\rangle|^2+O(q^2)](\delta_{\mu\nu}-\hat{q}_\mu\hat{q}_\nu), \tag{23}$$

where I allowed, in general, for some momentum dependence [the term $O(q^2)$]. The $O(q^2)$ term should be expected to appear in a more sophisticated approximation for the gauge-field dynamics than provided by the Eq. (20). To the lowest order, the integration over all other fields in Eq. (18) effectively yields the *Maxwell term* for the gauge field $\vec{a}$, with the stiffness inversely proportional to the expectation value of the *dual* loop condensate $\langle\Phi\rangle\sim\langle e^{i\theta}\rangle$ that reflects the phase of the XY model. This is the main result of this section. When the dSC is phase coherent and the vortex loops are finite in size, $\langle\Phi\rangle=0$, and $\vec{a}$ is infinitely stiff, and in a first approximation may be considered decoupled from spinons. When vortex loops blow up, $\langle\Phi\rangle\neq 0$, phase coherence is lost, and the spinons are minimally coupled to a massless gauge field. This is in agreement with the physical arguments advanced in Ref. 18.

At high temperature one can neglect the fluctuations in the imaginary time direction and deal with the purely 2D problem of point vortices and antivortices. This simplifies the analysis in that no gauge invariance needs to be ensured in the Eq. (17), so no dual angles are required.[32,33] One then ends up with the thermodynamic vortex fugacity playing the role of the dual condensate[15] and with the simpler sine-Gordon theory instead of the FLS. For an alternative derivation of the gauge-field dynamics at $T\neq 0$ and in continuum that is in full accord with the conclusions of this section I direct the reader to the Appendix B.

There is, however, an additional subtlety in going from the lattice to the continuum theory that is worth registering. The partition function for the XY model in the Eq. (17) has symmetry under $a_{i,\mu}\to a_{i,\mu}+2\pi n_{i,\mu}$, with $n_{i,\mu}$ integer, that is lost when a small chemical potential term [in passing to the Eq. (18)] is added. This implies that the summation over the integer vortex variables in Eq. (17) must yield a *compact* term for $\vec{a}$, which may be approximated with the Maxwell term, Eq. (23), in continuum. Possible effects of compactness of $\vec{a}$ on the picture developed in this paper are discussed in Sec. XII.

V. DYNAMICAL BREAKING OF CHIRAL SYMMETRY

The effective $T=0$ low-energy theory for the interacting system of d-wave quasiparticles and fluctuating vortex loops, after the integration over vortex loops is therefore

$$S[\Psi] = \int d^2\vec{r}\,d\tau\left\{\bar{\Psi}_1[\gamma_0(\partial_\tau+ia_0)+\gamma_1 v_f(\partial_x+ia_x) + \gamma_2|v_\Delta|(\partial_y+ia_y)]\Psi_1 + (1\to 2, x\leftrightarrow y) + \frac{1}{2|\langle\Phi\rangle|^2}[c^2(\vec{\nabla}\times\vec{a})^2_\tau + (\vec{\nabla}\times\vec{a})^2_r]\right\}, \tag{24}$$

where I omitted the higher derivative terms, and the terms quartic in Ψ. This is the standard QED_3, with two important caveats: (1) the coordinates x and y are exchanged for the second Dirac field, (2) there is an inherent anisotropy in the model, $v_f\neq v_\Delta\neq c$, where c is a characteristic velocity for the phase fluctuations.[37] First, let us consider the simpler isotropic limit of the theory, $v_f = v_\Delta = c$. There are sixteen 8×8 matrices then that either commute or anticommute with the three 8×8 γ matrices that appear in Eq. (24): diag$\{\gamma_0,\gamma_0\}$, diag$\{\gamma_1,\gamma_2\}$, diag$\{\gamma_2,\gamma_1\}$. First, there are eight block-diagonal Hermitean matrices

$$I\otimes I_4,\quad \sigma_3\otimes I_4,\quad I\otimes\gamma_{35},\quad \sigma_3\otimes\gamma_{35} \tag{25}$$

that commute, and

$$I\otimes\gamma_3,\quad \sigma_3\otimes\gamma_3,\quad I\otimes\gamma_5,\quad \sigma_3\otimes\gamma_5 \tag{26}$$

that anticommute with the γ matrices. Here, $\gamma_3 = \sigma_2\otimes\sigma_2$, $\gamma_5 = \sigma_3\otimes I$, $\gamma_{35} = i\gamma_3\gamma_5$, and $I_4 = I\otimes I$. Next, there are eight more block-off-diagonal Hermitean matrices

$$\sigma_1\otimes\frac{i}{\sqrt{2}}(\gamma_2-\gamma_1)\gamma_3,\quad \sigma_2\otimes\frac{i}{\sqrt{2}}(\gamma_2-\gamma_1)\gamma_3,$$

$$\sigma_1\otimes\frac{i}{\sqrt{2}}(\gamma_2-\gamma_1)\gamma_5,\quad \sigma_2\otimes\frac{i}{\sqrt{2}}(\gamma_2-\gamma_1)\gamma_5 \tag{27}$$

that commute, and

$$\sigma_1\otimes\frac{1}{\sqrt{2}}(\gamma_1-\gamma_2),\quad \sigma_2\otimes\frac{1}{\sqrt{2}}(\gamma_1-\gamma_2),$$

$$\sigma_1\otimes\frac{i}{\sqrt{2}}\gamma_0(\gamma_1+\gamma_2),\quad \sigma_2\otimes\frac{i}{\sqrt{2}}\gamma_0(\gamma_1+\gamma_2) \tag{28}$$

that anticommute with the γ matrices. I call these sixteen generators G_i, $i=1,\ldots,16$, in the above order. The isotropic QED_3 in the Eq. (24) is invariant under a global unitary transformation

$$\Psi\to U\Psi, \tag{29}$$

where

$$U=\exp\left(i\sum_{i=1}^{16}\theta_i G_i\right). \tag{30}$$

This follows immediately by observing that all the generators commute with the 8×8 matrices diag$\{\gamma_0,\gamma_0\}$diag$\{\gamma_1,\gamma_2\}$ and diag$\{\gamma_0,\gamma_0\}$diag$\{\gamma_2,\gamma_1\}$ by construction. The unitary transformations in relation (29) can be shown to form the Lie group U(4). Following the standard terminology in the field theory literature, I will refer to this symmetry of QED_3 as "chiral."

As a first step towards understanding of the meaning of the chiral symmetry in the present context, it will prove useful to consider how it may be broken. QED_3 is well known to have the chiral symmetry spontaneously broken[20] by the dynamical generation of the mass term in the action (24):

$$m\int d^2\vec{r}\,d\tau\sum_{i=1}^{2}\bar{\Psi}_i\Psi_i, \tag{31}$$

with $m\propto|\langle\Phi\rangle|^2$, i.e., proportional to the effective charge of QED_3. Containing just a single γ matrix, the mass term in the Eq. (31) breaks all the *anticommuting* generators, G_i with $i=5,6,7,8,13,14,15,16$. The chiral symmetry is reduced from U(4) to U(2)$\times$U(2), with eight generators preserved. The fermion mass is generated dynamically due to the coupling to the gauge field. To see this, neglect the wave-function renormalization and the vertex corrections (which can be rationalized in the limit of a large number of Dirac fields N), and write the self-energy as

$$\Sigma(q)=|\langle\Phi\rangle|^2\gamma_\nu\int\frac{d^3\vec{p}}{(2\pi)^3}\frac{D_{\nu\mu}(\vec{p}-\vec{q})\Sigma(p)}{p^2+\Sigma^2(p)}\gamma_\mu, \tag{32}$$

where $\vec{q}=(\omega,q_x,q_y)$. The gauge-field propagator in the transverse (Landau) gauge is

$$D_{\nu\mu}(\vec{p})=(\delta_{\nu\mu}-\hat{p}_\nu\hat{p}_\mu)/[p^2+\Pi(p)], \tag{33}$$

where $\Pi(p)$ is the self-consistently computed polarization. At $p\ll\Sigma(0)=m$, assuming a finite mass m gives

$$\Pi(p)=\frac{N|\langle\Phi\rangle|^2}{6\pi}\frac{p^2}{m}+O(p^4). \tag{34}$$

For the polarization at all momenta see Appendix C. Equation (32) was analyzed in Ref. 20 (see also Appendices C and E), and there is a solution with finite m for the number of Dirac fields $N<N_c=32/\pi^2=3.24$. Full numerical solution that includes the wave-function renormalization and vertex corrections confirms that $N_c\approx 3$,[38] almost independently of the choice of vertex. Lattice simulations give $3<N_c<4$,[39] or at least $N_c>2$.[40] It therefore seems reasonable to conclude that for $N=2$ the chiral symmetry in the isotropic QED_3 becomes spontaneously broken when the vortex loops unbind and $\langle\Phi\rangle\neq 0$.

Since the matrix γ_0 commutes with the electron-spinon transformation in the Eq. (9), it is easy to rewrite the mass term in QED_3 in terms of the original electron operators:

$$\begin{aligned} m\sum_{i=1}^{2}\bar{\Psi}_i\Psi_i\to mT\sum_{\vec{k}\approx\vec{K}_1,\omega_n}&\{[c_+^\dagger(\vec{k},\omega_n)c_+(\vec{k}-\vec{Q}_1,\omega_n)\\ &-c_-^\dagger(-\vec{k}+\vec{Q}_1,\omega_n)c_-(-\vec{k},\omega_n)]\\ &+[c_+^\dagger(\vec{k}-\vec{Q}_1,\omega_n)c_+(\vec{k},\omega_n)-c_-^\dagger(-\vec{k},\omega_n)\\ &\times c_-(-\vec{k}+\vec{Q}_1,\omega_n)]\}+(1\to 2). \end{aligned} \tag{35}$$

The reader will recognize this as the low-energy part of the *staggered* potential along the spin z axis

$$m\int d^2\vec{r}\,d\tau\sum_{\sigma=\pm,i=1,2}\cos(\vec{Q}_i\cdot\vec{r})\sigma c_\sigma^\dagger(\vec{r},\tau)c_\sigma(\vec{r},\tau), \tag{36}$$

so the mass in QED_3 is nothing but the spontaneously generated SDW order parameter. The periodicity of the SDW is set by the vectors $\vec{Q}_i$, and thus tied to the Fermi surface. The

SDW order is established as soon as the phase coherence is lost, and the charge $\langle\Phi\rangle\neq 0$. In the large-N approximation[20] one finds that

$$m\approx 16|\langle\Phi\rangle|^2\exp[-2\pi/\sqrt{(N_c/N-1)}]. \tag{37}$$

Since $N_c\approx 3$, for $N=2$ one finds that $m\sim 10^{-2}|\langle\Phi\rangle|^2$. This extreme "lightness" of fermions in QED_3 derives from the fact that the mass comes solely from the interaction with the soft gauge field.

