A model of a 2*d* non-Fermi liquid with *SO*(5) symmetry, AF order and a d-wave SC gap

2009 J. Phys. A: Math. Theor. 42 025402

(http://iopscience.iop.org/1751-8121/42/2/025402)

J. Phys. A: Math. Theor. **42** (2009) 025402 (47pp) doi:10.1088/1751-8113/42/2/025402

A model of a $2d$ non-Fermi liquid with $SO(5)$ symmetry, AF order and a d-wave SC gap

Eliot Kapit and André LeClair

Newman Laboratory, Cornell University, Ithaca, NY, USA

Received 9 July 2008
Published 2 December 2008
Online at stacks.iop.org/JPhysA/42/025402

Abstract
Demanding a consistent quantum field theory description of spin-$\frac{1}{2}$ particles near a circular Fermi surface in $2d$ leads to a unique fermionic theory with relevant quartic interactions which has an emergent Lorentz symmetry and automatically has an $Sp(4) = SO(5)$ internal symmetry. The interacting theory has a low-energy interacting fixed point and is thus a non-Landau/Fermi liquid. Anti-ferromagnetic (AF) and superconducting (SC) order parameters are bilinears in the fields and form the five-dimensional vector representation of $SO(5)$. An AF phase occurs at low doping which terminates in a first-order transition. We incorporate momentum-dependent scattering of Cooper pairs near the Fermi surface to 1-loop and derive a new kind of SC gap equation beyond mean field with a d-wave gap solution. Taking into account the renormalization group (RG) scaling properties near the low-energy fixed point, we calculate the complete phase diagram as a function of doping, which shows some universal geometric features. The d-wave SC dome terminates on the over-doped side at the fixed point of the RG, which is a quantum critical point. Optimal doping is estimated to occur just below $3/2\pi^2$. The critical temperature for SC at optimal doping is set mainly by the universal nodal Fermi velocity and lattice spacing, and is estimated to average around 140 K for LSCO. The pseudogap energy scale is identified with the RG scale of the coupling.

PACS numbers: 71.10.Ay, 11.10.−z, 11.30.−j

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In the renormalization group (RG) framework, Landau's theory of Fermi liquids is characterized by the irrelevance of the interactions of particles near the Fermi surface, in other words the low-energy fixed point is simply a free theory of fermions. The underlying reasons for the wide success of Landau/Fermi liquid theory are well understood [1–5], and consequently the known models of non-Landau/Fermi liquids are relatively rare. (Henceforth referred to simply as non-Fermi liquids.) An important exception is the Luttinger liquid and other related models consisting of quartic interactions of Dirac fields in $d = 1$ spatial dimension. Here the non-Fermi liquid behavior can be attributed to the fact that in $1d$, quartic interactions of Dirac fields are marginal operators in the RG sense. In higher dimensions quartic interactions of Dirac fermions are irrelevant and this is one of the reasons why candidate non-Fermi liquid models were not found in the previous works. Whereas more exotic non-Fermi liquid models have been proposed which typically involve gauge fields, the lack of non-Fermi liquid models in $2d$ appears paradoxical when one considers even the simplest models of itinerant electrons with quartic interactions, such as the Hubbard or t-J model, which are believed to be at strong coupling. Since such models have been proposed as good starting points for thinking about high-T_c superconductivity in the cuprates [6, 7, 9], it is certainly worthwhile to continue to try and construct relatively non-exotic models of continuum fermions with quartic interactions that have some resemblance to the Hubbard model and have non-Fermi liquid behavior in the normal state.

Though the search for a novel kind of non-Fermi liquid in $2d$ provided one of the main initial motivations for the formulation of the model that will be presented and analyzed in this work, the model turns out to have many unexpected bonus features, almost all of which are intrinsic to $2d$. We list the most prominent:

- The 4-fermion interaction is unique for spin-$\frac{1}{2}$ electrons and automatically has $SO(5)$ symmetry. In $2d$ the interactions are relevant and the model has a low-energy interacting fixed point with non-classical exponents which can be computed perturbatively.
- The model generalizes to N flavors, where $N = 2$ corresponds to spin-$\frac{1}{2}$ electrons, and has $Sp(2N)$ symmetry. Since $Sp(4) = SO(5)$ this provides an underlying framework based on a microscopic theory for exploring the ideas of Zhang based on $SO(5)$ [10, 11]. In particular, one can derive the effective Ginzburg–Landau theory.
- Because of the $SO(5)$ symmetry the model naturally has both anti-ferromagnetic (AF) and superconducting (SC) order parameters that form the five-dimensional vector representation of $SO(5)$. For repulsive interactions the model has AF order and no SC order in mean field approximation.
- When one incorporates momentum-dependent scattering to 1-loop to go beyond mean field, an attractive d-wave channel opens up and the momentum dependence of the gap can be calculated. This d-wave SC phase terminates on the over-doped side at the RG fixed point, which is a quantum critical point. Due to mathematical properties of the d-wave gap equation, it also terminates on the under-doped side yielding a 'dome'. Due to the properties of the RG flow, this attractive d-wave instability exists for arbitrarily strong repulsive interactions at short distances.
- Although the model may be at arbitrarily strong coupling at short distances, the low-energy fixed point is at a relatively small coupling $\approx 1/8$, and this renders the model perturbatively calculable. We are thus able to calculate the main features of the complete phase diagram as a function of a doping variable, including the phase boundary of the d-wave superconducting dome and estimate the optimal doping fraction, which is near $3/2\pi^2 \approx 0.15$. This phase diagram depends on a single parameter $0 < \gamma < 1$ which encodes the ratio of the strength of the coupling at short versus long distances. In figure 1 we summarize the results of our calculations for $\gamma = 1$, which corresponds to infinite coupling at short distances. The overall scale of temperature is set by the universal nodal Fermi velocity and the lattice spacing.

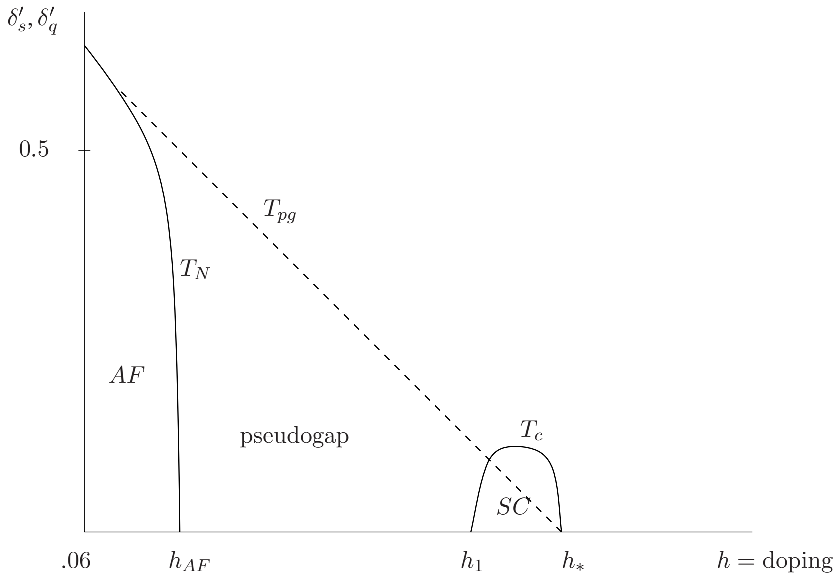


Figure 1. *Calculated* phase diagram as a function of hole doping, which depends on a single parameter $0 < \gamma < 1$ determined by the strength of the interaction at short distances. We set $\gamma = 1$ corresponding to infinitely strong coupling. The vertical axis represents the low-energy scale relative to the cut-off. What is shown are solutions δ'_s and δ'_q of the AF and d-wave gap equations (109) and (120) in units of the cut-off Λ_c. The critical temperatures are proportional to these gaps with constants of order unity which we estimate, equations (167) and (169). The overall temperature scale is determined by the universal nodal Fermi velocity and lattice spacing (171). The running coupling g in the gap equations is in terms of x in equation (147), where x is the inverse dimensionless coupling. The hole-hoping $h(x)$ is the 1-loop-corrected expression (160). The straight line $T_{\rm pg}$ is the renormalization group scale corresponding to the energy scale of the coupling, equation (144), and represents the boundary to the pseudogap region. The AF transition point at $h_{\rm AF} = 3/4\pi^2$ is first order. The SC transition at $h_* = 3/2\pi^2$ is second order and corresponds to the fixed point of the renormalization group, i.e. a quantum critical point. The transition point $h_1 \approx 0.13$ is not universal, but relies on mathematical properties of the gap equation.

Many of the above properties were highlighted on a list of the most important features of high-T_c superconductivity compiled early on in the subject [6, 9]. Because of the importance of the CuO_2 planes, high T_c is believed to be essentially a $2d$ phenomenon. This, along with the detailed properties of the solution of our model, in particular the phase diagram, led us to propose it as a model of high-T_c superconductivity [12]. If our theory turns out to be the correct description, it reveals that the phenomenon of high-T_c superconductivity is remarkably universal, with a single energy scale, and its main features follow from the existence of the low-energy fixed point in $2d$. It is truly a beautiful phenomenon that has managed to realize some subtle theoretical loopholes in the usual requirements of unitarity, the spin-statistics theorem, and the Mermin–Wagner theorem, which are only possible in $2d$. Our theory represents a significant departure from the models considered thus far in connection with high T_c, which are reviewed in [13–16], along with reviews of experimental results. On the other hand, we believe it represents a particular scaling limit of the Hubbard model at and just below half-filling, and is in this sense conservative in comparison with other more exotic ideas, and is thus in line with the early ideas concerning the role of AF order and the Heisenberg and Hubbard models [6–8]. However, our model is not simply a direct scaling limit of the Hubbard model with no attention paid to the Fermi surface, since the latter only has an $SO(4) = SU(2) \otimes SU(2)$ symmetry, whereas our theory has the $SO(5)$ symmetry. In our theory the 'fermion sign' problem is solved by doing analytic, perturbative calculations in a fermionic theory from the beginning, and relatively simple 1-loop calculations already reveal the main features.

Irregardless of whether our model has been exactly realized in the laboratory, it can serve as a useful tool for exploring many of the paradigms in the area of strongly correlated electrons and also for developing new methods. For instance, we develop new gap equations that take into account higher order scattering of Cooper pairs near the Fermi surface. Our analysis shows clearly how in $2d$ one can obtain a momentum-dependent gap with a d-wave structure from a rotationally invariant continuum field theory, i.e. without an explicit lattice that breaks the rotational symmetry. This is interesting especially since the precise origin of the d-wave symmetry of the SC gap has been unclear. We also show how to introduce doping in terms of the coupling and RG scale, and a small non-zero temperature as a relativistic mass coupling.

For the remainder of this introduction we outline the organization of the paper and summarize our main results. In section 2 we motivate the model by showing how it can approximately describe particles and holes near a circular Fermi surface. The manner in which we expand around the Fermi surface is in the same spirit as in [1–5] but differs in some important ways. For a single spin-less fermion one thereby obtains a free Hamiltonian of particles and holes with a massless, i.e. relativistic dispersion relation. In section 3 we insist on a local quantum field description of the effective theory near the Fermi surface with a consistent quantization. Since the particles are massless, the only known candidate field theories are either Dirac or 'symplectic' fermions, which differ primarily by being first order versus second order in spacetime derivatives, respectively. For the remainder of the paper we focus on symplectic fermions since unlike the Dirac fermions, the interactions are relevant. The model was first proposed in this context by one of us [17], where the groundwork was done on the low-energy non-Fermi liquid fixed point and in part the AF properties; at the time the SC properties were unknown. As explained in this paper, the central idea of this previous work, that the AF order parameter is bilinear in symplectic fermion fields and that the low-energy RG fixed point describes a quantum critical point, appears to be correct; however, as we will see, the quantum critical point terminates the SC rather than AF phase. Quantum critical points in the context of high T_c were emphasized earlier by Vojta and Sachdev [19]. The issue of the unitarity of our theory was mostly resolved in [18] by noting that the Hamiltonian is pseudo-Hermitian and this is sufficient for a unitary time evolution. In this paper, the expansion around the Fermi surface provides a new view on the pseudo-Hermiticity and it is explained how it is related to the kinematics of particles versus holes. The critical exponents were computed to 2-loops in [18], which corrected some errors in [17]. In this paper we analyze many more properties, in particular the AF and d-wave SC ordering properties for the first time.

Since the consistency of the quantization of a fermionic theory with a Lagrangian that is second order in time derivatives is at the heart of the unitarity issue, in section 4 we work out in detail the $d = 0$ dimensional quantum-mechanical case where all the subtle consistency issues are present. In this section we also construct the conserved charges for the $Sp(2N)$ symmetry. In section 5 the field theory version in d spatial dimensions is defined and spin and charge are identified for the case of $N = 2$. In this section we also define the $SO(5)$-order parameters for AF and SC order.

In section 6 we sketch an argument that the resistivity is linear in temperature in the limit of no interactions. In the following section we consider small thermal perturbations near $T = 0$. By comparing with the specific heat of a degenerate electron gas, we argue that a small non-zero temperature can be incorporated as a coupling in the Lagrangian corresponding to relativistic mass $m = \alpha T$ and we estimate the constant α.

In section 8 the mean field analysis is carried out with potential competition between AF and SC order. As we explain, these two phases actually do not compete in our model in this approximation. As a check of the formalism, we reproduce some of the basic features of the BCS theory for an s-wave gap in the case of an attractive coupling in section 9. For repulsive interactions we find only AF order is possible in mean field approximation and this is studied in section 10. There we first argue that this phase must be anti-ferromagnetic by comparing our model with the low-energy nonlinear sigma model description of the Heisenberg anti-ferromagnet. This gives another motivation for our model at half-filling away from the circular Fermi surface, and explains how the same model can interpolate between an SC phase and an AF one. We argue that the AF phase terminates in a first-order phase transition. The AF gap is then the solution of a transcendental equation that is analyzed in various limits.

In section 11, orbital symmetries of momentum-dependent gaps are studied in a model-independent way and we explain how a d-wave gap can arise. This analysis is based on a gap equation which is derived in appendix A. In section 12 we compute the 1-loop contributions to the scattering of Cooper pairs and show that at low energies the d-wave channel is attractive if the number of components $N < 3$. Since the theory is free for $N = 1$, this means that only the physically relevant $N = 2$ case has d-wave SC. This also means that the d-wave pairing cannot be studied with large N methods.

Section 13 is devoted to describing our RG prescription which is specific to $2d$. This is necessary for a proper understanding of the phase diagram. In section 14 we present global features of the phase diagram, which is characterized by some universal geometric relations, and bears a striking resemblance to the cuprates. The SC phase terminates at a second-order phase transition precisely at the low-energy RG fixed point, and is thus a quantum critical point. We also estimate optimal hole doping. In section 15 we present detailed numerical solutions to the AF and SC d-wave gap equations at non-zero temperature. For reasonable values of the lattice spacing and universal nodal Fermi velocity, we estimate $T_c \approx 140\,\mathrm{K}$ on average for SC in LSCO. In section 16 we describe the interpretation of the pseudogap within our model.

Although we do not give a complete and rigorous derivation of our model from lattice fermion models, in order to motivate and point out relations, we have collected some known results about the latter in appendix B.

2. Expansion around the Fermi surface

2.1. Kinematics

Let us first ignore spin and consider a single species of fermion described by the free Hamiltonian in momentum space

$$H = \int (\mathrm{d}^d\mathbf{k})(\varepsilon(\mathbf{k}) - \mu)c^\dagger_{\mathbf{k}}c_{\mathbf{k}}, \tag{1}$$

where μ is the chemical potential, we have defined $(\mathrm{d}^d\mathbf{k}) \equiv \mathrm{d}^d\mathbf{k}/(2\pi)^d$, and

$$\left\{c^\dagger_{\mathbf{k}}, c_{\mathbf{k}'}\right\} = (2\pi)^d\delta^{(d)}(\mathbf{k} - \mathbf{k}'). \tag{2}$$

At finite density and zero temperature, all states with $\varepsilon \leqslant \varepsilon_F$ are filled, where the Fermi energy ε_F depends on the density, and at zero temperature $\mu = \varepsilon_F$. The Fermi surface $\mathcal{S}_F$ is the manifold of points $\mathbf{k}_F$ satisfying $\varepsilon(\mathbf{k}_F) = \varepsilon_F$.

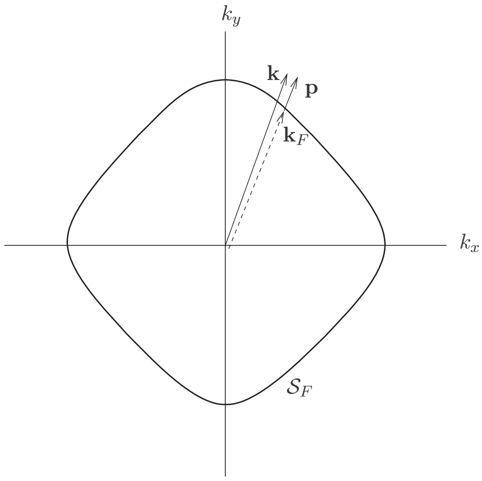


Figure 2. Expansion around the Fermi surface.

We wish to consider a band of energies near ε_F as shown in figure 2 for $d = 2$. Let $\mathbf{k}$ be any wave vector in such a band, and let $r(\mathbf{k})$ denote a ray from the origin to infinity along the direction of $\mathbf{k}$. We further assume that the Fermi surface is sufficiently smooth, such that $r(\mathbf{k})$ intersects $\mathcal{S}_F$ only once. The latter implies that $\mathbf{k}$ can be *uniquely* expressed as

$$\mathbf{k} = \mathbf{k}_F(\mathbf{k}) + \mathbf{p}(\mathbf{k}), \tag{3}$$

where $\mathbf{k}_F(\mathbf{k})$ is the vector from the origin to the intersection of $r(\mathbf{k})$ with $\mathcal{S}_F$. Whereas the two vectors $\mathbf{k}$ and $\mathbf{k}_F(\mathbf{k})$ are by construction parallel, the vector $\mathbf{p}(\mathbf{k})$ is either parallel or anti-parallel to $\mathbf{k}$. Let us fix $\mathbf{p}$ to be a small vector parallel to $\mathbf{k}$, i.e. pointing radially outward, as shown in figure 2. Since now $\mathbf{k}_F(\mathbf{k})$ is uniquely determined by $\mathbf{p}$, we may write $\mathbf{k}_F(\mathbf{p})$. Furthermore, since the particles below the Fermi surface correspond to $-\mathbf{p}$, the energies near the Fermi surface are approximately given by

$$\varepsilon(\mathbf{k}) = \varepsilon_F \pm \mathbf{p} \cdot \mathbf{v}_F(\mathbf{k}), \tag{4}$$

where $\pm$ corresponds to above or below the Fermi surface, and

$$\mathbf{v}_F(\mathbf{k}) = \vec{\nabla}\varepsilon(\mathbf{k})|_{\mathbf{k}_F} \tag{5}$$

is the Fermi velocity normal to $\mathcal{S}_F$.

Let us now assume that the Fermi surface is rotationally invariant, i.e. $\varepsilon(\mathbf{k})$ depends only on $|\mathbf{k}|$. In $2d$ the Fermi surface $\mathcal{S}_F$ is thus a circle. This leads to the simplification that $k_F = |\mathbf{k}_F(\mathbf{k})|$ and $v_F = |\mathbf{v}_F(\mathbf{k})|$ are independent of $\mathbf{k}$. For any $\mathbf{k}$ in the band,

$$\mathbf{k} = (k_F \pm p)\widehat{p}, \tag{6}$$

where $p = |\mathbf{p}|$ and $\mathbf{p} = p\widehat{p}$. Furthermore, since $\mathbf{v}_F$ is normal to $\mathcal{S}_F$, the energies are linear in $|\mathbf{p}|$:

$$\varepsilon(\mathbf{k}) = \varepsilon_F \pm v_F|\mathbf{p}|. \tag{7}$$

For non-relativistic particles with $\varepsilon(\mathbf{k}) = \mathbf{k}^2/2m_*$, $v_F = k_F/m_*$.

Since the map from $\mathbf{k}$ to $\mathbf{p}$ is one-to-one, we can define the following operators:

$$a_\mathbf{p} = c_{\mathbf{k}_F+\mathbf{p}}, \qquad b_\mathbf{p} = c^\dagger_{\mathbf{k}_F-\mathbf{p}}, \tag{8}$$

where it is implicit that $\mathbf{k}_F$ depends on $\mathbf{p}$. The above is a canonical transformation, since

$$\{a^\dagger_{\mathbf{p}}, a_{\mathbf{p}'}\} = \{b^\dagger_{\mathbf{p}}, b_{\mathbf{p}'}\} = (2\pi^d)\delta^{(d)}(\mathbf{p} - \mathbf{p}'). \tag{9}$$

After normal ordering, the Hamiltonian for the particles in the band is defined to be

$$H = \int_{|\mathbf{p}|<\Lambda_c} (\mathrm{d}^d\mathbf{p})\big[(v_F|\mathbf{p}| - \widehat{\mu})a^\dagger_{\mathbf{p}}a_{\mathbf{p}} + (v_F|\mathbf{p}| + \widehat{\mu})b^\dagger_{\mathbf{p}}b_{\mathbf{p}}\big], \tag{10}$$

where $\widehat{\mu} = \mu - \varepsilon_F$ is zero at zero temperature. For the remainder of this paper we mostly set $\widehat{\mu} = 0$. Since we are only interested in a band of energies near the Fermi surface, we have introduced a cut-off Λ_c. The vacuum $|0\rangle$ is defined to satisfy $a_{\mathbf{p}}|0\rangle = b_{\mathbf{p}}|0\rangle = 0$. This corresponds to $c_{\mathbf{k}_F+\mathbf{p}}|0\rangle = c^\dagger_{\mathbf{k}_F-\mathbf{p}}|0\rangle = 0$, which correctly implies that all states with $\varepsilon < \varepsilon_F$ are filled. The $a_{\mathbf{p}}$ and $b_{\mathbf{p}}$ thus correspond to particles and holes, respectively.