Breaking of chiral symmetry in QED_3 also implies that the energies of spinons have become complex and finite in the phase incoherent state with $\langle\Phi\rangle\neq 0$. In the simplest approximation the electron propagator may be computed as a product of the spinon and the gauge-field factors, so a spinon "gap" should imply a charge gap as well, i.e., the system becomes an insulator.[41] In Sec. IX I discuss how spinons should actually be confined in the insulating state. Staggered magnetization, charge gap, and the spinon confinement when taken together imply that the state with broken chiral symmetry is nothing but the standard, albeit a weak, SDW. It seems natural to assume then that this state is continuously connected to the antiferromagnet near half-filling in cuprates. This expectation is further corroborated by considering the effect of Coulomb interactions, which is done in Sec. VIII.

It has been already mentioned that we have some freedom in choosing the representation of the γ matrices. In particular, it was the specific choice of γ_0 that led to the cos-SDW order parameter displayed in Eqs. (35) and (36). In the next section I discuss how "rotating" the cos-SDW by the broken chiral generators leads to a different insulating states.

VI. MORE ON CHIRAL SYMMETRY: THE SPACE OF INSULATORS

In discussing the pattern of chiral symmetry breaking in QED_3 one needs to distinguish at least two different cases. The isotropic theory ($v_\Delta=v_f$) has the full U(4) symmetry in its massless phase, so the mass term breaks eight of its sixteen generators. In cuprates,[42] however, $v_f/v_\Delta\sim 10$, and even with $m=0$ the symmetry is only $\mathrm{U}(2)\times\mathrm{U}(2)$, generated by the block-diagonal G_i $i=1,\ldots,8$. How such a large anisotropy affects the value of N_c is a nontrivial problem, and is addressed in a separate publication.[43] Here I will consider only the effect of anisotropy on chiral symmetry, and assume it is reduced to $\mathrm{U}(2)\times\mathrm{U}(2)$. It suffices then to look at each Dirac component in QED_3 separately, i.e., consider just the 4×4 representation of the γ matrices, as defined right below Eq. (5).

It can be easily shown that any matrix that anticommutes with both M_1 and M_2 and squares to the unit matrix may be chosen as γ_0, and will lead to a representation of the γ matrices such as in the Eq. (6). The mass term $\sim m\Psi^\dagger\gamma_0\Psi$ in the action would then gap the quasiparticles, in analogy to the standard relativistic Dirac equation. The problem of different chiral orders is therefore nothing else but finding all the ways in which d-wave quasiparticles can spontaneously acquire such a "relativistic mass." It will be useful to introduce "directions" in the space of broken symmetry states, as a set of linearly independent matrices that anticommute with M_1 and M_2, and square to one. It is easy to show that in the 4×4 representation there are only four such matrices

$$\tilde\gamma_0,\tilde\gamma_3,\tilde\gamma_5,i\tilde\gamma_1\tilde\gamma_2, \tag{38}$$

with $\tilde\gamma_0=\gamma_0$, and where $\tilde\gamma_1=-iM_1$, $\tilde\gamma_2=iM_2$, $\tilde\gamma_3=\sigma_3\otimes\sigma_2$, and $\tilde\gamma_5=\sigma_2\otimes I$. In principle, any of these four if used instead of our γ_0 in the construction of the Dirac theory in Eq. (6) and in the mass term would give a relativistic gap to Dirac fermions. The last matrix,

$$i\tilde\gamma_1\tilde\gamma_2=I\otimes\sigma_2, \tag{39}$$

however, being a product of two $\tilde\gamma$ matrices does not break the chiral symmetry, and is believed not to be spontaneously generated in QED_3.[20,44] I therefore focus on the remaining three. Choosing one among $\{\tilde\gamma_0,\tilde\gamma_3,\tilde\gamma_5\}$ as the γ_0 matrix in the mass term reduces the $\mathrm{SU}_c(2)$ subgroup of U(2) $[=\mathrm{U}(1)\times\mathrm{SU}_c(2)]$, generated by $\{\gamma_3,\gamma_5,\gamma_{35}\}$, to $\mathrm{U}_c(1)$. The two anticommuting generators of the $\mathrm{SU}_c(2)$ that are broken then rotate the chosen order parameter towards the two remaining "directions" in the chiral space. For example, for our choice of $\tilde\gamma_0=\gamma_0$, it is γ_{35} that remains unbroken in the cos-SDW phase, whereas the broken generators rotate the cos-SDW order parameter as

$$e^{i\theta\gamma_i}\tilde\gamma_0 e^{-i\theta\gamma_i}=\cos(2\theta)\tilde\gamma_0-\sin(2\theta)\tilde\gamma_i,\quad i=3,5. \tag{40}$$

Choosing $i=5$, for example, for both Dirac fields rotates the cos-SDW in the Eq. (35) into

$$m\int d^2\vec r\,d\tau\sum_{\sigma=\pm,i=1,2}\cos(\vec Q_i\cdot\vec r+2\theta)\sigma c^\dagger_\sigma(\vec r,\tau)c_\sigma(\vec r,\tau). \tag{41}$$

Chiral rotations generated by γ_5 thus correspond to *sliding* modes of the SDW. γ_3, on the other hand, rotates $\tilde\gamma_0$ towards the direction of $\tilde\gamma_3$, which describes an additional particle-particle pairing potential between the neutral spinons, with the opposite sign for the diagonally opposed nodes. This may be understood as an additional p-wave component of pairing between the spinons, so the state described by $\tilde\gamma_3$ order parameter may be called the $d+ip$ state.[45] This state preserves the superconducting U(1) symmetry and the translational invariance, but breaks the spin-rotational invariance and is odd under parity. Since $\tilde\gamma_3$ *does not* commute with the electron-spinon transformation (9), however, the $d+ip$ state cannot be simply expressed in terms of electronic operators, as it was proved possible for the SDW states. The relationship between the directions in the order parameter space $\{\tilde\gamma_0,\tilde\gamma_3,\tilde\gamma_5\}$, and the chiral generators may be summarized pictorially as on Fig. 3.

It is instructive to look more closely at the origin of the U(2) symmetry (per Dirac component) that appears in the low-energy theory of the dSC. First, the transformations in the U(1) subgroup of $\mathrm{U}(2)=\mathrm{U}(1)\times SU_c(2)$ are analogous to the spin rotations around the z axis. To see this, consider

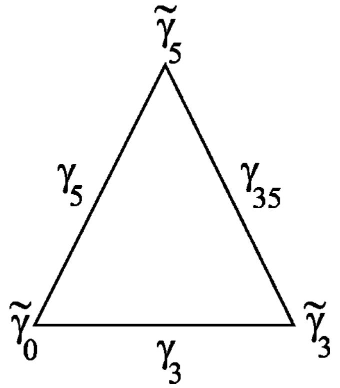

FIG. 3. The corners of the triangle represent the three chiral directions in the space of insulating states that descend from the *d*-wave superconductor. At the side opposite to a particular direction lies the corresponding unbroken chiral generator, while the remaining two broken generators rotate the chosen insulator towards two other directions.

the conserved current that corresponds to the U(1) subgroup, $J_{i,\mu}=\overline{\Psi}_i\gamma_\mu\Psi_i$, so that the conserved charge is simply the z component of the total spin of the low-energy quasiparticles, one charge per each pair of nodes. Of course, the quasiparticle action (2) has full SO(3) spin symmetry, and this is not to imply that a part of it is broken in the dSC. It only means that in writing the full action (2) in terms of the Dirac fields (3) only the subgroup of spin rotations around the z axis is represented by a simple 4×4 matrices that act on Ψ. The rest is still present, but not that obvious in our choice of the Dirac fields, which was made to make the chiral symmetry manifest. (For a complementary representation that is fully rotationally symmetric at the expense of chiral symmetry see Appendix D.) The U(1) subsymmetry is therefore *always* present, both in the superconducting and insulating states. The $SU_c(2)$ factor is more interesting. The conserved currents (per pair of nodes) in the dSC that correspond to this symmetry are $J^{\Gamma}_{i,\mu}=\overline{\Psi}_i\gamma_\mu\Gamma\Psi_i$, where $\Gamma=\gamma_3,\gamma_5,\gamma_{35}$. As we have seen already, the γ_5 generator simply translates in the diagonal direction. The corresponding conserved charge may be written as

$$Q_i^5=\int d\vec{r}\, d\tau J^5_{i,0}=T\sum_{\sigma,\omega_n,\vec{k}\approx\pm\vec{K}_i}\pm c^\dagger_\sigma(\vec{k},\omega_n)c_\sigma(\vec{k},\omega_n), \tag{42}$$

and may be identified with the quasiparticle momentum along $\vec{K}_i$. More precisely, under the translation of the original electron operators $c_\sigma(\vec{k},\omega)\rightarrow e^{i\vec{k}\cdot\vec{R}}c_\sigma(\vec{k},\omega)$, the spinon field transforms as

$$\Psi_i(\vec{r},\tau)\rightarrow e^{i(\vec{K}_i\cdot\vec{R})\gamma_5}\Psi_i(\vec{r}+\vec{R},\tau), \tag{43}$$

where $\vec{k}=\vec{K}_i+\vec{q}$. The low-energy theory therefore has more symmetry than the original action (2), as the chiral rotation by γ_5 and the translations of Ψ separately are still the symmetries of the theory (24), while only when combined as above do they represent an ordinary translation. Nevertheless, breaking of chiral generator γ_5 implies breaking of the translational symmetry in the theory. The remaining two generators of the chiral $SU_c(2)$, γ_3 and γ_{35}, on the other hand, are not related to any spatial symmetry. They should be understood as "internal," and approximate, symmetries of the dSC that emerge at low energies. They rotate the translationally invariant $d+ip$ state into a SDW, and therefore connect the two fundamentally different types of insulators.