There is an approximation made in obtaining the above Hamiltonian having to do with the density of states, and this is crucial to understanding how our expansion differs from previous works. In the rotationally invariant case, for particles above the Fermi surface:

$$\int \mathrm{d}^d\mathbf{k} = \int \mathrm{d}\Omega \int \mathrm{d}p(k_F + p)^{d-1}, \tag{11}$$

where $d\Omega$ are angular integrals. Note that due to equation (6) the angular integrals for $\mathbf{k}$ and $\mathbf{p}$ are identical. At least two approximations to the above are meaningful. The first favors low energies where one approximates $k_F + p \approx k_F$. This is the approximation that is commonly made in the literature [1, 2]. On the other hand, expanding out the $(k_F + p)^{d-1}$, at high energies the leading term is p^{d-1}, and is the most sensitive to the short-distance physics and spatial dimensionality. A possible shortcoming of the first, low-energy approximation is that the high-energy physics is discarded from the beginning. It cannot be recovered by the RG flow to low energies since the latter is irreversible. In the physical problem we are considering, the short-distance physics of the strong Coulomb repulsion is known to be important for understanding the AF phase, so it makes sense to adopt an approximation that favors high energies from the beginning and to then incorporate their effects by an RG flow to lower energies. We thus keep the most important term at short distances and set $\mathrm{d}^d\mathbf{k} = \mathrm{d}^d\mathbf{p}$, i.e. $k_F + p \approx p$. This is in line with the usual RG idea that it is important to fix the high-energy physics as accurately as possible, and then flow down to lower energies. Finally, our choice is necessary for the $2d$ effective field theory description in the following section. However, one should not conclude that every theory with a circular Fermi surface can be described by a relativistic field theory. One signature of a relativistic description is a density of states that is linear in energy: $\int \mathrm{d}^2\mathbf{p} = 2\pi v_F \int \mathrm{d}\varepsilon\, \varepsilon$.

The above expansion around the Fermi surface is thus not identical to the expansion in [1–5], where the integration over $\mathbf{p}$ is taken to be normal to $\mathcal{S}_F$ times the angular integrations, and it is assumed that $p \ll k_F$. This leads to the choice $\int \mathrm{d}^d\mathbf{k} = \int \mathrm{d}\Omega \int \mathrm{d}p\, k_F^{d-1}$, and the constant k_F is absorbed into the definitions of the operators. Thus in the approach followed in [1–5], although the angular integrals obviously depend on d, the scaling analysis of the p dependence leads to marginal 4-fermion interactions for any d, and the resulting theory is effectively one-dimensional, or a collection of such theories, one for each angular direction. Notably, it was not possible to obtain a non-Fermi liquid based on 4-fermion interactions in this approach [2].

In contrast, in the approach developed in this paper there is a strong dependence on d, as in other critical phenomena, and this will turn out to be very important. In particular, it leads to a non-Fermi liquid in $2d$. There are other important justifications for this choice. In particular,

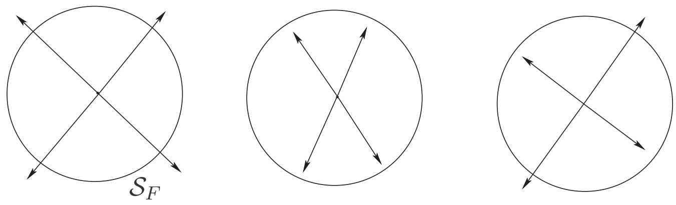


Figure 3. Allowed processes.

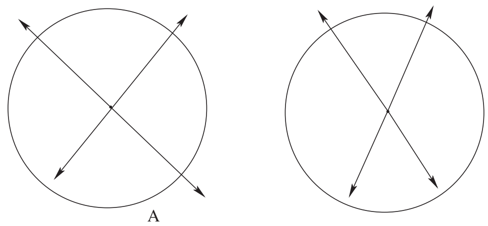


Figure 4. Unallowed processes.

near half-filling where the interacting lattice model can be mapped to the Heisenberg antiferromagnet, there is known to be a relativistic description of the low-energy, long-wavelength limit in terms of the $O(3)$ nonlinear sigma model. It will be shown in section 10 that our choice of field theory near the circular Fermi surface can be extrapolated to half-filling in that an independent derivation of it can be provided exactly at half-filling.

For general processes, physical momentum conservation of the $\mathbf{k}$'s is not equivalent to $\mathbf{p}$ conservation. However, consider a zero-momentum process proportional to $\delta(\sum_i \mathbf{k}_i)$. If the $\mathbf{k}$'s are all exactly on the Fermi surface, then this implies $\sum_i \mathbf{k}_F(\mathbf{k}_i) = 0$. For even numbers of particles, since all vectors on the Fermi surface have the same length, this is satisfied by pairs of particles with opposite $\mathbf{k}_F$. Allowing now small deviations $\mathbf{p}_i$ from the Fermi surface, one has

$$\delta\left(\sum_i \mathbf{k}_i\right) = \delta\left(\sum_{\text{particles}} \mathbf{p}_i - \sum_{\text{holes}} \mathbf{p}_i\right). \tag{12}$$

Because of the particle/hole transformation for the b's in equation (8), this is equivalent to overall $\mathbf{p}$ conservation. Note that by construction it is not possible for the momentum of a particle and a hole to add up to zero, so in the above δ-function, holes are paired with other holes, and particles with other particles. Therefore, spatial translational invariance of our local field theory will ensure physical momentum conservation of the $\mathbf{k}$'s for this class of processes.

The important allowed processes are shown in figure 3. Examples of an unallowed process are shown in figure 4. The distinction between the allowed and unallowed processes can be made explicit by introducing an operator C that distinguishes particles and holes:

$$C a_{\mathbf{p}} C = a_{\mathbf{p}}, \qquad C b_{\mathbf{p}} C = -b_{\mathbf{p}}, \tag{13}$$

where C is a unitary operator satisfying $C = C^\dagger$ so that $C^2 = 1$. An eigenstate with pairs of particles and/or pairs of holes is then required to have $C = 1$. We will return to this in connection with the pseudo-Hermiticity of symplectic fermions in the sequel.

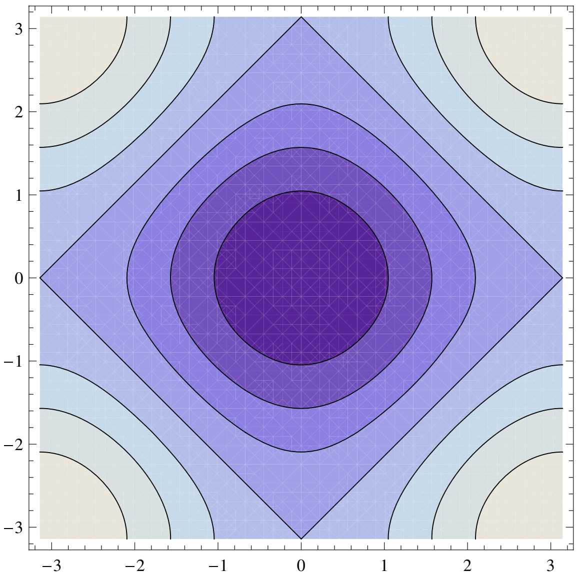


Figure 5. Fermi surface contours for free lattice fermions in $2d$ with lattice spacing $a = 1$.

2.2. *Lattice fermions*

In the following, our field theory model will be related to lattice models of itinerant electrons such as the Hubbard model, although we do not claim a precise equivalence. The known $2d$ square lattice model results we will need to make the comparison are all contained in appendix B. In this section we consider only the free, hopping term. In momentum space the 1-particle energy is (B.5)

$$\varepsilon_{\mathbf{k}} = -2t(\cos k_x a + \cos k_y a), \tag{14}$$

where a is the lattice spacing. Equal energy contours in the first Brillouin zone are shown in figure 5. The Fermi surface at half-filling is the square diamond with corners on the x, y axes. Note that one does not have to be very far below half-filling for the contours to be approximately circular. The free local field theory model in the following section can thus be viewed as an approximate effective theory for free particles on the lattice below half-filling.

An important point is that our model is not simply a direct continuum limit of the lattice model since, without additional care, the latter does not take into account the Fermi surface at finite density. For instance, whereas the Hubbard model has at most an $SO(4) = SU(2) \otimes SU(2)$ symmetry [36], our continuum model has the larger $SO(5)$ symmetry. Furthermore, as will be explained in section 10, the success of our model can be attributed to the fact that an alternative justification of it can be given right at half-filling so that it can actually interpolate between half-filling and below.

3. Requirements on the free local field theory

The main requirements we impose for a local field theory description of the last section are as follows:

(i) The theory has a Lagrangian description with a consistent quantization.

(ii) In momentum space the Hamiltonian reduces to equation (10) for particles and holes of energy $v_F|\mathbf{p}|$. The latter is a relativistic dispersion relation for massless particles.

In order to motivate our arguments, let us start from non-relativistic particles with $\varepsilon_{\mathbf{k}} = \mathbf{k}^2/2m_*$. The second-quantized description consists of a single field $\Psi(\mathbf{x}, t)$ with Lagrangian

$$L = \int \mathrm{d}^d\mathbf{x} \left(\mathrm{i}\Psi^\dagger \partial_t \Psi - \frac{1}{2m_*} \vec{\nabla}\Psi^\dagger \cdot \vec{\nabla}\Psi \right). \tag{15}$$

The field has the momentum space expansion

$$\Psi(\mathbf{x}, t) = \int (\mathrm{d}^d\mathbf{k}) c_{\mathbf{k}}\, \mathrm{e}^{-\mathrm{i}\varepsilon_{\mathbf{k}} + \mathrm{i}\mathbf{k}\cdot\mathbf{x}}. \tag{16}$$

Expanding around the Fermi surface one finds

$$\Psi(\mathbf{x}, t) = \mathrm{e}^{-\mathrm{i}\varepsilon_F t} \int (\mathrm{d}^d\mathbf{p}) \widetilde{\Psi}_{\mathbf{p}}(\mathbf{x}, t)\, \mathrm{e}^{\mathrm{i}\mathbf{k}_F(\mathbf{p})\cdot\mathbf{x}}, \tag{17}$$

where

$$\widetilde{\Psi}_{\mathbf{p}}(\mathbf{x}, t) = a_{\mathbf{p}}\, \mathrm{e}^{-\mathrm{i}v_F|\mathbf{p}|t + \mathrm{i}\mathbf{p}\cdot\mathbf{x}} + b^\dagger_{\mathbf{p}}\, \mathrm{e}^{\mathrm{i}v_F|\mathbf{p}|t - \mathrm{i}\mathbf{p}\cdot\mathbf{x}}. \tag{18}$$

We wish to find an effective theory for $\widetilde{\Psi}$, which satisfies the relativistic wave equation:

$$\left(\partial_t^2 - v_F^2 \vec{\nabla}^2\right) \widetilde{\Psi}_{\mathbf{p}}(\mathbf{x}, t) = 0. \tag{19}$$

Thus, due to the kinematics of the expansion around the Fermi surface we identify an emergent Lorentz symmetry. In $2d$ this Lorentz symmetry is $SO(3)$. The case of $d = 1$ is special in that the Fermi surface consists of only two disconnected points $\mathbf{k}_F = \pm k_F$, and decomposition (17) naturally separates into left and right movers. For higher dimensions there is no such separation since all points on the Fermi surface are related by spatial rotations and are continuously connected. There are many additional reasons why $d = 1$ is the exceptional case, and these will be pointed out where appropriate in the sequel since some of the literature attempts to draw analogies between $1d$ and $2d$.