The reader should also note that in the action (24) the order parameter can be rotated independently for the first and the second Dirac field. Any linear combination of $\tilde{\gamma}_0$, $\tilde{\gamma}_3$, and $\tilde{\gamma}_5$ is a regular order parameter too. Since $\tilde{\gamma}_5$ is just a sin-SDW, the fundamentally different states are just the SDW $d+ip$ state. This, however, already leads to a variety of insulating phases. For example, one can choose the cos-SDW for the first Dirac field $(\vec{Q}_1)$, while being in the $d+ip$ state for the second. This would correspond to a one-dimensional SDW along one of the diagonals.

With the velocity anisotropy neglected QED_3 has a larger U(4) symmetry,with 16 generators G_i. The mass term now breaks all eight anticommuting generators, and the chiral manifold of insulating states becomes larger. For instance, rotating the cos-SDW with $\theta=\pi/4$ and the generator $G=G_{13}-G_{15}$ leads to a uniform state with an additional s component of pairing between spinons, $d+is$.[46] Interestingly, rotating the cos-SDW by block-off-diagonal generators may also lead to *charge stripes*. For example, taking $\theta=\pi/4$ and G_{15} rotates the 8×8 cos-SDW order parameter $I\otimes\tilde{\gamma}_0$ into $(1/2)\sigma_1\otimes(\gamma_1+\gamma_2)$. When written in terms of the electronic operators, this order parameter corresponds to the *one-dimensional* charge density wave with the periodicity $\vec{P}_b=\vec{K}_2+\vec{K}_1$, and with residual pairing correlations in the orthogonal a-axis direction. It has been known that stripes indeed occur in some high-T_c materials.[47] Here they emerge as insulating cousins of the d-wave state in the isotropic limit of the theory. It is also interesting that stripes seem always to be accompanied by the residual pairing correlations, so one can think of them as weakly coupled one-dimensional systems on the verge of becoming phase coherent.

VII. REDUCTION OF CHIRAL SYMMETRY BY THE IRRELEVANT TERMS

We saw that the low-energy theory of dSC has chiral U(2) symmetry per Dirac component, which when spontaneously broken leads to emergence of the SDW or the $d+ip$ insulators. This enlarged symmetry arises only at low energies, and the irrelevant terms omitted in the Eq. (6) reduce the U(2) to $U(1)\times U_c(1)$. In this section I show the higher-order derivatives and the Hubbard repulsion reduce the chiral $SU_c(2)$ symmetry to just translations, generated by γ_5. However, we will also find that if both perturbations are weak it will actually be the SDW solution that is energetically preferred.

Let us first consider the higher derivative terms in the Eq. (6). Since $\xi(\vec{k}-\vec{Q}_1)=\xi(\vec{q}-\vec{K}_1)$, and $\xi(\vec{q}-\vec{K}_1)=\xi(\vec{K}_1-\vec{q})$, and analogously for $\Delta(\vec{k})$, one can write the second-order derivatives in Eq. (6) as

$$S_1 = -i\int d^2\vec{r}\, d\tau \bar{\Psi}_1' \gamma_5 [\gamma_1 \xi''(\partial^2) - \gamma_2 \Delta''(\partial^2)] \Psi_1' + (1\to 2, x\leftrightarrow y), \tag{44}$$

where ξ'' and Δ'' are the functions coming from the expansion of $\xi(\vec{k})$ and $\Delta(\vec{k})$ around $\vec{K}_{1(2)}$, respectively. Their specific forms are model dependent, and will not be of importance here. What is important is that S_1 manifestly breaks the part of the chiral symmetry generated by γ_3 and γ_{35}, while preserving only the translational invariance, generated by γ_5. One can easily prove that the same holds to all orders in the gradient expansion.

Next, consider the Hubbard-like short-range repulsion term, in the continuum notation,

$$H_U = U\int d^2\vec{x} [n_+(\vec{x}) + n_-(\vec{x})]^2, \tag{45}$$

with $U>0$. Retaining again only the excitations near the four nodes, one can write this as

$$S_U = U\int d^2\vec{r}\, d\tau \left(i\sum_{i=1,2} \bar{\Psi}_i' \gamma_5 \gamma_1 \Psi_i' \right)^2. \tag{46}$$

The reader is probably not surprised that it is again only γ_5 that remains the symmetry generator. This is because γ_5 in our formalism is related to translations, which are always the exact symmetry of the action (2). So both the higher derivative terms and the quartic repulsion term reduce the chiral $SU_c(2)$ subgroup of U(2) to $U_c(1)$, the translations. One could therefore naively expect that it is the $d+ip$ state, which is translationally symmetric, that may be preferred by these perturbations. To decide on this, however, it is not enough just to know the symmetry of the action, since the new terms S_1 and S_U may turn out to disfavor the $d+ip$ state. Assuming that both ξ'' and Δ'' terms are small, one finds that the contribution to the energy of the SDW (or $d+ip$) state is of second order in the S_1. The interaction term, on the other hand, yields

$$\langle S_U \rangle_0 = -U\sum_i \langle \bar{\Psi}_i' \Psi_i' \rangle_0^2, \tag{47}$$

with the average taken over the massive QED$_3$ with $\gamma_0 = \sigma_1 \otimes I$ (cos-SDW). The result, of course, is the same for the sin-SDW, or for any linear combination of the cos-SDW and the sin-SDW. Alternatively, if one assumes the $d+ip$ ordering, one finds that S_U gives then a *positive* contribution to its energy, to first order in U. Although S_U is only translationally symmetric, it actually *inhibits* the formation of the translationally invariant state, and prefers the ordering to be in the "orthogonal" direction, i.e., the SDW.

If both the interaction and the gradient terms are weak, it will therefore always be the SDW solution that is energetically preferred. This is because both the repulsive and the higher-derivative terms are equally irrelevant by power counting (and have the engineering dimension -1); the gain in energy due to SDW is of first order only in U. The gradient terms affect the energy of the SDW only to the next order. So at long enough length scales one can alway neglect higher-derivative terms as compared to the repulsive interaction, which then serves to select the SDW over the $d+ip$ insulator.

VIII. MEAN-FIELD THEORY WITH REPULSIVE INTERACTION

The message from the preceding section is that the quartic term that represents a short-range repulsion, although irrelevant, at low but finite energies is still finite, and it breaks the chiral symmetry in favor of the SDW state. This is its first important role. The second is that once the chiral symmetry is dynamically broken by unbinding of vortex loops, the quartic term affects the size of the order parameter, and therefore sets the scale for the value of the SDW transition temperature. In this section I formulate the simplest mean-field theory of the chiral symmetry breaking in presence of the repulsion term, and demonstrate that it drastically increases the value of the SDW order parameter at $T=0$.

We have seen that unbinding of vortex loops leads to weak SDW order, but with the order parameter orders of magnitude smaller than the coupling constant $|\langle \Phi \rangle|^2$. Assuming that the dual condensate as a function of doping x should be of the same order of magnitude as the superfluid density on the other side of the transition (at $x=x_u$), $|\langle \Phi(x_u-\delta)\rangle|^2 \sim \rho_{\mathrm{sf}}(x_u+\delta)$, and that Uemura scaling[48] $T_c(x) \propto \rho_{sf}(x)$ is obeyed, the identification of the size of the SDW order parameter with the transition temperature $T_{\mathrm{SDW}}(x)$ suggests that $T_{\mathrm{SDW}}(x_u-\delta) \ll T_c(x_u+\delta)$. The difference in the relevant scales for the superconducting and the SDW orderings is in accord with the known phase diagram in the underdoped regime. Starting from half-filling, with increased doping the antiferromagnetic order is quickly lost, and only at a larger doping does the dSC appear. I attribute the absence of the obvious SDW order very near the superconducting phase to the inherent weakness of the spontaneous chiral symmetry breaking in dSC. Assuming that the weak SDW smoothly evolves into the commensurate antiferromagnet near half-filling, the obvious problem then becomes the following: how should one understand the dramatic increase of $T_{\mathrm{SDW}}(x)$ near half-filling, all the way up to ~ 300 K?

The answer is provided by the observation that although the repulsion U is irrelevant if weak enough, it enhances the SDW order once it becomes spontaneously generated through the interaction with the gauge field. To show this I will consider the mean-field theory of QED$_3$ with the additional S_U quartic term. First, notice that in the Hartree-Fock approximation the S_U term gets replaced by the effective quadratic term

$$S_U \to -U\langle \Psi_i' \bar{\Psi}_j' \rangle_0 \int d^2\vec{r}\, d\tau\, \mathrm{Tr}(\bar{\Psi}_i' \gamma_5 \gamma_1 \gamma_5 \gamma_1 \Psi_j'), \tag{48}$$

with the average to be calculated self-consistently within the theory quadratic in fermionic fields. The above term corresponds to the decoupling in the exchange (Fock) channel,

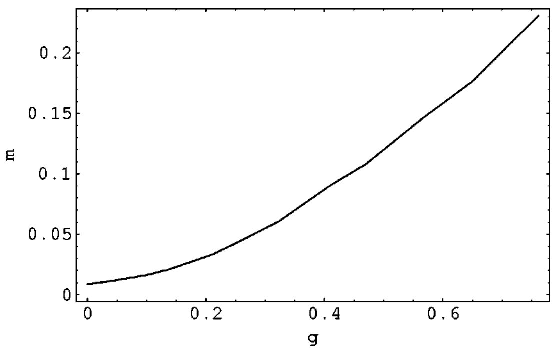

FIG. 4. The SDW order parameter m in units of $|\langle\Phi\rangle|^2$ as a function of dimensionless short-range repulsion $g = U|\langle\Phi\rangle|^2/(2\pi)^2$.