There are only two known candidate field theories which differ in whether the Lagrangian is first or second order in derivatives. First consider the case of first order. One then needs to factor the operator $\partial_t^2 - \vec{\nabla}^2$ into two first-order multiples. The only way to accomplish this is to promote $\widetilde{\Psi}$ to a multi-component field and introduce a matrix representation $\gamma^\mu, \mu = 0, 1, \ldots, d$, of the Clifford algebra:

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}, \tag{20}$$

where $\eta^{\mu\nu} = \mathrm{diag}(1, -1, ., -1)$. One then has

$$\sum_{\mu,\nu} \gamma^\mu \gamma^\nu \partial_\mu \partial_\nu = \sum_\mu \partial^\mu \partial_\mu = \partial_t^2 - \vec{\nabla}^2. \tag{21}$$

(We have adopted the relativistic notation $x^\mu = (x^0, x^1, \ldots, x^d) = (t, \mathbf{x})$ and $\partial_\mu = \frac{\partial}{\partial x^\mu}$, and henceforth, repeated indices are implicitly summed over.) The Lagrangian is then the standard first-order Dirac Lagrangian:

$$L = \int \mathrm{d}^d\mathbf{x}\, \mathrm{i}\overline{\psi}\gamma^\mu \partial_\mu \psi, \tag{22}$$

where $\overline{\psi} = \psi^\dagger \gamma^0$. The smallest representation of the Clifford algebra is two-dimensional: $\gamma^0 = \sigma_z, \gamma^1 = \mathrm{i}\sigma_x$ and $\gamma^2 = \mathrm{i}\sigma_y$, where $\vec{\sigma}$ are the standard Pauli matrices. In this simplest case, although there is a doubling of components, they are constrained by the Dirac equation of motion and the spectrum still consists of one species of particles and holes with the Hamiltonian equation (10).

Interactions are relevant to the low-energy physics if the operator characterizing them has scaling dimension less than $d + 1$. The classical scaling dimension of the Dirac field ψ is $d/2$ in d spatial dimensions, thus a quartic interaction has dimension $2 \cdot d$ and is thus only relevant for $d < 1$. It is not even perturbatively renormalizable for $d > 1$. Thus, the Dirac theory should lead to ordinary Landau–Fermi liquid behavior in $2d$. It is noteworthy that once again $d = 1$ is special, and this helps to explain how for instance the $1d$ Hubbard model can be mapped onto interacting Dirac fermions, and the low-energy fixed point found using special bosonization techniques and spin–charge separation [35]. Simply based on the fact that the interactions in the Hubbard model are strong in $2d$, one can rule out a description in terms of Dirac fields with quartic interactions since the latter are irrelevant. Furthermore, it is already understood that one normally needs additional special properties in order to obtain the first-order Dirac theory. For example, it is known to arise when one expands around special nodes (Dirac points) on the Fermi surface for a hexagonal lattice [21], as in graphene, and the multiple components of the Dirac field correspond to different sub-lattices. There is no reason to expect this here for a square lattice, as in the cuprates.

The other candidate field theory is of second order in derivatives, with kinetic term

$$S = \int \mathrm{d}t\, \mathrm{d}^d\mathbf{x}\, \partial_\mu \chi^- \partial^\mu \chi^+. \tag{23}$$

For $\chi^\pm$ fermionic (Grassman) fields, this is a very unconventional theory, since it potentially has problems with the spin-statistics theorem and unitarity; in high-energy elementary particle theory it usually corresponds to ghost fields. These issues will be discussed in detail and resolved completely in the following two sections. Here, let us give the main arguments for why this should be the right starting point:

(i) As shown in the following two sections, the free theory in momentum space corresponds precisely to the Hamiltonian (10) for particles and holes near a circular Fermi surface. This is of course a perfectly Hermitian theory with no negative norm states.

(ii) The fundamental field χ has scaling dimension $(d-1)/2$ and thus quartic interactions have dimension $2(d-1)$ which is actually relevant for $d < 3$. Thus it can have non-Fermi liquid behavior.

(iii) Although we are led to consider this model for the nearly circular Fermi surface below half-filling, a simple argument leads to the same model at half-filling. It is well known that a low-energy description of excitations above the staggered AF state is described by the $O(3)$ nonlinear sigma model for a field $\vec{\phi}$ constrained to have constant length with the action

$$S = \int \mathrm{d}t\, \mathrm{d}^d\mathbf{x}\, \partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi}. \tag{24}$$

(See appendix B.) In our model the anti-ferromagnetic order parameter $\vec{\phi}$ is bilinear in the fields $\vec{\phi} = \chi^- \vec{\sigma} \chi^+ / \sqrt{2}$. The nonlinear constraint on the $\vec{\phi}$ fields follow from imposing a similar constraint on the χ fields: $\chi^- \chi^+ =$ constant. This was pointed out in [17]. Inserting this into the above action one finds that one obtains the second-order action (23) for the χ fields up to irrelevant operators (equation (104)). Thus the symplectic fermion model with interactions can in principle describe AF order, and in the sequel we will show that this is indeed the case. This is explained in more detail in section 10.

Since v_F only serves to convert dimensions of time and space, it plays the role of the speed of light; we can set it to unity since it can always be restored by dimensional analysis.

The important point here is that v_F is fixed and universal in our model, in particular it does not depend on the coupling. In the following the doping will be related to the coupling, so v_F does not depend on doping either. For high-T_c materials, we believe the 'speed of light' has actually already been measured [20]. The Fermi surface at half-filling is closest to the nearly circular surface below it in the nodal $\mathbf{k} = (0, 0)$ to (π, π) direction, so our v_F should correspond to this nodal Fermi velocity. Remarkably, the latter was measured to be universal at low energies, i.e. independent of doping in [20]. Whereas this universality of v_F has not been explained theoretically up to now, it is a necessary aspect of our theory. Taking the slope $\partial E/\partial k$ of the curves in [20] we estimate $v_F \approx 1.4\,\text{ev Å} = 210\,\text{km s}^{-1}$ for LSCO. As we will show in section 15, this gives very reasonable estimates of T_c. Another signature that the system may be in a relativistic regime is a density of states that is linear in energy, as explained in the last section.

4. Symplectic fermion quantum mechanics

As stated above, symplectic fermions are primarily characterized by a Lagrangian that is second-order in space *and* time derivatives. Since this is unfamiliar to most readers and there are some delicate issues in the quantization of such theories, let us first start with the simplest case of $d = 0$ quantum mechanics.

4.1. Canonical quantization

In order to draw comparisons, let us first consider a first-order Lagrangian as in the Dirac theory:

$$L = \sum_{\alpha=1}^{N} \left(\mathrm{i} c_\alpha^\dagger \partial_t c_\alpha - \omega c_\alpha^\dagger c_\alpha \right). \tag{25}$$

It is well understood that this Lagrangian has two consistent quantizations, i.e. one can impose either canonical commutation relations $\left[c_\alpha, c_\beta^\dagger\right] = \delta_{\alpha,\beta}$ or canonical anti-commutation relations $\left\{c_\alpha, c_\beta^\dagger\right\} = \delta_{\alpha,\beta}$. In both cases the Hamiltonian is $H = \sum_\alpha \omega c_\alpha^\dagger c_\alpha$. It is clear that both options are possible since equation (15) is a proper second quantized description of either bosons or fermions. For future reference we note that the model has a manifest $SU(N)$ symmetry.

The second-order bosonic version of the above is just the ordinary harmonic oscillator with $L = ((\partial_t q)^2 - \omega^2)/2$. Since the first-order Lagrangian can be consistently quantized as a fermion or boson, one expects that the second-order case should also be quantizable as a fermion, and as we now describe, this is indeed the case. In order to have a fermionic version, we need at least 2 degrees of freedom since fermionic variables square to zero. Let us therefore consider the Lagrangian

$$L = \dot{\chi}^- \dot{\chi}^+ - \omega^2 \chi^- \chi^+, \tag{26}$$

where χ are Grassman variables:

$$\{\chi^i, \chi^j\} = 0, \tag{27}$$

which implies $(\chi^-)^2 = (\chi^+)^2 = 0$, and we have defined $\dot{\chi} = \partial_t \chi$. The canonical momenta are $p^- = \partial L/\partial \dot{\chi}^- = \dot{\chi}^+$ and $p^+ = \partial L/\partial \dot{\chi}^+ = -\dot{\chi}^-$, which leads to the canonical anti-commutation relations

$$\{\chi^-, \dot{\chi}^+\} = -\{\chi^+, \dot{\chi}^-\} = \mathrm{i}. \tag{28}$$

The canonical Hamiltonian is simply

$$H = \dot{\chi}^- \dot{\chi}^+ + \omega^2 \chi^- \chi^+. \tag{29}$$

The equation of motion is $\left(\partial_t^2 + \omega^2\right)\chi = 0$. Because this is second order, the mode expansion involves both positive and negative frequencies:

$$\begin{aligned}
\chi^-(t) &= \frac{1}{\sqrt{2\omega}}\left(a^\dagger\, \mathrm{e}^{-\mathrm{i}\omega t} + b\, \mathrm{e}^{\mathrm{i}\omega t}\right),\\
\chi^+(t) &= \frac{1}{\sqrt{2\omega}}\left(-b^\dagger\, \mathrm{e}^{-\mathrm{i}\omega t} + a\, \mathrm{e}^{\mathrm{i}\omega t}\right).
\end{aligned} \tag{30}$$

The canonical anti-commutation relations (28) then require

$$\{a, a^\dagger\} = \{b, b^\dagger\} = 1 \tag{31}$$

with all other anti-commutators equal to zero. The Hamiltonian is

$$H = \omega(a^\dagger a + b^\dagger b - 1). \tag{32}$$

4.2. Pseudo-Hermiticity

The only subtle aspect of the above quantization is the extra minus sign in the expansion of χ^+ in equation (30), which was necessary in order to have the canonical relations (31). This minus sign implies that χ^+ is not the Hermitian conjugate of χ^-. One can understand this feature more clearly, and also keep track of it, with the operator C that distinguishes particles and holes in equation (13):

$$\chi^+ = C(\chi^-)^\dagger C. \tag{33}$$

In terms of the original χ variables, the Hamiltonian is pseudo-Hermitian, $H^\dagger = CHC$. However, after using the equations of motion and expressing it in terms of a, b's, since it is quadratic in b's, the Hamiltonian (32) is actually Hermitian. This issue will be revisited when interactions are introduced in the following section.

4.3. Symmetries

We now study the symmetries of the N-copy theory. Introduce variables χ_α^i, $i = -, +$, $\alpha = 1, 2, \ldots, N$ and define the Lagrangian

$$L = \frac{1}{2}\sum_{i,j,\alpha} \epsilon_{ij}\left(\dot{\chi}_\alpha^i \dot{\chi}_\alpha^j - \omega^2 \chi_\alpha^i \chi_\alpha^j\right), \tag{34}$$

where ϵ_{ij} is the 2×2 anti-symmetric matrix $\epsilon_{-+} = -\epsilon_{+-} = 1$. The Hamiltonian is

$$H = \frac{1}{2}\sum_{i,j,\alpha} \epsilon_{ij}\left(\dot{\chi}_\alpha^i \dot{\chi}_\alpha^j + \omega^2 \chi_\alpha^i \chi_\alpha^j\right). \tag{35}$$

Arrange χ_α^i into a $2N$-component vector and consider the transformation $\chi \to M\chi$, where M is a $2N$-dimensional matrix. Then the Lagrangian is invariant if $M^t \epsilon_N M = \epsilon_N$ where $\epsilon_N = \epsilon \otimes 1_N$, and M^t is the transpose. This implies that M is an element of the group $Sp(2N)$ of dimension $N(2N+1)$. Interestingly, of the classical Lie groups, $Sp(2N)$ is the only one that does not play any known role in elementary particle physics [25]. Note that the bosonic version of the theory with $2N$ real components has the symmetry $O(2N)$ with dimension $N(2N-1)+1$, thus the fermionic version always has a larger symmetry. The N-component symplectic fermion also has a larger symmetry than the N-component first-order fermionic action which has a $U(N)$ symmetry or an $O(2N)$ symmetry if the complex fermions are rewritten in terms of $2N$ real fields.

The conserved charges that generate the $N(2N+1)$-dimensional Lie algebra of $Sp(2N)$ are easily constructed. Define

$$\begin{aligned} Q^0_{\alpha\beta} &= -\mathrm{i}\big(\chi^-_\alpha\dot\chi^+_\beta + \chi^+_\beta\dot\chi^-_\alpha\big), \\ Q^-_{\alpha\beta} &= -\mathrm{i}\big(\chi^-_\alpha\dot\chi^-_\beta + \chi^-_\beta\dot\chi^-_\alpha\big), \\ Q^+_{\alpha\beta} &= \mathrm{i}\big(\chi^+_\alpha\dot\chi^+_\beta + \chi^+_\beta\dot\chi^+_\alpha\big). \end{aligned} \tag{36}$$

One can easily verify that the Hamiltonian H commutes with all the charges Q using

$$\begin{aligned} \big\{\chi^-_\alpha, \dot\chi^+_\beta\big\} &= -\big\{\chi^+_\alpha, \dot\chi^-_\beta\big\} = \mathrm{i}\delta_{\alpha\beta}, \\ \big\{\chi^i_\alpha, \chi^j_\beta\big\} &= \big\{\dot\chi^i_\alpha, \dot\chi^j_\beta\big\} = 0. \end{aligned} \tag{37}$$

In 'momentum' space, the charges are

$$\begin{aligned} Q^0_{\alpha\beta} &= a^\dagger_\alpha a_\beta - b^\dagger_\beta b_\alpha, \\ Q^-_{\alpha\beta} &= a^\dagger_\alpha b_\beta + a^\dagger_\beta b_\alpha, \\ Q^+_{\alpha\beta} &= b^\dagger_\alpha a_\beta + b^\dagger_\beta a_\alpha \end{aligned} \tag{38}$$

and they satisfy the Hermiticity properties

$$\big(Q^0_{\alpha\beta}\big)^\dagger = Q^0_{\beta\alpha}, \qquad \big(Q^-_{\alpha\beta}\big)^\dagger = Q^+_{\alpha\beta}. \tag{39}$$

5. Field theory version with spin and interactions

5.1. Lagrangian and Hamiltonian

The field theory version in d spatial dimensions follows straightforwardly from the above $d = 0$ case with the addition of spatial or momentum integrals. Introducing fields $\chi^i_\alpha(\mathbf{x}, t)$, the action is

$$S = \frac{1}{2}\int \mathrm{d}t\, \mathrm{d}^d\mathbf{x} \sum_{i,j,\alpha} \epsilon_{ij}\big(\partial^\mu\chi^i_\alpha\partial_\mu\chi^j_\alpha - m^2\chi^i_\alpha\chi^j_\alpha\big) \tag{40}$$

and the equations of motion are

$$(\partial^\mu\partial_\mu + m^2)\chi = 0. \tag{41}$$

The momentum space expansion is

$$\begin{aligned} \chi^-(\mathbf{x}, t) &= \int \frac{(\mathrm{d}^d\mathbf{p})}{\sqrt{2\omega_\mathbf{p}}}\big(a^\dagger_\mathbf{p}\,\mathrm{e}^{-\mathrm{i}p\cdot x} + b_\mathbf{p}\,\mathrm{e}^{\mathrm{i}p\cdot x}\big), \\ \chi^+(\mathbf{x}, t) &= \int \frac{(\mathrm{d}^d\mathbf{p})}{\sqrt{2\omega_\mathbf{p}}}\big(-b^\dagger_\mathbf{p}\,\mathrm{e}^{-\mathrm{i}p\cdot x} + a_\mathbf{p}\,\mathrm{e}^{\mathrm{i}p\cdot x}\big), \end{aligned} \tag{42}$$

where $\omega_\mathbf{p} = \sqrt{\mathbf{p}^2 + m^2}$ and $p\cdot x \equiv \omega_\mathbf{p} t - \mathbf{p}\cdot\mathbf{x}$. (We do not display the α indices since they just correspond to N identical copies.) The canonical anti-commutation relations are

$$\{\chi^-(\mathbf{x}, t), \dot\chi^+(\mathbf{x}', t)\} = -\{\chi^+(\mathbf{x}, t), \dot\chi^-(\mathbf{x}', t)\} = \mathrm{i}\delta^{(d)}(\mathbf{x} - \mathbf{x}') \tag{43}$$

which in momentum space leads to

$$\big\{a_\mathbf{p}, a^\dagger_{\mathbf{p}'}\big\} = \big\{b_\mathbf{p}, b^\dagger_{\mathbf{p}'}\big\} = (2\pi)^d\delta^{(d)}(\mathbf{p} - \mathbf{p}'). \tag{44}$$

(All other anti-commutators are zero.)

The Hamiltonian is

$$H = \int (\mathrm{d}^d \mathbf{p}) \sum_{\alpha=1}^{N} \omega_{\mathbf{p}} \big(a^{\dagger}_{\mathbf{p},\alpha} a_{\mathbf{p},\alpha} + b^{\dagger}_{\mathbf{p},\alpha} b_{\mathbf{p},\alpha} \big). \tag{45}$$

In the limit $m \to 0$, we obtain the effective Hamiltonian near the Fermi surface in equation (10), as desired. The mass m in this section is an infrared regulator and is unrelated to the non-relativistic mass m_* above. In section 7 we will show that it can be viewed as proportional to the temperature.

The field theory has the same $Sp(2N)$ symmetry as the quantum-mechanical version. The expressions for the conserved charges are identical to equations (36) and (38) with additional integrals over $\mathbf{x}$ or $\mathbf{p}$.

There is a unique 4-fermion interaction that preserves the $Sp(2N)$ symmetry:

$$S_{\text{int}} = -\pi^2 g \int \mathrm{d}t\, \mathrm{d}^d \mathbf{x} (\chi^- \epsilon_N \chi^+)^2 = -4\pi^2 g \int \mathrm{d}t\, \mathrm{d}^d \mathbf{x} \left(\sum_{\alpha} \chi^-_{\alpha} \chi^+_{\alpha} \right)^2. \tag{46}$$

(We have included an overall π^2 so that the RG equations below for g have no π's; our convention is the same as in [18].) For $N = 2$, even without imposing the $Sp(4)$ symmetry, there is a unique interaction due to fermionic statistics since there are only four independent fields, which implies that higher order terms beyond quartic interactions are zero by Fermi statistics. This interaction is automatically $SO(5)$ invariant. Positive g corresponds to repulsive interactions. Since the field χ has classical scaling dimension $(d-1)/2$, the interaction is a dimension $2(d-1)$ operator which is relevant for $d < 3$ as previously mentioned.

5.2. *Pseudo-Hermiticity*

The symplectic fermion action (40) has a Lorentz invariance if χ is understood to be a Lorentz scalar. Since χ is fermionic, to a particle physicist this model would appear to violate the spin-statistics theorem. There are two separate aspects of this issue. First of all, in the condensed-matter context, rotational spin is an internal flavor symmetry (spin-$\frac{1}{2}$ fermions corresponds to $N = 2$) which is not viewed as embedded in the Lorentz group. This implies that spin-$\frac{1}{2}$ particles are not forced to be described by the first-order Dirac theory. There is no violation of the spin-statistics connection in our theory since we are quantizing spin-$\frac{1}{2}$ particles with fermionic fields. The potential problem rather has to do with unitarity, as explained in the quantum-mechanical case studied in the last section, and is manifested in the pseudo-Hermiticity property (33). This explains how the proof of the spin-statistics connection is circumvented: the proof assumes that the Hamiltonian is built out of fields and their Hermitian conjugates and thus does not allow for different fields being related by pseudo-Hermitian conjugation [5]. Furthermore, in the end the free Hamiltonian in momentum space is a perfectly Hermitian theory with no negative norm states.

Whereas the free theory is Hermitian in momentum space, for the interacting theory, it follows from (33) that the Hamiltonian is pseudo-Hermitian:

$$H^{\dagger} = CHC, \tag{47}$$

where the unitary operator $C = C^{\dagger}$ and $C^2 = 1$. This sort of generalization of Hermiticity was understood to give a consistent quantum mechanics long ago by Pauli [22], and more recently in connection with $\mathcal{PT}$-symmetric quantum mechanics [23, 24]. Let us summarize the main properties enjoyed by pseudo-Hermitian Hamiltonians:

(i) Define a C-Hermitian conjugation as follows:

$$A^{\dagger_c} = CA^\dagger C. \tag{48}$$

Then the usual rules are satisfied:

$$(AB)^{\dagger_c} = B^{\dagger_c}A^{\dagger_c}, \qquad (aA + bB)^{\dagger_c} = a^* A^{\dagger_c} + b^* B^{\dagger_c}. \tag{49}$$

(ii) Define a C-conjugate inner product:

$$\langle \psi' | \psi \rangle_c \equiv \langle \psi' | C | \psi \rangle. \tag{50}$$

Then time-evolution is unitary:

$$\langle \psi'(t) | \psi(t) \rangle_c = \langle \psi' | \, \mathrm{e}^{\mathrm{i}H^\dagger t} C \, \mathrm{e}^{-\mathrm{i}Ht} | \psi \rangle = \langle \psi'(0) | \psi(0) \rangle_c. \tag{51}$$

(iii) The eigenvalues of H are real:

$$(E - E^*) \langle \psi_E | \psi_E \rangle_c = \langle \psi_E | (CH - H^\dagger C) | \psi_E \rangle = 0. \tag{52}$$

(iv) Diagonal matrix elements of C-pseudo-Hermitian operators $A = A^{\dagger_c}$ are real. This will be important in the sequel since it guarantees the reality of vacuum expectation values of pseudo-Hermitian order parameters. For our model

$$H^{\dagger_c} = H, \qquad (\chi^-)^{\dagger_c} = \chi^+, \qquad (\chi^- \chi^+)^{\dagger_c} = \chi^- \chi^+. \tag{53}$$

In the present work there is a new aspect of the pseudo-Hermiticity that relates to the kinematics of the expansion around the Fermi surface. As explained in section 2, for the 4-particle processes near the Fermi surface, conservation of $\mathbf{p}$ relative to the Fermi surface is equivalent to the conservation of the physical $\mathbf{k}$ momentum if particles are paired with particles and holes with holes. Since $C = \pm 1$ for particles versus holes, on physical grounds we should restrict to eigenstates with even numbers of holes and even numbers of particles with $C = 1$. Let $|\psi_E\rangle$ denote an eigenstate of H which is also an eigenstate of C. Then $H^\dagger |\psi_E\rangle = C^2 H |\psi_E\rangle = H|\psi_E\rangle$. Thus for the eigenstates of interest, $H = H^\dagger$.

5.3. Charge and spin

In order to describe spin-$\frac{1}{2}$ electrons, as usual we treat the spin as a flavor and thus consider the $N = 2$ theory. The symmetry of the free theory is $Sp(4) = SO(5)$. A subgroup of this large (ten-dimensional) symmetry can be identified with rotational spin and charge.