since the direct (Hartree) term vanishes. Therefore in the Hartree-Fock approximation, after the electron-spinon transformation,

$$S_U \to U\langle\Psi_i\bar{\Psi}_i\rangle_0 \int d^2\vec{r}\, d\tau\, \bar{\Psi}_i\Psi_i. \tag{49}$$

Assuming a uniform $\chi = -U\langle\Psi_i\bar{\Psi}_i\rangle_0$, and treating the gauge-field fluctuations in the large-N approximation leads to two coupled equations for χ and for the momentum-dependent fermion self-energy

$$\chi = U\int \frac{d^3\vec{k}}{(2\pi)^3}\frac{\Sigma(k)}{k^2+\Sigma^2(k)}, \tag{50}$$

$$\Sigma(q) = \chi + \int \frac{d^3\vec{k}}{(2\pi)^3}\frac{2|\langle\Phi\rangle|^2\Sigma(k)}{[k^2+\Sigma^2(k)][p^2+\Pi(p)]}, \tag{51}$$

with $\vec{p} = \vec{k} - \vec{q}$. When $U=0$ these reduce to the Eq. (32), which leads to $N_c = 32/\pi^2$. When loops are bound and $\langle\Phi\rangle = 0$, on the other hand, $\Sigma(q) = \chi$, and the Eq. (50) allows a nontrivial solution only when the dimensionless coupling $g = U\Lambda/(2\pi)^2 > 1$, where Λ is the uv cutoff, $\Lambda < T^*$. Assuming that long-range SDW order and dSC do not coexist, I take that $g<1$ in the superconducting phase, so that the quartic coupling is there irrelevant. With $\langle\Phi\rangle \neq 0$, however, small g ceases to be irrelevant, since there is now a small mass scale to effectively cut off its flow. Since $\Sigma(q)$ is quickly damped for $q \gg |\langle\Phi\rangle|^2$, one can take the uv cutoff in the above equations to be $\Lambda \sim |\langle\Phi\rangle|^2$. The above equations were studied before[49,50] in the context of gauged Nambu–Jona Lasinio model of chiral symmetry breaking in particle physics. Here I solve the equations numerically for $N=2$, as discussed in Appendix E. The result is presented in Fig. 4. The main point is that as the superconducting phase is more and more disordered and the dual condensate grows, the presence of a moderate repulsion between electrons increases the SDW order parameter at $T=0$ by one to two orders of magnitude. Recalling the above argument that compares T_{SDW} to the superconducting T_c on the other side of the superconductor-insulator transition, this appears to be in qualitative agreement with the generic behavior observed in underdoped cuprates.

IX. CONFINEMENT OF SPINONS

In the superconducting state, the electrically neutral low-energy spinons represented by the fermionic field Ψ in QED_3 are well-defined excitations. This effective spin-charge separation implicit in the superconducting state was emphasized in Ref. 51, and more recently in Ref. 7. One may therefore naturally wonder if this form of spin-charge separation will survive once the superconductivity is lost via unbinding of vortex loops. The answer seems to be *no*. It is believed that chiral symmetry breaking and confinement go together in QED_3.[20] The qualitative argument why it should be so is provided by the low-momentum form of the polarization tensor in the Eq. (33):[52,53] $\Pi(q) \sim q^2/m$ for $q \ll m$, so in two dimensions spinons are at large distances bound by a logarithmic potential. One may independently arrive at the same conclusion by analytically continuing the fermion propagator in the broken symmetry phase to real frequencies[54] to find that its poles lie at complex energies with both real and imaginary parts proportional to the chiral mass. The chiral symmetry breaking and confinement of spinons seem therefore to go hand in hand in QED_3, so the states with broken chiral symmetry, including most importantly the SDW, should not have well-defined fermionic excitations even above the mass "gap."

Disappearance of spinons from the spectrum in the insulating phase, if required, imposes a rather nontrivial constraint on a candidate theory for underdoped cuprates. For example, one could imagine a completely different mechanism of chiral symmetry breaking in dSC: even without the gauge field, simply increasing the quartic coupling U above a certain value $[U_c\Lambda/(2\pi)^2 = 1$ in the Hartree-Fock approximation$]$ would open the gap for spinons and lead to SDW order. This would be analogous to the chiral symmetry breaking in the Nambu–Jona Lasinio and related models.[55–58] The crucial difference, however, is that such a mechanism would yield well-defined spinon excitations at energies above the gap, in the insulating state. The integrity of the gapped spinons is assured essentially by the Landau phase space arguments. Such a "deconfined" antiferromagnet was dubbed AF* and studied in Ref. 24, for example. From this point of view it becomes a nontrivial problem to understand how spinons could be removed from the spectrum. In a QED_3 this is accomplished via the same nonperturbative mechanism that yields chiral symmetry breaking, described by Eq. (32), for example.

Having said all this, it needs to be realized that in a weak SDW confinement of spinons is effective only over very large distances, $L \gg 1/m$. At intermediate scales, the polarization $\Pi(q) \sim q$, so the potential between spinons is $\sim 1/r$, and at intermediate distances $1/m \gg L \gg 1/|\langle\Phi\rangle|^2$ spinons will appear effectively deconfined. In this sense it is still meaningful to think about underdoped cuprates as exhibiting an effective spin-charge separation. Computing the electron spectral function by taking the gauge-field fluctuations into account in the large-N approximation,[18] which suppresses the dynamical symmetry breaking, for example, gives results in qualitative agreement with the experiment.[59] As one continues to underdope, however, the SDW order parameter grows, and spinons become more strongly confined. In the strong antiferromagnet at half-filling therefore, one may expect spinons to be confined already at atomic distances.

X. EXPERIMENT

The principal consequence of QED$_3$ theory of underdoped cuprates is, of course, the antiferromagnetism itself. All the materials that become d-wave superconductors with doping are insulating antiferromagnets in its parent state. Furthermore, the sharp spectral features in the dSC should become very broad in the insulator, since there is a soft (propagator $\sim 1/q^2$) gauge field in the problem. Nevertheless, an insulator that derives from a dSC should partially inherit the d-wave form for its "gap," except for its finite value in the nodal directions. This is in very good agreement with the ARPES measurement on the insulating $Ca_2CuO_2Cl_2$, and $Sr_2CuO_2Cl_2$, in its parent state.[8,9] In Fig. 5 I compare the ARPES data for the gap measured from the top of the lower Hubbard band in the insulating state with the simplest functional form consistent with the chiral mass: at the remnant Fermi surface $\omega=(\{E_{\max}[\cos(k_x)-\cos(k_y)]/2\}^2+E_{\min}^2)^{1/2}$, where the chiral mass $m=E_{\min}=75$ meV is chosen to be the $T=0$ sublattice magnetization for $J=125$ meV. The best fit is obtained then for $E_{\max}=420$ meV. The quality of the fit is actually not very sensitive to some variations in $E_{\min}$ and the corresponding $E_{\max}$.

The key prediction of this work is that the above "gapped d-wave" form of the insulating gap is a generic feature of the insulating state. Upon underdoping, ARPES should show the standard d-wave gap for sharp quasiparticles in the superconducting phase, which should evolve into a gapped d-wave form for broad ARPES shape in the insulating state, with the gap increasing as one approaches half-filling. The rounding of the data at low energies should therefore be intrinsic to the insulating state, and should weaken with doping. Although the initial experiment on $Ca_2CuO_2Cl_2$ (Ref. 8) only indicated such rounding, later measurements on $Sr_2CuO_2Cl_2$ with higher resolution[9] clearly showed the deviation from the simple d-wave cusp at lowest energy. More recent measurements[60] indicate that the rounding of the data at low energies is a robust feature. It would clearly be desirable to perform a systematic study of this effect at variable doping.

It may also be worth mentioning that some signs of the gap rounding in the insulator may be observable already in the superconducting state. In Bi2212,[61] for example, as one underdopes, the d-wave gap continues to show the cusp at zero energy, but with the slope (velocity v_Δ) *decreasing*, in spite of the increase in the overall gap magnitude in the (π,π) direction. It is tempting to interpret this effect as a precursor of the dynamical mass generation. A detailed study of this effect and of the spectral features in the insulator is deferred to a future work.

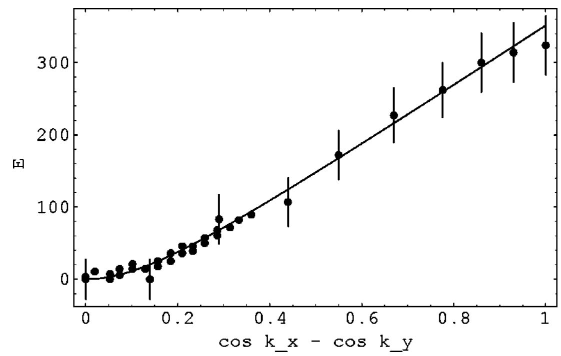

FIG. 5. ARPES results for $Ca_2CuO_2Cl_2$ (bars) and $Sr_2CuO_2Cl_2$ (dots) with $E=E(k)-E(\pi/2,\pi/2)$ in meV. The line is the function described in the text.

XI. CONCLUSION AND DISCUSSION

In summary, I have shown that the minimal theory that describes the unbinding of vortex defects in the d-wave superconductor at $T=0$ is the two-component, $(2+1)$-dimensional QED, with the vortex condensate playing the role of "charge." With the loss of phase coherence, the d-wave superconductor suffers the spontaneous breaking of the low-energy "chiral" symmetry, which results in a weak SDW order. It was argued that with underdoping this SDW smoothly evolves into the strong antiferromagnet near half-filling, with the selection and the increase of the SDW order parameter being provided by the repulsion between electrons. I argued that spinons are marginally confined in a weak SDW, and may appear effectively deconfined over intermediate length scales in the pseudogap regime. Finally, it was proposed that the rounded d-wave form of the "gap" in the insulating $Ca_2CuO_2Cl_2$ observed by Ronning *et al.*[8,9] may be a consequence of the chiral mass for the approximate spinon excitations, as implied by QED$_3$.

The present theory is similar in spirit to the approaches of Refs. 7 and 24, in that it attempts to understand the phase diagram of underdoped high-temperature superconductors beginning from the superconducting phase. It differs, however, in its conclusions of the ground state that results from unbinding of topological defects in the d-wave state. Whereas it was argued in Ref. 7 and 24 that the relevant description of this process is provided by the Ising (Z_2) gauge theory, and that the resulting state may show spin-charge separation, I argued that unbinding of defects of unit vorticity leads to the dynamical symmetry breaking in the low-energy theory, and the accompanying confinement of spinons in the insulating state. In fact, if one *demands* that the insulating state near half-filling is the standard antiferromagnet with spin-1 excitations and confined spinons, the form of a single theory that would be able to describe both the dSC and the insulator becomes severely restricted. QED$_3$ in this paper is one such theory.

A variation of QED$_3$ as an effective theory for underdoped cuprates has also been considered before,[62–65] as the theory of low-energy fluctuations around the π-flux phase in the large-N version of the Heisenberg model. In that approach the gauge invariance reflects the local particle number conservation at half-filling, and the gauge field has no dynamics on its own. As a result, the gauge field is necessarily compact, and the theory is infinitely strongly coupled. Not much is definitely known about such a lattice gauge theory, which greatly diminishes its utility. Nevertheless, it was argued that neglecting the instanton configurations would restore the antiferromagnetic order at half-filling, via spontaneous breaking of a different chiral symmetry, which in this case is actually an enlarged spin rotational symmetry.[63,64] While this logic may at first appear close to the one in the present work, there are crucial differences. First, I begin from the superconducting state, away from half-filling, with the gauge field describing vortex fluctuations. As a result, the gauge field is weakly coupled near the dSC-SDW transition. Also, the SDW phase that obtained from chiral symmetry breaking may be incommensurate, and the approximate chiral symmetry of the low-energy theory is unrelated to spin rotations.

Nevertheless, it may be possible to understand QED_3 as a low-energy description of the microscopic *t-J* model of cuprates. Starting from the mean-field slave-boson theory of the *t-J* model and integrating the constraints of no double occupancy, for example, leads to an effective theory of the form quite similar to QED_3,[66] but with the Volovik field $\vec{v}$ only. Including vortices would then be expected to introduce the gauge field $\vec{a}$, as shown in this paper. The point is that irrespectively from the underlying microscopic model the theory of the fluctuating dSC should assume the QED_3 form. Values of the parameters, however, may strongly depend on the microscopic physics: the bare stiffness K in the XY model for the phase fluctuations [Eq. (12)], for example, should be proportional to doping x in the doped Mott insulator.[66] Also, the charge of quasiparticles [the coefficient in the last term in the Eq. (10)] would be expected to change from unity to $\sim x$, at small dopings.

There exist further parallels between QED_3 and the gauge theory of the *t-J* model. One may formulate a representation of the *t-J* model with a U(1) gauge field that minimally couples to spinons and holons. It was argued[64] that the effect of holons would be to screen the temporal component of the gauge field, which then may be shown to halve the critical number of spinon species for the chiral instability, $N_c \to N_c/2$. In that way one could avoid the chiral transition at $N=2$ (assuming that $N_c \approx 3$), and have a spin liquid as the ground state in the underdoped regime instead. The tacit assumption, however, is that uncondensed bosons (holons) at $T=0$ may exist in a compressible state. If the system becomes insulating with the loss of phase coherence, however, bosons would become incompressible and the above argument breaks down. This is indeed the case in QED_3: with the proliferation of vortices the system becomes insulating, and all the components of the gauge field become massles. The same conclusion would be reached within the gauge theory of the *t-J* model if one would consider the incompressible state of slave bosons.[67]

The present work shares the same philosophy with the recent works,[18,45] where the massless U(1) gauge field as an effective description of unbound vortex loops was also considered. While the authors[18] considered the large-N limit of QED_3, and thus precluded chiral symmetry breaking, my main point is that at $T=0$ the spontaneous formation of the chiral condensate is nothing else but the SDW instability of the *d*-wave superconductor. The results of Ref. 18 may therefore be understood as applying to the finite-T phase much below the pseudogap scale T^* in Fig. 1.

The problem of phase disordering of dSC has also been recently studied by Ye.[68] Working in the Anderson gauge in which $\phi_{sA}=\phi_s$, $\phi_{sB}=0$ in the Eq. (8), the author concluded that the gauge field $\vec{a}$ is always massive when charge fluctuations are included. It is easy to see that this is a direct consequence of the gauge choice: in the Anderson gauge $\vec{a}=\vec{v}$, and not only $\vec{v}$, but $\vec{a}$ too is ultimately coupled to the charge current. In my gauge-invariant approach, on the other hand, $\vec{a}$ is completely decoupled from charge, and couples only to spin. Inclusion of charge fluctuations therefore does not make $\vec{a}$ massive, but simply adds an irrelevant quartic coupling to the QED_3 Lagrangian.

The intimate relationship between *d*-wave superconductivity and antiferromagnetism is also the main theme of the SO(5) theory of Zhang.[69] The present work echoes some of that general idea, but is based on entirely different physical principles. In particular, although there should be a direct dSC-SDW transition in the phase diagram, this appears unrelated to the SO(5) symmetry, but comes as a consequence of the *chiral* symmetry that emerges at low energies in the *d*-wave superconducting state. It is the spontaneous breaking of this hidden approximate symmetry that implies then the breaking of the spin rotational symmetry in the SDW phase.

The marginal confinement of spinons we found in the weak SDW phase is very much in line with the speculations of Laughlin[70,71] on parallels between antiferromagnetism and confinement in strong interactions. In fact, QED_3 shows precisely how chiral symmetry breaking, i.e., SDW ordering, binds spinons into spin-1 objects. Deconfinement in this theory seems indeed tantamount to the absence of chiral symmetry breaking. In this context, it may be interesting to note that the $d+id$ state, which would correspond to the $i\tilde{\gamma}_1\tilde{\gamma}_2$ matrix in Eq. (38), could lead to deconfined spinons. This state is outside of the chiral manifold, and it is believed that it is not spontaneously induced in the QED_3,[20] because of the Chern-Simons term that becomes generated for the gauge field. With the Chern-Simons term, on the other hand, the gauge-field propagator behaves like $\sim q$ at low momenta, and thus spinons may become deconfined.[72] Chiral symmetry breaking in QED_3 is therefore nothing by the effective description of the spinon confinement.

It is also interesting to note that were the critical number of fermions $N_c<2$, the result of phase disordering of dSC would be quite different. Instead of symmetry breaking and confinement one would find a gapless, chirally symmetric state, in which spinons would be deconfined. This is again because the polarization tensor would then be $\sim q$ at low momenta, i.e., the interaction between spinons would be $\sim 1/r$ at large distances. This state would be similar in spirit to the "nodal liquid,"[7] or analogous to the "algebraic Fermi liquids"[18,59,73] proposed in literature as candidates for the pseudogap phase. It has been proposed recently that $N_c=3/2$ exactly,[74] although all the actual calculations based on Schwinger-Dyson formalism lead to $N_c>3$. If N_c is indeed that low, phase disordering of the dSC would first lead to the deconfined pseudogap phase, which only later would turn into the confined SDW phase, presumably due to the repulsive quartic term, which is know to increase N_c.[49,50] At this time it is hard to say which one of these two scenarios is realized in cuprates.

The main point made in this paper is that unbinding of vortex loops in a d-wave superconductor at $T=0$ results in SDW order. It then appears natural to assume that the cores of fluctuating vortices are already in the insulating state. This speculation is in accord with the recent scanning tunneling microscopy and neutron scattering experiments,[75–77] the SO(5) proposal,[69,78] and the mean-field[79] and the finite size QED_3 calculations.[80] The superconductor-insulator transition would then be the result of the decrease of the bare stiffness K in the XY model with underdoping, since $K\sim x$ in the doped Mott insulator.[66]

XII. FURTHER PROBLEMS

I finish with a tentative list of problems opened by this work.

(1) The role of strong anisotropy $v_f/v_\Delta \gg 1$ that exists in cuprates is unclear. In particular, since anisotropy on the bare level is marginal, it may affect the value of N_c. The preliminary results, indicate, however, that weak anisotropy is irrelevant, so that one would expect N_c to be unaffected by it.[43]

(2) The nature of the various phase transitions in the theory is also of interest. Whereas one expects that gapless quasiparticles do not change the Kosterlitz-Thouless universality class of the finite temperature superconducting transition, the nature of chiral symmetry breaking at finite temperature and its possible interplay with the Néel transition is far less clear.[81] In particular, in relation to Uemura scaling,[48] one would like to understand the quantum superconductor-insulator criticality and how it may be affected by gapless spinons.

(3) Can long-range SDW and SC order coexist? In the approximation employed in the present work, the gauge field $\vec{a}$ is considered decoupled from spinons in the dSC phase. This is likely to underestimate the effect of $\vec{a}$, and a better approximation for the gauge-field propagator is needed to study its effect *inside* the dSC. This could be important in light of recent experimental data[77,82] that may be interpreted as indicating the coexistence of the SDW and SC orders in some compounds.[83]

(4) The present work also points to a new route towards a deconfined phase in two dimensions: lowering N_c below two would allow for an insulating phase with deconfined spinons. At present, however, it is not clear how to achieve this within QED_3, unless the Schwinger-Dyson equations systematically overestimate N_c.[74]

(5) The computation of the electron propagator within QED_3 is an important problem.[41] This would be necessary for a detailed comparison of the theory with the ARPES measurements.

(6) As mentioned at the end of Sec. IV, on a lattice, the gauge field $\vec{a}$ appears to be compact, in contrast to the Volovik field $\vec{v}$. The effect of the compact nature of $\vec{a}$ on chiral symmetry breaking in QED_3 is at present poorly understood. It has been argued that coupling to gapless spinons makes the single instanton-antiinstanton pair that derives from compactness of $\vec{a}$ bound above the certain number of spinon components N_{inst},[63,84] N_{inst} may be made smaller than N_c for chiral symmetry breaking by a large anisotropy,[73] for example. It is unclear, however, whether this conclusion survives the effects of screening by other pairs.[85] Also, even if the instantons can be made irrelevant above N_c, below N_c one would expect them to become relevant again with the opening of the spinon "gap." This in turn could have profound consequences for the spinon confinement. It would obviously be desirable to shed some light on these pressing issues.