It was pointed out in [18] that there are potentially two ways to identify electronic spin, and the focus in that work was the $SU(2)$ subalgebra that exists for all N and acts on the $\pm$ indices of $\chi_\alpha^\pm$. It turns out that the other identification is more natural in the present context since, as explained in section 2, the two components corresponding to the $\pm$ indices are already necessary for the expansion of a single spin-less fermion near the Fermi surface. There is actually an analog of the identification of spin and charge for arbitrary N. Let $M = \mathrm{e}^{-m}$ be an element of the group $Sp(2N)$ in the defining $2N$-dimensional representation. Using the relation

$$M^t \epsilon_N M = \epsilon_N, \tag{54}$$

the elements of the Lie algebra satisfy

$$m^t \epsilon_N = -\epsilon_N m, \tag{55}$$

and a basis is the following: $m \in \{1 \otimes a, \sigma_x \otimes a_x, \sigma_y \otimes a_y, \sigma_z \otimes a_z\}$, where σ_i are the Pauli matrices, a is an $N \times N$ dimensional anti-symmetric matrix and a_i are $N \times N$ symmetric matrices [25]. Clearly, $Sp(2N)$ has an $SU(2)^{\otimes N}$ sub-algebra generated by $\sigma_i \otimes I^{(\alpha)}$, where

$I^{(\alpha)} = \mathrm{diag}(0, 0, \ldots, 1, 0, \ldots, 0)$ with the 1 in the αth entry. For $N = 2$ this corresponds to an $SO(4) = SU(2) \otimes SU(2)$ symmetry. However, since this sub-algebra does not mix the flavors, it is not the right sub-group for identification with spin as a flavor. The correct identification is the $SU(N)$ sub-algebra generated by $1 \otimes a$ and $\sigma_z \otimes a_z$ where now a_z is also traceless. There is also a $U(1)$ which commutes with the $SU(N)$ corresponding to $\sigma_z \otimes 1_N$. The $SU(N)$ is generated by the charges $Q^{(0)}_{\alpha\beta}$, which form a closed algebra:

$$\left[Q^0_{\alpha\beta}, Q^0_{\alpha'\beta'}\right] = \mathrm{i}\left(\delta_{\alpha\beta'} Q^0_{\alpha'\beta} - \delta_{\beta\alpha'} Q^0_{\alpha\beta'}\right). \tag{56}$$

The $U(1)$ we identify with electric charge is generated by $Q_e = \sum_\alpha Q^0_{\alpha\alpha}$.

Let us now return to the physically interesting case of $N = 2$ and label the two components as $\alpha = \uparrow, \downarrow$, corresponding to up and down spins. The $SU(2)$ spin symmetry is generated by

$$\begin{aligned} Q_z &= \frac{1}{2}\left(Q^0_{\uparrow\uparrow} - Q^0_{\downarrow\downarrow}\right) = -\frac{\mathrm{i}}{2}\int \mathrm{d}^d\mathbf{x}\left(\chi^-_\uparrow \dot\chi^+_\uparrow + \chi^+_\uparrow \dot\chi^-_\uparrow - \chi^-_\downarrow \dot\chi^+_\downarrow - \chi^+_\downarrow \dot\chi^-_\downarrow\right) \\ &= \int (\mathrm{d}^d\mathbf{p})\left(\frac{1}{2}\left(a^\dagger_{\mathbf{p}\uparrow} a_{\mathbf{p}\uparrow} - a^\dagger_{\mathbf{p}\downarrow} a_{\mathbf{p}\downarrow}\right) - \frac{1}{2}\left(b^\dagger_{\mathbf{p}\uparrow} b_{\mathbf{p}\uparrow} - b^\dagger_{\mathbf{p}\downarrow} b_{\mathbf{p}\downarrow}\right)\right), \\ Q_+ &= \frac{1}{\sqrt{2}} Q^0_{\uparrow\downarrow} = -\frac{\mathrm{i}}{\sqrt{2}}\int \mathrm{d}^d\mathbf{x}\left(\chi^-_\uparrow \dot\chi^+_\downarrow + \chi^+_\downarrow \dot\chi^-_\uparrow\right) = \frac{1}{\sqrt{2}}\int (\mathrm{d}^d\mathbf{p})\left(a^\dagger_{\mathbf{p}\uparrow} a_{\mathbf{p}\downarrow} - b^\dagger_{\mathbf{p}\downarrow} b_{\mathbf{p}\uparrow}\right), \\ Q_- &= \frac{1}{\sqrt{2}} Q^0_{\downarrow\uparrow} = -\frac{\mathrm{i}}{\sqrt{2}}\int \mathrm{d}^d\mathbf{x}\left(\chi^-_\downarrow \dot\chi^+_\uparrow + \chi^+_\uparrow \dot\chi^-_\downarrow\right) = \frac{1}{\sqrt{2}}\int (\mathrm{d}^d\mathbf{p})\left(a^\dagger_{\mathbf{p}\downarrow} a_{\mathbf{p}\uparrow} - b^\dagger_{\mathbf{p}\uparrow} b_{\mathbf{p}\downarrow}\right) \end{aligned} \tag{57}$$

satisfying the $SU(2)$ Lie algebra:

$$[Q_z, Q_\pm] = \pm Q_\pm, \qquad [Q_+, Q_-] = Q_z. \tag{58}$$

(As before $\dot\chi = \partial_t \chi$.) The $U(1)$ charge is generated by

$$\begin{aligned} Q_e = Q^0_{\uparrow\uparrow} + Q^0_{\downarrow\downarrow} &= -\mathrm{i}\int \mathrm{d}^d\mathbf{x}\left(\chi^-_\uparrow \dot\chi^+_\uparrow + \chi^+_\uparrow \dot\chi^-_\uparrow + \chi^-_\downarrow \dot\chi^+_\downarrow + \chi^+_\downarrow \dot\chi^-_\downarrow\right) \\ &= \int (\mathrm{d}^d\mathbf{p})\left(\left(a^\dagger_{\mathbf{p}\uparrow} a_{\mathbf{p}\uparrow} + a^\dagger_{\mathbf{p}\downarrow} a_{\mathbf{p}\downarrow}\right) - \left(b^\dagger_{\mathbf{p}\uparrow} b_{\mathbf{p}\uparrow} + b^\dagger_{\mathbf{p}\downarrow} b_{\mathbf{p}\downarrow}\right)\right). \end{aligned} \tag{59}$$

The conserved electric current corresponding to the above charge is

$$J^e_\mu = -\mathrm{i}\sum_{\alpha=\uparrow,\downarrow}\left(\chi^-_\alpha \partial_\mu \chi^+_\alpha + \chi^+_\alpha \partial_\mu \chi^-_\alpha\right). \tag{60}$$

The fields $\chi^\pm$ have electric charge $Q_e = \pm 1$. Commutations of the fields with the $SU(2)$ generators show that $\left(\chi^-_\uparrow, \chi^-_\downarrow\right)$ form a doublet whereas $\left(\chi^+_\uparrow, \chi^+_\downarrow\right)$ is the conjugate:

$$\left[Q_z, \chi^\pm_\uparrow\right] = \mp\tfrac{1}{2}\chi^\pm_\uparrow, \qquad \left[Q_z, \chi^\pm_\downarrow\right] = \pm\tfrac{1}{2}\chi^\pm_\downarrow. \tag{61}$$

Note that the above identification is consistent with a being particles and b holes: $Q_z|a_\uparrow\rangle = \frac{1}{2}|a_\uparrow\rangle$ whereas $Q_z|b_\uparrow\rangle = -\frac{1}{2}|b_\uparrow\rangle$. Also, the a-particles have $Q_e = 1$ whereas the b have $Q_e = -1$.

The additional six conserved charges with electric charge ± 2 that complete the $SO(5)$ Lie algebra are

$$Q^\pm_{\uparrow\uparrow} = -2\mathrm{i}\int \mathrm{d}^d\mathbf{x}\,\chi^\pm_\uparrow \dot\chi^\pm_\uparrow, \qquad Q^\pm_{\downarrow\downarrow} = -2\mathrm{i}\int \mathrm{d}^d\mathbf{x}\,\chi^\pm_\downarrow \dot\chi^\pm_\downarrow, \tag{62}$$

$$Q^\pm_{\uparrow\downarrow} = -\mathrm{i}\int \mathrm{d}^d\mathbf{x}\left(\chi^\pm_\uparrow \dot\chi^\pm_\downarrow + \chi^\pm_\downarrow \dot\chi^\pm_\uparrow\right). \tag{63}$$

These symmetries flip the charge and spin of the fields, for instance

$$\left[Q^+_{\uparrow\uparrow}, \chi^-_\uparrow\right] = 2\chi^+_\uparrow, \qquad \left[Q^+_{\uparrow\downarrow}, \chi^-_\uparrow\right] = \chi^+_\downarrow, \qquad \left[Q^+_{\uparrow\downarrow}, \chi^-_\downarrow\right] = \chi^+_\uparrow. \tag{64}$$

There is no real separation of spin and charge degrees of freedom in our model at this stage, unlike Dirac fermions in $1d$, where this separation relies on bosonization. However, as we explain in section 8, the Goldstone bosons are electrically neutral but carry spin quantum numbers.

5.4. $SO(5)$-order parameters

The four fields $\chi_\uparrow^\pm$, $\chi_\downarrow^\pm$ are in the four-dimensional spinor representation of $SO(5)$. (See, e.g. [25].) The important operators are bilinears, which decompose as $\mathbf{4}\otimes\mathbf{4} = \mathbf{1}\oplus\mathbf{5}\oplus\mathbf{10}$, where $\mathbf{1}$ is the singlet $\chi^-\chi^+$, the $\mathbf{5}$ is the vector representation and $\mathbf{10}$ is the adjoint. The $\mathbf{10}$ corresponds to the currents constructed above. The $\mathbf{5}$ will serve as the order parameters of our model in the sequel. A triplet of fields $\vec{\phi} = (\phi_x, \phi_y, \phi_z)$ transforming under the three-dimensional vector representation of the spin $SU(2)$ is the following:

$$\phi^+ = \chi_\uparrow^-\chi_\downarrow^+, \qquad \phi^- = \chi_\downarrow^-\chi_\uparrow^+, \qquad \phi_z = \tfrac{1}{\sqrt{2}}\left(\chi_\uparrow^-\chi_\uparrow^+ - \chi_\downarrow^-\chi_\downarrow^+\right), \tag{65}$$

where $\phi^\pm = (\phi_x \pm \mathrm{i}\phi_y)/\sqrt{2}$. Note that these fields are electrically neutral. Let us also define two $SU(2)$ singlets that carry electric charge $Q_e = \pm 2$:

$$\phi_e^+ = \chi_\uparrow^+\chi_\downarrow^+, \qquad \phi_e^- = \chi_\downarrow^-\chi_\uparrow^-. \tag{66}$$