ACKNOWLEDGMENTS

The author has benefited greatly from discussions and exchanges with S. Dodge, V. Gusynin, M. P. A. Fisher, E. Fradkin, M. Franz, K. Kavehmaryan D. J. Lee, B. Marston, M. Reenders, F. Ronning, S. Sachdev, B. Seradjeh, S. Sondhi, Z. Tešanović, C. Wu, W.-C. Wu, and S.-C Zhang. This work was supported by NSERC of Canada and the Research Corporation.

APPENDIX A

I present the self-consistent mean-field theory (20) of the lattice superconductor[34] and use it to approximately compute the correlator appearing in Eq. (20). By the Bogoliubov inequality,

$$Z_{xy} \geq Z_0 e^{-\langle H-H_0\rangle_0}, \tag{A1}$$

where Z_{xy} is the partition function in the dual form (21) with a finite "inverse temperature" x, and the average in the exponent is performed over a *local* mean-field Hamiltonian

$$H_0=-h\sum \cos\theta_i+\frac{1}{8K\pi^2}\sum (\vec{\nabla}\times\vec{\Phi})^2 + \frac{m^2}{4K\pi^2}\sum \vec{\Phi}^2. \tag{A2}$$

The optimal values of the parameters h and m that maximize the right-hand side in the Bogoliubov inequality are then determined by the equations

$$h=\frac{6A}{x}\frac{I_1(h)}{I_0(h)}, \tag{A3}$$

$$m^2=\frac{K\pi^2}{3}\frac{I_1(h)}{I_0(h)}h, \tag{A4}$$

$$A=\exp\left[-\frac{2K\pi^2}{3}\int\frac{d^3\vec{k}}{(2\pi)^3}\frac{1}{F(k)+m^2}\right], \tag{A5}$$

where $F(k)=\Sigma_\nu(e^{ik_\nu}-1)^2$, and the integral over $\vec{k}$ is taken over $(-\pi,\pi)$. I_0 and I_1 are the Bessel functions. These equations can be solved graphically, and describe a discontinuous transition from the phase with $h=m=0$ (bound vortex loops), to the condensed phase $h\neq 0$, $m\neq 0$ (infinitely large vortex loops).[86]

The requisite average in the Eq. (20) is easy to compute in the mean-field theory that has different sites decoupled:

$$\langle\cos(\theta_i-\theta_{i+\widehat{nu}}-\Phi_{i,\nu})\rangle_0=|\langle e^{i\theta_i}\rangle_0|^2\langle e^{-i\Phi_{i,\nu}}\rangle_0. \tag{A6}$$

Since, $\langle e^{-i\Phi_{i,\nu}}\rangle_0=A$ and finite, we conclude that

$$\langle\cos(\theta_i-\theta_{i+\nu}-2\pi\Phi_{i,\nu})\rangle_0\propto h^2, \tag{A7}$$

i.e., finite *only* in the ordered phase of the dual theory (20), i.e., in the disordered phase of the original XY model.

APPENDIX B

Here I provide a different derivation of the dynamics of the gauge field $\vec{a}$ at $T\neq 0$ starting from the Hamiltonian for the Coulomb plasma. Assume a collection of N_+ (N_-) vortices (antivortices) at the positions $\{\vec{r}_i\}$. The Hamiltonian of the vortex system is

$$H_v=\frac{1}{2}\sum_{i=1}^{N}q_iq_jv(\vec{r}_i-\vec{r}_j), \tag{B1}$$

where $v(\vec{r})\approx-\ln|\vec{r}|$, at large distances, and $N=N^++N^-$, $q_i=\pm 1$. The partition function of the vortex system Z_v can then be written as

$$Z_v=\sum_{N_{A,B}^{+,-}=0}^{\infty}\frac{N^+!}{N_A^+!N_B^+!}\frac{N^-!}{N_A^-!N_B^-!}\frac{(y/2)^N}{N^+!N^-!}\int\prod_{i=1}^{N}d\vec{r}_ie^{-H_v/T}, \tag{B2}$$

where $N^{+(-)}=N_A^{+(-)}+N_B^{+(-)}$ and y is the bare vortex fugacity. The combinatorial factors serve to ensure that in Z_v one sums over *all* possible divisions of vortices and antivortices into groups A and B, and divides by the number of combinations. With this symmetrization the symmetry between up and down spin in the original Hamiltonian (2) will be preserved in the Dirac theory for neutral spinons. This also guarantees that on average there is an equal number of vortices (and antivortices) in both groups.

Next, introduce the vorticity densities in Z_v by inserting the unity

$$1=\int D[\rho_A]\delta\left(\rho_A(\vec{r})-\sum_{i=1}^{N_A}q_{iA}\delta(\vec{r}-\vec{r}_{iA})\right), \tag{B3}$$

and similarly for B. The gauge field then becomes

$$(\vec{\nabla}\times\vec{a}(\vec{r}))_\tau=\pi[\rho_A(\vec{r})-\rho_B(\vec{r})], \tag{B4}$$

in the transverse gauge $\vec{\nabla}\cdot\vec{a}=0$, and the index denotes the τ component. $\vec{v}$ is defined the same way except with the plus sign between ρ_A and ρ_B.

By introducing two auxiliary fields Φ_A and Φ_B to enforce the constraints, after the integration over the densities the partition function may be written as

$$\begin{aligned}Z_v=&\sum_{N_{A,B}^{+,-}=0}^{\infty}\frac{(y/2)^N}{N_A^+!N_A^-!N_B^+!N_B^-!}\int D[\vec{a},\vec{v},\Phi_+,\Phi_-]\\&\times\exp\Bigg\{-\Bigg[\frac{1}{2\pi^2T}\int d\vec{r}\,d\vec{r}\,'B(\vec{r})v(\vec{r}-\vec{r}\,')B(\vec{r}\,')\\&+\frac{i}{2\pi}\int d\vec{r}[B(\vec{r})\Phi_+(\vec{r})+b(r)\Phi_-(\vec{r})]\\&-\sum_{i\alpha,\alpha=A,B}\ln\int d\vec{r}\exp[iq_{i\alpha}\Phi_\alpha(\vec{r})]\Bigg]\Bigg\},\end{aligned} \tag{B5}$$

where $\Phi_{+,-}=\Phi_A\pm\Phi_B$, $B(\vec{r})=(\vec{\nabla}\times\vec{v})_\tau$, and $b(\vec{r})=(\vec{\nabla}\times\vec{a})_\tau$. Performing the summations yields

$$\begin{aligned}Z_v=&\int D[\vec{a},\vec{v},\Phi_+,\Phi_-]\exp\Bigg\{-\Bigg[\frac{1}{2\pi^2T}\int d\vec{r}\,d\vec{r}\,'B(\vec{r})v(\vec{r}\\&-\vec{r}\,')B(\vec{r}\,')+\frac{i}{2\pi}\int d\vec{r}[B(\vec{r})\Phi_+(\vec{r})+b(\vec{r})\Phi_-(\vec{r})]\\&-y\int d\vec{r}[\cos\Phi_A(\vec{r})+\cos\Phi_B(\vec{r})]\Bigg]\Bigg\}.\end{aligned} \tag{B6}$$

Finally, neglecting the coupling to the charge current, the Gaussian integration over $\vec{v}$ (i.e., B) gives

$$\begin{aligned}Z_v=&\int D[\vec{a},\Phi_+,\Phi_-]\exp\Bigg[-\int d\vec{r}\Bigg(T/2(\nabla\Phi_+)^2\\&+\frac{i}{\pi}b(\vec{r})\Phi_-(\vec{r})-2y\cos[\Phi_+(\vec{r})]\cos[\Phi_-(\vec{r})]\Bigg)\Bigg],\end{aligned} \tag{B7}$$

where I also have rescaled the Φ fields by a factor of 2. The last expression is then analogous to the $T=0$ expression in the Eq. (18) with x finite and without the dual angles $\theta_{A,B}$. By introducing a source term in the action, $\sim i\int j(r)b(r)/\pi$, and integrating over b, one readily finds

$$\langle[\vec{\nabla}\times\vec{a}(\vec{r})]_\tau[\vec{\nabla}\times\vec{a}(\vec{r}\,')]_\tau\rangle=\langle y\rangle\delta(\vec{r}-\vec{r}\,'), \tag{B8}$$

where $\langle y\rangle=y\pi^2\langle\exp(i\Phi_+)\rangle$, with the average to be taken at $\Phi_-=\vec{a}\equiv 0$. One recognizes $\langle y\rangle$ as the renormalized, or running, fugacity in the Kosterlitz-Thouless scaling, which signals the appearance of free vortices. $\langle y\rangle$ plays the role analogous to the vortex loop condensate in $2+1$ dimensions, in providing a mass for the field Φ_+ in Eq. (B7). This implies the Maxwell term at $T\neq 0$ for the τ component of $\vec{\nabla}\times\vec{a}$ once fluctuating vortices are integrated out.

APPENDIX C

For completeness, here I outline the derivation of the result that chiral symmetry in isotropic massless QED3 is spontaneously broken for $N<N_c$, with N_c finite, at any value of the coupling constant. Rescaling the momenta $p/m\to p$ and self-energies $\Sigma(p)/m\to m$ and $\Pi(p)/m^2 \to\Pi(p)$, after taking the limit $q\to 0$ in the Eq. (32) we find

$$1=\frac{|\langle\Phi\rangle|^2}{\pi^2 m}\int_0^{\Lambda/m}dp\,\frac{p^2\Sigma(p)}{[p^2+\Sigma^2(p)][p^2+\Pi(p)]}, \tag{C1}$$

where the polarization is now

$$\Pi(p)=\frac{N|\langle\Phi\rangle|^2}{4\pi m}f(p), \tag{C2}$$

with

$$f(p)=\left[2+\frac{p^2-4}{p}\sin^{-1}\left(\frac{p}{\sqrt{4+p^2}}\right)\right], \tag{C3}$$

to the leading order in N.[20] We see that the right-hand side of the Eq. (C1) is a decreasing function of m, so for $m\neq 0$ solution to exist we just need the right-hand side to be greater than 1 for $m=0$. This is satisfied for $N<N_c$, where

$$N_c=4\int_0^\infty dp\,\frac{p^2\Sigma(p)}{[p^2+\Sigma^2(p)]f(p)}. \tag{C4}$$

As defined, $\Sigma(0)=1$, and one expects $\Sigma(p)$ to vanish at large momenta. Also, $f(p)\approx\pi p/2$ for $p\gg 1$, so the integrand at large argument behaves like $\sim\Sigma(p)/p$. N_c is therefore finite, and independent of the coupling constant $\langle\Phi\rangle$. Its precise value in the large-N approximation will depend only on the function $\Sigma(p)$ at $N=N_c$, and can be obtained by solving the differential equation equivalent to the integral equation (C1) (Ref. 20) (see Appendix E). This yields $N_c=32/\pi^2$, not far from the results of other more elaborate computations that go beyond the leading order in N.[38,39]

APPENDIX D

Here I discuss a different representation of the quasiparticle action, more in line with the previous work.[7] This should serve to underline the difference between the approximate chiral $SU_c(2)$ symmetry, and the exact spin rotational $SO(3)$, also present in dSC. It is only the latter that will appear in the different version of the theory considered here and in Ref. 7, while the chiral symmetry will remain completely obscured.

I start again from the same quasiparticle action in the Eq. (2), but now introduce the four-component field as

$$\Psi'^{\dagger}_{1(2)}(\vec q,\omega_n)=(c^\dagger_+(\vec k,\omega_n),c_-(-\vec k,-\omega_n),c^\dagger_-(\vec k,\omega_n),\\ -c_+(-\vec k,-\omega_n)). \tag{D1}$$

By linearizing the spectrum and by retaining only the modes near the four nodes, the continuum theory may again be written as

$$S[\Psi']=\int d^2\vec r\int_0^\beta d\tau\,\Psi'^{\dagger}_1[\partial_\tau+M_1 v_f\partial_x+M_2 v_\Delta\partial_y]\Psi'_1\\ +(1\to 2,x\leftrightarrow y), \tag{D2}$$

but this time with a different form of the matrices M_1 and M_2: $M_1=-iI\otimes\sigma_3$ and $M_2=iI\otimes\sigma_1$. Introducing $\gamma_0=\sigma_3\otimes\sigma_2$, for example, the theory becomes

$$S[\Psi']=\int d^2\vec r\int_0^\beta d\tau\,\bar\Psi'_1[\gamma_0\partial_\tau+\gamma_1 v_f\partial_x+\gamma_2 v_\Delta\partial_y]\Psi'_1\\ +(1\to 2,x\leftrightarrow y), \tag{D3}$$

with $\gamma_1=\sigma_3\otimes\sigma_1$ and $\gamma_2=\sigma_3\otimes\sigma_3$. It is interesting to consider the generators of the global $U(2)=U(1)\times SU(2)$ symmetry per Dirac component present in this representation of the theory. They are $I_4=I\otimes I$, $\gamma_3=\sigma_1\otimes I$, $\gamma_5=-\sigma_2\otimes I$, and $\gamma_{35}=\sigma_3\otimes I$, respectively. One may recognize the $U(1)$ factor as representing now the continuous translations, since under a translation $c_\sigma(\vec k,\omega)\to e^{i\vec k\cdot\vec R}c_\sigma(\vec k,\omega)$, the Dirac field now transforms as

$$\Psi'_i(\vec r,\tau)\to e^{i\vec K_i\cdot\vec R}\Psi'_i(\vec r+\vec R,\tau). \tag{D4}$$

The $SU(2)$ operators, on the other hand, are nothing but the *spin rotations*. In fact, the above $U(2)$ is an exact symmetry of the Hamiltonian (2), and is present even if all higher order derivatives are retained.

Including the coupling to vortex loops via massless gauge field in the above representation of the problem then may spontaneously induce only the $d+ip$ insulator. This breaks two of the above generators, which then simply rotate the spin axis. Translational symetry is, on the other hand, always preserved in this formulation, and the SDW remains invisible.

APPENDIX E

Here I provide the details behind the numerical solution of Eqs. (50) and (51). Since we are interested only in the qualitative effect of the U term, it will suffice to assume that the fermion mass is small, $m\ll|\langle\Phi\rangle|^2$, so that one can neglect the p^2 term compared to $\Pi(p)$ in Eq. (51), and take

$$\Pi(p)=\frac{N|\langle\Phi\rangle|^2}{8}p, \tag{E1}$$

appropriate for $p\gg m$. This approximation is known to lead to even quantitatively good result for the mass for N as low as unity, when $U=0$.[20] Evaluating the angular integrals then gives

$$\Sigma(q)=\chi+\frac{8}{N\pi^2 q}\int_0^\Lambda dk\frac{k\Sigma(k)[k-(k-q)\theta(k-q)]}{k^2+\Sigma^2(k)}. \tag{E2}$$

Differentiating twice, one finds that this integral equation is equivalent to the differential equation:[20]

$$\frac{d}{dq}\left(q^2\frac{d}{dq}\Sigma(q)\right)=-\frac{8}{N\pi^2}\frac{q^2\Sigma(q)}{q^2+\Sigma^2(q)}, \tag{E3}$$

with the boundary condition

$$\Lambda\Sigma'(\Lambda)+\Sigma(\Lambda)=\chi, \tag{E4}$$

and with

$$\chi=\frac{U}{(2\pi)^2}\int_0^\Lambda dq\frac{q^2\Sigma(q)}{q^2+\Sigma^2(q)}. \tag{E5}$$

Here I take $\Lambda=|\langle\Phi\rangle|^2$.

The above equations may now be studies by assuming $q \gg \Sigma(q)$, which leads to a linear equation that can be exactly solved.[87] This yields, for example, the well-known transition line in the g-N plane: $g_c(N)=(1/4)(1+\sqrt{1-(N_c/N)^2}$, for $N>N_c$, $g_c\leqslant 1/4$ for $N=N_c$, with $N_c=32/\pi^2$. To determine the size of $\Sigma(0)$, however, one needs to solve the full nonlinear equation. This may be accomplished, for example, by choosing a value for χ, assuming $\Sigma(\Lambda)$ next, and then iterating back to find $\Sigma(q)$ for $0<q<\Lambda$. The solution is found by tuning $\Sigma(\Lambda)$ to achieve $\Sigma(0)$ finite. One then computes the value of $g=U\Lambda/(2\pi)^2$ from the assumed χ and the found $\Sigma(q)$. This procedure leads to Fig. 4.

[1] P. W. Anderson, *The Theory of Superconductivity in the High-T_c Cuprates* (Princeton University Press, 1997).
[2] J. R. Schrieffer, *Theory of Superconductivity* (W. A. Benjamin, Inc., New York, 1964).
[3] W.N. Hardy, D.A. Bonn, D.C. Morgan, Ruixing Liang, and Kuan Zhang, Phys. Rev. Lett. **70**, 3999 (1993).
[4] D.A. Wollman, D.J. Van Harlingen, J. Gianpintzakis, and D.M. Ginsberg, Phys. Rev. Lett. **74**, 797 (1995).
[5] A. Hosseini, R. Harris, Saeid Kamal, P. Dosanjh, J. Preston, Ruixing Liang, W.N. Hardy, and D.A. Bonn, Phys. Rev. B **60**, 1349 (1999).
[6] A. Kaminski, J. Mesot, H. Fretwell, J.C. Campuzano, M.R. Norman, M. Randeria, H. Ding, T. Sato, T. Takahashi, T. Mochiku, K. Kadowaki, and H. Hoechst, Phys. Rev. Lett. **84**, 1788 (2000).
[7] L. Balents, M.P.A. Fisher, and C. Nayak, Int. J. Mod. Phys. B **10**, 1033 (1998); Phys. Rev. B **60**, 1654 (1999); for a review, see M.P.A. Fisher, cond-mat/9806164 (unpublished).
[8] F. Ronning, C. Kim, D.L. Feng, D.S. Marshal, A.G. Loeser, L.L. Miller, J.N. Eckstein, I. Božović, and Z.-X. Shen, Science **282**, 2067 (1998).
[9] F. Ronning, C. Kim, A. Damascelli, N.P. Armitage, D.H. Lu, K.M. Shen, L.L. Miller, Y.J. Kim, M.A. Kastner, R.J. Birgenau, and Z.-X. Shen, Physica C **341-348**, 2087 (2000).
[10] I.F. Herbut, Phys. Rev. Lett. **85**, 1532 (2000).
[11] S. Doniach and M. Inui, Phys. Rev. B **41**, 6668 (1990).
[12] V.J. Emery and S.A. Kivelson, Nature (London) **374**, 434 (1995).
[13] J. Corson, R. Mallozzi, J. Orenstein, J.N. Eckstein, and I. Božović, Nature (London) **398**, 221 (1999).
[14] Z.A. Xu, N.P. Ong, Y. Wang, T. Kakeshita, and S. Uchida, Nature (London) **406**, 486 (2000); Y. Wang, Z.A. Xu, T. Kakeshita, S. Uchida, S. Ono, Y. Ando, and N.P. Ong, Phys. Rev. B **64**, 224519 (2001).
[15] I.F. Herbut, Phys. Rev. Lett. **88**, 047006 (2002).
[16] H. Kleinert, *Gauge Fields in Condensed Matter* (World Scientific, Singapore, 1989).
[17] C. Dasgupta and B.I. Halperin, Phys. Rev. Lett. **47**, 1556 (1981); A. Nguyen and A. Sudbo, Phys. Rev. B **60**, 15307 (1999); N.D. Antunes and L.M.A. Bettencourt, Phys. Rev. Lett. **81**, 3083 (1998); G.A. Williams, *ibid.* **82**, 1201 (1999).
[18] M. Franz and Z. Tešanović, Phys. Rev. Lett. **87**, 257003 (2001); M. Franz, Z. Tešanović, and O. Vafek, cond-mat/0203333 (unpublished).
[19] T. Valla, A.V. Fedorov, P.D. Johnson, Q. Li, G.D. Gu, and N. Koshizuka, Phys. Rev. Lett. **85**, 828 (2000).
[20] R.D. Pisarski, Phys. Rev. D **29**, 2423 (1984); T.W. Appelquist, M. Bowick, D. Karabali, and L.C.R. Wijewardhana, *ibid.* **33**, 3704 (1986); T.W. Appelquist, D. Nash, and L.C.R. Wijewardhana, Phys. Rev. Lett. **60**, 2575 (1988).
[21] B. Keimer, A. Aharony, A. Auerbach, R.J. Birgenau, A. Cassanho, Y. Endoh, R.W. Erwin, M.A. Kastner, and G. Shirane, Phys. Rev. B **45**, 7430 (1992).
[22] X.G. Wen, Phys. Rev. B **44**, 2664 (1991).
[23] C. Mudry and E. Fradkin, Phys. Rev. B **49**, 5200 (1994); **50**, 11409 (1994).
[24] T. Senthil and M.P.A. Fisher, Phys. Rev. B **62**, 7850 (2000); C. Lannert, M.P.A. Fisher, and T. Senthil, *ibid.* **63**, 134510 (2001).
[25] M.A. Kastner, R.J. Birgeneau, G. Shirane, and Y. Endoh, Rev. Mod. Phys. **70**, 897 (1998).
[26] J.M. Kosterlitz and D.J. Thouless, J. Phys. C **6**, 1181 (1973); J.M. Kosterlitz, *ibid.* **7**, 1046 (1974); P.B. Wiegmann *ibid.* **11**, 1583 (1978).
[27] M. Franz and Z. Tešanović, Phys. Rev. Lett. **84**, 554 (2000); O. Vafek, A. Melikyan, M. Franz, and Z. Tešanović, Phys. Rev. B **63**, 134509 (2001).
[28] See also, P.W. Anderson, cond-mat/9812063 (unpublished).
[29] G.E. Volovik, Pisma Zh. Éksp. Teor. Fiz. **58**, 457 (1993) [JETP Lett. **58**, 469 (1993)].
[30] J. Villain, J. Phys. (Paris) **36**, 581 (1975).
[31] Definition of $\vec{v}$ here differs slightly from the one below Eq. (10), in that it does not contain the spin-wave contribution that has already been integrated out in arriving at Eq. (13).
[32] M.P.A. Fisher and D.H. Lee, Phys. Rev. B **39**, 2756 (1989).
[33] I.F. Herbut, Phys. Rev. B **57**, 13729 (1998).
[34] M. Peskin, Ann. Phys. (N.Y.) **113**, 122 (1978); T. Banks, R. Meyerson, and J. Kogut, Nucl. Phys. B **129**, 493 (1977); M. Stone and P.R. Thomas, *ibid.* **144**, 513 (1978).
[35] I.F. Herbut, J. Phys. A **30**, 423 (1997).
[36] P. Olsson and S. Teitel, Phys. Rev. Lett. **80**, 1964 (1998).
[37] M. Randeria, in *Bose-Einstein Condensation*, edited by A. Griffin, D. Snoke, and S. Stringari (Cambridge University Press, 1996), p. 355.
[38] P. Maris, Phys. Rev. D **54**, 4049 (1995), and references therein.
[39] E. Dagotto, J.B. Kogut, and A. Kocić, Phys. Rev. Lett. **62**, 1083 (1989); Nucl. Phys. B **334**, 279 (1990).