Using the commutation relations of the $SO(5)$ charges with the χ-fields, one finds that the five order parameters

$$\vec{\Phi} = \left(\phi_x, \phi_y, \phi_z, \phi_e^+, \phi_e^-\right) \tag{67}$$

transform under the five-dimensional vector representation of $SO(5)$. For instance $\left[Q_{\uparrow\uparrow}^+, \phi_e^-\right] = 2\phi^-$ and $\left[Q_{\uparrow\uparrow}^+, \phi^+\right] = 2\phi_e^+$. The ordering of fermionic operators was chosen such that $(\vec{\Phi})^{\dagger_c} = \vec{\Phi}$, more precisely $\vec{\phi}^{\dagger_c} = \vec{\phi}$ and $\left(\phi_e^+\right)^{\dagger_c} = \phi_e^-$, which guarantees that vacuum expectation values of $\vec{\phi}$ are real and those of $\phi_e^\pm$ are complex conjugates. An $SO(5)$ invariant is the following:

$$\vec{\Phi}\cdot\vec{\Phi} \equiv \phi^+\phi^- + \phi^-\phi^+ + \phi_z^2 - \phi_e^+\phi_e^- - \phi_e^-\phi_e^+. \tag{68}$$

Finally, the interaction Lagrangian density equation (46) can be expressed as

$$\mathcal{L}_{\mathrm{int}} = \frac{8\pi^2 g}{5}\vec{\Phi}\cdot\vec{\Phi} = -8\pi^2 g\,\chi_\uparrow^-\chi_\uparrow^+\chi_\downarrow^-\chi_\downarrow^+. \tag{69}$$

5.5. Low-energy fixed point

The Feynman rules for the theory are the same as for a Lorentz-invariant scalar with ϕ^4 interaction, which in practice is quite simple [5, 26], *except for some all important minus signs*. In the following we mainly work in Euclidean space $t \to -\mathrm{i}t$, since this simplifies solving the gap equations and is also appropriate for finite temperature. Where appropriate we will return to real time (Minkowski space) for certain physical quantities. In particular, the propagators in Euclidean space are

$$\left\langle\chi_\alpha^-(x)\chi_\beta^+(0)\right\rangle = -\left\langle\chi_\alpha^+(x)\chi_\beta^-(0)\right\rangle = -\delta_{\alpha\beta}\int\frac{\mathrm{d}^D p}{(2\pi)^D}\,\mathrm{e}^{-\mathrm{i}p\cdot x}\frac{1}{p^2+m^2}, \tag{70}$$

where $D = d+1$, and they respect causality. In the usual Euclidean conventions the 4-vertex for the interaction in (46) is $-4\pi^2 g$. (See figure 9.)

Our RG prescription will be specialized precisely to $2d$ and is described in detail in section 13 after some 1-loop diagrams are explicitly calculated. In order to obtain a clear physical picture, it will turn out to be important to carry out the RG directly in $2d$. However, the RG can also be studied perturbatively in an epsilon expansion around $d = 3$. For future reference we summarize the main results obtained in [17, 18] from the epsilon expansion. We will not use these results very much in the sequel; they are included as a guide to the sign and strength of the anomalous corrections to scaling. Since the perturbative expansion differs from that of bosonic scalars only by fermionic minus signs, the standard methods for the bosonic $O(M)$ vector models apply. Specifically, these are models of an M-vector of scalar fields $\vec{\phi}$ with interaction $(\vec{\phi}\cdot\vec{\phi})^2$. In $D = d + 1 = 3$, the low-energy theory is the Wilson–Fisher fixed point [27] describing classical $3d$ magnets. The anomalous dimensions of the most important operators were computed to 2-loops in [18]. As it turns out, most of the results can be obtained from known results for the $O(M)$ models analytically continued to $M = -2N$. Besides the χ field itself and mass term $\chi^-\chi^+$ (which will be the thermal perturbation), an additional important class of operators are the bilinears that correspond to order parameters, in particular the $\vec{\phi}$ of the $SO(5)$-order parameters when $N = 2$, of the form $\chi^-\vec{\sigma}\chi^+$ where $\vec{\sigma}$ is a Pauli matrix. These have no analog that is known to be physically meaningful for the $O(M)$ models. It should be emphasized that the quantum critical exponents of our $N = 2$ model are completely unrelated to the usual Wilson–Fisher exponents of the bosonic $O(3)$ model, since here the $O(3)$ vector is a composite bilinear operator and its scaling dimension must be calculated in the fundamental χ-theory.

Let $[\![X]\!]$ denote the scaling dimension of X in inverse length, i.e. energy units. One can also define a correlation length exponent based on the mass m:

$$\xi \sim m^{-\nu}, \tag{71}$$

where $\nu = 1/[\![m]\!]$:

$$\nu^{-1} = (d + 1 - [\![\chi^-\chi^+]\!])/2. \tag{72}$$

Specializing the results in [18] to $N = 2$ and $2d$ one obtains

$$[\![\chi]\!] \approx \tfrac{15}{32}, \qquad [\![\chi^-\chi^+]\!] \approx \tfrac{5}{8}, \qquad [\![\chi^-\vec{\sigma}\chi^+]\!] \approx \tfrac{3}{2}, \qquad \nu \approx \tfrac{16}{19}. \tag{73}$$

Note that whereas $\chi^-\chi^+$ decreases in dimension from the classical value 1, $\chi^-\vec{\sigma}\chi^+$ increases.

Some remarks concerning the $d = 1$ case are again appropriate. It is well known that the low-energy fixed points of the bosonic $O(M)$ vector models do not extend down to $d = 1$, i.e. $D = 2$. For general M this can be viewed as a manifestation of the Mermin–Wagner result [29], which states that spontaneous symmetry breaking in a D-dimensional Euclidean field theory is not possible in $D = 2$. On the other hand, the models do have a conformally invariant fixed point for $-2 < M < 2$ in $2D$ [28]. Since our N-component symplectic fermion model is formally equivalent in perturbation theory to $O(-2N)$, this suggests that our models do not extend to $d = 1$ except possibly for $-1 < N < 1$. This helps to explain why for instance it has never been considered before in connection with lattice fermion models in $1d$.

6. Resistivity in the normal state

In this section we give a rough calculation of the temperature dependence of the conductivity when the interactions are negligible. The rigorous study of this question requires a full finite-temperature treatment of the Kubo formula, which is known to be quite subtle, and therefore beyond the scope of this paper. We hope to return to this in a future work, however in this section we present the following non-rigorous scaling argument.

Start with a version of the zero-temperature Kubo formula

$$\sigma_{ij}(\mathbf{q}, \omega) = \frac{1}{\omega} \int_0^\infty \mathrm{d}t\, \mathrm{e}^{\mathrm{i}\omega t} \int \mathrm{d}^2\mathbf{x}\, \mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}} \langle [J_i(\mathbf{x}, t), J_j(0)] \rangle, \tag{74}$$

where $i, j = x, y$ are the spatial components of the currents J_μ given in (60).

It is well known that the frequency $\omega = 0$, which corresponds to the DC conductivity, is a delicate limit. Let us first set $\omega = 0$ in the integrand. Using the propagators in equation (70) one finds

$$\int \mathrm{d}t \int \mathrm{d}^2\mathbf{x}\, \langle \chi^- \partial_\mu \chi^+(x) \chi^- \partial_\nu \chi^+(0) \rangle = \int_0^{\Lambda_c} \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{p_\mu p_\nu}{p^4}, \tag{75}$$

where we have set $m = 0$. The above integral is proportional to Λ_c. Since the conductivity σ is dimensionless in $2d$, one must have that it is proportional to Λ_c/T where T is the temperature. For reasons that are unclear, this apparently amounts to setting the overall $1/\omega$ in equation (74) equal to T. Taking into account the two spin components, and the four terms of the above form, one obtains

$$\sigma_{xx} = \frac{4}{3\pi^2} \frac{\Lambda_c}{T}. \tag{76}$$

Therefore, the resistivity $(1/\sigma)$ is linear in the temperature T.

7. Thermal perturbations and anomalous specific heat

The parameters of our model thus far are the Fermi velocity v_F, which we have set equal to 1, the cut-off Λ_c, the coupling g and the infrared regulator mass m. The Fermi energy ε_F can be viewed as implicit in the cut-off if the latter is taken to be the frequency $\omega_D = v_F k_F$. The density of the free electron gas is only a function of k_F:

$$\frac{N}{V} = 2 \int_0^{k_F} \frac{\mathrm{d}^d\mathbf{k}}{(2\pi)^d} = \frac{4\pi^{d/2}}{\Gamma(d/2)} k_F^d. \tag{77}$$

Thus, the density can be varied by varying the cut-off, and since g is proportional to the cut-off in $2d$, equivalently by varying g. (See the discussion of the RG in section 13.)

In this section we suggest how to introduce a *small* non-zero temperature as the mass m. This is similar to how temperature appears in the Landau theory for continuous phase transitions in $O(M)$ magnets [30], where there also temperature corresponds to a coupling in an effective action. However, there are some important differences, since in the latter the coupling is $T - T_c$, whereas here it will be proportional to T^2. Since we have built our model by expanding around the zero-temperature Fermi surface, it is not obviously consistent to incorporate a finite temperature by starting over and formulating a finite-temperature version of the χ fields with Matsubara frequency summations, ignoring that they are effective fields. Rather, since a small non-zero temperature amounts to a small distortion of the Fermi surface, it could correspond to an additional coupling in the Lagrangian. If this is correct, then it should be possible to obtain known results in the limit where the interaction is turned off. Suggestive of this possibility is the fact that in natural units the leading contribution to the specific heat of a degenerate electron gas can be expressed entirely in terms of k_F and T, i.e. the dependence on the electron mass is only through the Fermi velocity $v_F = k_F/m_*$. We emphasize that this way of introducing temperature is expected to be valid for temperatures near zero, and cannot replace a full-fledged finite-temperature formalism at arbitrary temperatures. It will however be useful for exploring the temperature dependence of low-temperature gaps in the sequel.

In Euclidean space, introducing a finite temperature T is known to correspond to compactifying the Euclidean time to a circle of circumference $1/T$. Let us consider adding a thermal perturbation to the Euclidean action of the form

$$\delta S = \int \mathrm{d}t\, \mathrm{d}^d\mathbf{x}\, g_T \mathcal{O}_T(x), \tag{78}$$

where g_T is a coupling and $\mathcal{O}_T$ is the 'thermal operator'. Since the Euclidean functional integral is over e^{-S}, the correction to the free energy $F = -T \log Z$ to lowest order in g_T is $\delta F = T g_T \int \mathrm{d}t\, \mathrm{d}^d\mathbf{x} \langle \mathcal{O}_T \rangle$. Since the Euclidean spacetime volume is $V^{(D)} = V/T$, where V is the spatial volume, one finds

$$\frac{\delta F}{V} = g_T \langle \mathcal{O}_T \rangle. \tag{79}$$

This leads us to identify the thermal perturbation with a mass term:

$$g_T = (\alpha T)^2, \qquad \mathcal{O}_T = \chi^- \chi^+ \equiv \sum_\alpha \chi^-_\alpha \chi^+_\alpha. \tag{80}$$

The above relation is consistent with dimensional analysis for α, a dimensionless parameter. The specific heat C_V at constant volume is $C_V = -T \frac{\partial^2 F}{\partial T^2}$. Since the propagator goes as $-1/p^2$ as $m \to 0$ (see equation (70)) one has

$$\langle \mathcal{O}_T \rangle = \sum_{\alpha=\uparrow,\downarrow} \left\langle \chi^-_\alpha \chi^+_\alpha \right\rangle = -2 \int_0^{\Lambda_c} \frac{\mathrm{d}^D p}{(2\pi)^D} \frac{1}{p^2}, \tag{81}$$

where a factor of 2 comes from spin up and down. If the cut-off Λ_c is equated with the frequency k_F, then this leads to specific heat that is linear in T:

$$\frac{C_V}{V} = \frac{\alpha^2 T}{(d-1) 2^{d-1} \pi^{(d+1)/2} \Gamma\left(\frac{d+1}{2}\right)} \frac{k_F^{d-1}}{v_F}. \tag{82}$$

(We have re-introduced the Fermi velocity for the sake of comparison.) The above dependence on T, k_F and v_F is the correct one, i.e. it is the same as for a non-relativistic degenerate electron gas near $T = 0$. Repeating the standard calculation of the specific heat of the electron gas to order T/ε_F (see for instance [30]) for arbitrary d, expressing the result in terms of k_F instead of the density, and requiring the result to match equation (82) fixes the constant α:

$$\alpha^2 = \frac{\pi^2 (d-1) \Gamma\left(\frac{d+1}{2}\right)}{3}. \tag{83}$$

For $d = 2, \alpha = \pi^{5/4}/\sqrt{6} \approx 1.7$. The case $d = 1$ is noteworthy since $\alpha = 0$, and will be commented on below.