[40] J. Alexandre, K. Farakos, S.J. Hands, G. Koutsoumbas, and S.E. Morrison, Phys. Rev. D **64**, 034502 (2001).
[41] For a discussion of the problem of electron propagator in QED$_3$, see D.V. Khveschenko, Phys. Rev. B **65**, 235111 (2002); and cond-mat/0204040 (unpublished); see also the second reference in Ref. 18 and M. Franz *et al.*, cond-mat/0204536 (unpublished).
[42] M. Chiao, R.W. Hill, C. Lupien, L. Taillefer, P. Lambert, R. Gagnon, and P. Fournier, Phys. Rev. B **62**, 3554 (2000).
[43] D.J. Lee and I.F. Herbut, cond-mat/0201088, Phys. Rev. B (to be published 1 September 2002); O. Vafek *et al.*, cond-mat/0203047 (unpublished).
[44] C. Vafa and E. Witten, Commun. Math. Phys. **95**, 257 (1984).
[45] Z. Tešanović, O. Vafek, and M. Franz, Phys. Rev. B **65**, 180511 (2002).
[46] I thank to B. Seradjeh for pointing this out to me.
[47] For a recent review, see J. Orenstein and A.J. Millis, Science **288**, 468 (2000).
[48] T. Schneider and J. M. Singer, *Phase Transition Approach to High Temperature Superconductivity* (Imperial College Press, 2000).
[49] M. Carena, T.E. Clark, and C.E.M. Wagner, Nucl. Phys. B **356**, 117 (1991).
[50] V. Gusynin, A. Hams, and M. Reenders, Phys. Rev. D **63**, 045025 (2001).
[51] D.S. Rokhsar and S. Kivelson, Phys. Rev. B **41**, 11693 (1990).
[52] M. Gopfert and G. Mack, Commun. Math. Phys. **82**, 545 (1982).
[53] C.J. Burden, J. Praschifka, and C.D. Roberts, Phys. Rev. D **46**, 2695 (1992).
[54] P. Maris, Phys. Rev. D **52**, 6087 (1995).
[55] Y. Nambu and G. Jona Lasinio, Phys. Rev. **122**, 345 (1961).
[56] B. Rosenstein, B.J. Warr, and S.H. Park, Phys. Rev. Lett. **62**, 1433 (1989).
[57] G.W. Semenoff and L.C.R. Wijewardhana, Phys. Rev. Lett. **63**, 2633 (1989).
[58] M. Reenders, Phys. Rev. B **66**, 024501 (2002).
[59] For a similar calculation, but within an apparently different physical picture see also W. Rantner and X.G. Wen, Phys. Rev. Lett. **86**, 3871 (2001); and cond-mat/0105540 (unpublished).
[60] Filip Ronning (private communication).
[61] J. Mesot, M.R. Norman, H. Ding, M. Randeria, J.C. Campuzano, A. Paramekanti, H.M. Fretwell, A. Kaminski, T. Takeuchi, T. Yokoya, T. Sato, T. Takehashi, T. Mochiku, and K. Kadowaki, Phys. Rev. Lett. **83**, 840 (1999).
[62] I. Affleck and J.B. Marston, Phys. Rev. B **37**, 3774 (1988); J.B. Marston and I. Affleck, *ibid.* **39**, 11538 (1989).
[63] J.B. Marston, Phys. Rev. Lett. **64**, 1166 (1990).
[64] D.H. Kim and P.A. Lee, Ann. Phys. (N.Y.) **272**, 130 (1999) and references therein.
[65] N. Nagaosa and P.A. Lee, Phys. Rev. B **61**, 9166 (2000).
[66] D.H. Lee, Phys. Rev. Lett. **84**, 2694 (2000).
[67] Y.B. Kim and Z. Wang, Europhys. Lett. **50**, 656 (2000).
[68] J. Ye, Phys. Rev. B **65**, 214505 (2002).
[69] S.-C. Zhang, Science **275**, 1089 (1997).
[70] R.B. Laughlin, Adv. Phys. **47**, 943 (1998); cond-mat/9802180 (unpublished).
[71] See also Ref. 23, and the book by E. Fradkin, *Field Theories of Condensed Matter Systems* (Addison-Wesley, Redwood City, 1991), Chap. 6.
[72] See also, E. Fradkin and F.A. Schaposnik, Phys. Rev. Lett. **66**, 276 (1991).
[73] X.G. Wen, Phys. Rev. B **65**, 165113 (2002).
[74] T.W. Appelquist, A.G. Cohen, and M. Schmaltz, Phys. Rev. D **60**, 045003 (1999).
[75] Ch. Renner, B. Revaz, K. Kadowaki, I. Maggio-Aprile, and O. Fischer, Phys. Rev. Lett. **80**, 3606 (1998).
[76] J.E. Hoffman, E.W. Hudson, K.M. Lang, V. Madhavan, H. Eisaki, S. Uchida, and J.C. Davis, Science **295**, 466 (2002).
[77] B. Lake, H.M. Ronnow, N.B. Christensen, G. Aeppli, K. Lefmann, D.F. McMorrow, P. Vorderwisch, P. Smeibidl, N. Mangkorntong, T. Sasegawa, M. Nohara, H. Tagaki, and T.E. Mason, Nature (London) **415**, 299 (2002).
[78] D.P. Arovas, A.J. Berlinsky, C. Kallin, and S.-C. Zhang, Phys. Rev. Lett. **79**, 2871 (1997).
[79] J-X. Zhu and C.S. Ting, Phys. Rev. Lett. **87**, 147002 (2001).
[80] M. Franz, D.E. Sheehy, and Z. Tešanović, Phys Rev. Lett. **88**, 257005 (2002).
[81] I.J.R. Aitchison and C.D. Fosco, Nucl. Phys. B **578**, 199 (2000) and references therein.
[82] R.I. Miller, R.F. Kiefl, J.H. Brewer, J.E. Sonier, J. Chakhalian, S. Dunsiger, G.D. Morris, A.N. Price, D.A. Bonn, W.H. Hardy, and R. Liang, Phys. Rev. Lett. **88**, 137002 (2002).
[83] E. Demler, S. Sachdev, and Y. Zhang, Phys. Rev. Lett. **87**, 067202 (2001).
[84] H. Kleinert, F.S. Nogueira, and A. Sudbo, Phys. Rev. Lett. **88**, 232001 (2002); L.B. Ioffe and A.I. Larkin, Phys. Rev. B **39**, 8988 (1989).
[85] S. Sachdev and K. Park, Ann. Phys. (N.Y.) **298**, 58 (2002).
[86] R. Balian, J.M. Drouffe, and C. Itzykson, Phys. Rev. D **10**, 3376 (1974).
[87] V. A. Miransky, *Dynamical Symmetry Breaking in Quantum Field Theories* (World Scientific, Singapore, 1993), Chap. 10, and references therein.