The k_F dependence of the specific heat in equation (82) is a direct consequence of the scaling dimension $d-1$ of $\mathcal{O}_T$, which is twice the scaling dimension of χ. When one includes the quartic interaction, at the low-energy fixed point this scaling dimension has anomalous corrections and this should lead to anomalous T dependence of the specific heat. The scaling dimension of $\mathcal{O}_T$ was computed perburbatively in an epsilon-expansion [17, 18], and the results summarized in the last section for arbitrary N. For $N = 2$, $\mathcal{O}_T$ has scaling dimension approximately equal to $5/8$ in the epsilon expansion.

We can provide a naive estimate of the anomalous T dependence of the specific heat. Let $[\![\mathcal{O}_T]\!]$ denote the scaling dimension of $\mathcal{O}_T$. Then if we assume the only effect of the anomalous corrections is to replace k_F^{d-1} by $k_F^{[\![\mathcal{O}_T]\!]}$, then since C_V/V has scaling dimension d, this requires $C_V \propto T^{d-[\![\mathcal{O}_T]\!]}$. It is not clear that our previous assumption is correct however, since in general C_V could contain terms $m^x k_F^y$ with $x + y = d$. Nevertheless, using our estimate of $5/8$ for the dimension of $\mathcal{O}_T$, this gives $C_V \propto T^{11/8}$. Thus $C_V/T \propto T^{3/8}$. Though the exponent 3/8 should perhaps not be taken as very accurate, the general point is that since $[\![\mathcal{O}_T]\!]$ is shifted downward to 5/8 from the classical value 1, this shows that C_V/T should vanish as $T \to 0$. This shift downward is entirely due to a fermionic minus sign [17]. We hope to study this more carefully in future work.

Finally, it is important to note that the manner in which we have introduced temperature allows, at least computationally, for phase transitions that break the continuous $SU(2)$ and $U(1)$ symmetries in $d = 2$. The Mermin–Wagner result [29] is the statement that spontaneous symmetry breaking is not possible in a (1 + 1)-dimensional (space plus time) quantum-mechanical system at zero temperature because of infrared divergences that plague the existence of Goldstone bosons. An example of such a divergence is in equation (81) for $D = 2$. In the Matsubara approach to finite temperature, time is compactified into a circle of circumference $1/T$ and the discrete Matsubara frequencies are summed over. Thus the arguments of the theorem in principle apply to a finite-temperature system in 2+1 dimensions at each Matsubara frequency. On the other hand, in our model temperature appears as a coupling in the theory, as in classical statistical mechanics in $3d$. Furthermore, as discussed in section 5, our model breaks down in $d = 1$, since the fixed point is lost, and a manifestation of this is the vanishing of α in $d = 1$.

In the following, wherever m is non-zero, it should be thought of as representing a small non-zero temperature.

8. Mean field analysis

Because of the $SO(5)$ symmetry, the interaction term in the Lagrangian can be expressed in terms of either the magnetic or electric order parameters:

$$\mathcal{L}_{\text{int}} = -8\pi^2 g \chi_\uparrow^- \chi_\uparrow^+ \chi_\downarrow^- \chi_\downarrow^+ = 8\pi^2 g \vec{\phi}\cdot\vec{\phi}/3 = -8\pi^2 g \phi_e^+ \phi_e^- . \tag{84}$$

This implies that magnetic and SC order may in principle compete. In this section we study this in mean field approximation.

Introduce auxiliary fields $\vec{s}, q^\pm$ coupled to the order parameters with the action:

$$S_{\text{aux}} = \int \mathrm{d}t\, \mathrm{d}^d\mathbf{x} \left(\sqrt{2}\vec{s}\cdot\vec{\phi} - \frac{1}{8\pi^2 g_s}\vec{s}\cdot\vec{s} + q^+\phi_e^- + q^-\phi_e^+ - \frac{1}{8\pi^2 g_q} q^+ q^- \right). \tag{85}$$

Variations $\delta S_{\text{aux}} = 0$ imply

$$q^\pm = 8\pi^2 g_q \phi_e^\pm, \qquad \vec{s} = 8\pi^2 g_s \vec{\phi}/\sqrt{2}. \tag{86}$$

Plugging this back into the action, one finds that the interaction is recovered if

$$g_q - 3g_s/2 = -g. \tag{87}$$

The effective action for the auxiliary $\vec{s}, q^\pm$ fields follows from performing the fermionic Gaussian integrals over the χ fields. Let us pass to Euclidean space with the usual prescription $t \to -\mathrm{i}t$, $\mathrm{i}S \to -S$. We will refer to the $D = d + 1$ Euclidean coordinates as simply x. Then the effective action S_{eff} is defined as

$$\mathrm{e}^{-S_{\text{eff}}(s,q)} = \int D\chi\, \mathrm{e}^{-S_{\text{aux}}(\chi,s,q) - S_{\text{free}}(\chi)}, \tag{88}$$

where S_{free} is the free action for the χ fields. The result is

$$S_{\text{eff}} = \int \mathrm{d}^D x \left(\frac{1}{8\pi^2 g_s}\vec{s}\cdot\vec{s} + \frac{1}{8\pi^2 g_q} q^+ q^- \right) - \frac{1}{2}\mathrm{Tr}\log \mathrm{A}, \tag{89}$$

where the operator A is a differential operator that depends on $\vec{s}, q^{\pm}$. For constant $\vec{s}$ and $q^{\pm}$, the $\mathrm{Tr}\log \mathrm{A}$ can be computed in D-dimensional Euclidean momentum space since the derivatives are diagonal: $\langle p|\partial|p\rangle = \mathrm{i}p\langle p|p\rangle$. For constant fields it is meaningful to define the effective potential $V_{\mathrm{eff}} = S_{\mathrm{eff}}/V^{(D)}$, where $V^{(D)}$ is the D-dimensional volume. At finite temperature for instance $V^{(D)} = V\beta$, where V is the usual d-dimensional volume and β is the inverse temperature. Using $\langle p|p\rangle = V^{(D)}/(2\pi)^D$, one obtains

$$V_{\mathrm{eff}} = \frac{1}{8\pi^2 g_s}\vec{s}\cdot\vec{s} + \frac{1}{8\pi^2 g_q}q^+q^- - \frac{1}{2}\int \frac{\mathrm{d}^D p}{(2\pi)^D}\mathrm{Tr}\log \mathrm{A(p)}. \tag{90}$$

In the basis $\left(\chi_{\uparrow}^-, \chi_{\uparrow}^+, \chi_{\downarrow}^-, \chi_{\downarrow}^+\right)$ the anti-symmetric matrix A is the following:

$$A(p) = \begin{pmatrix} 0 & p^2+m^2-s_z & q^+ & -\sqrt{2}s^- \\ -p^2-m^2+s_z & 0 & \sqrt{2}s^+ & -q^- \\ -q^+ & -\sqrt{2}s^+ & 0 & p^2+m^2+s_z \\ \sqrt{2}s^- & q^- & -p^2-m^2-s_z & 0 \end{pmatrix}. \tag{91}$$

Using $\mathrm{Tr}\log \mathrm{A} = \log \mathrm{Det A}$, one finds

$$V_{\mathrm{eff}} = \frac{1}{8\pi^2 g_s}\vec{s}\cdot\vec{s} + \frac{1}{8\pi^2 g_q}q^+q^- - \int \frac{\mathrm{d}^D p}{(2\pi)^D}\log((p^2+m^2)^2 + q^+q^- - \vec{s}\cdot\vec{s}). \tag{92}$$

The gap equations follow from setting the variation of V_{eff} with respect to $\vec{s}$ and $q^{\pm}$ separately equal to zero. The result is the following for $m = 0$:

$$\begin{aligned} \vec{s} &= -8\pi^2 g_s \int \frac{\mathrm{d}^D p}{(2\pi)^D}\frac{\vec{s}}{p^4 + q^+q^- - s^2}, \\ q^{\pm} &= 8\pi^2 g_q \int \frac{\mathrm{d}^D p}{(2\pi)^D}\frac{q^{\pm}}{p^4 + q^+q^- - s^2}. \end{aligned} \tag{93}$$

There is no simultaneous solution with both $q^{\pm}$ and $\vec{s}$ non-zero unless one fine tunes to the $SO(5)$-invariant point $g_s = -g_q$. In fact, there is no true competition between AF and SC order in these equations, and the distinction between g_s and g_q is somewhat fictitious, since for one sign of the coupling there are AF solutions and no SC solutions, and vice versa if the sign is flipped. We thus consider solutions with either pure SC order ($s = 0$) or pure magnetic order ($q = 0$). To further clarify the structure of the gap equation in the sequel, let us separate the spatial and temporal parts of the D-dimensional momentum vector p as $p = (\omega, \mathbf{k})$. (This notation is different from that of section 2 where there $\mathbf{k}$ was a physical momentum; here and henceforth it is relative to the Fermi surface.) We also restore the mass m. For pure SC order one then obtains the gap equation

$$1 = 8\pi^2 g_q \int \frac{\mathrm{d}\omega\, \mathrm{d}^d\mathbf{k}}{(2\pi)^{d+1}}\frac{1}{(\omega^2+\mathbf{k}^2+m^2)^2 + q^2}, \tag{94}$$

where $q^+ = q^- = q$. For pure magnetic order one instead has

$$1 = -8\pi^2 g_s \int \frac{\mathrm{d}\omega\, \mathrm{d}^d\mathbf{k}}{(2\pi)^{d+1}}\frac{1}{(\omega^2+\mathbf{k}^2+m^2)^2 - s^2}. \tag{95}$$

It is important to note the asymmetry in the signs of the above SC versus magnetic gap equations, which is ultimately traced to the signs in the $SO(5)$ invariant (68). Thus, although the model has an $SO(5)$ symmetry that rotates the SC and magnetic order parameters, the gap equations are not invariant under the exchange of s and q. This implies that the AF gap is related by symmetry to a conventional s-wave gap obtained when one flips the sign of the coupling, described in the following section.

Let us now consider the implications of Goldstone's theorem. A non-zero vacuum expectation value in any one direction of the 5-vector $\vec{\phi}$ preserves an $SO(4)$ subgroup of $SO(5)$. Since $\dim(SO(5)) - \dim(SO(4)) = 4$, there are potentially four Goldstone bosons. One of these is associated with the $U(1)$, and when coupled to the electromagnetic field is eaten up by the Anderson–Higgs mechanism. This leaves a spin-1 triplet of Goldstone bosons associated with the 3-vector $\vec{\phi}$, and these are the closest thing to spinons in our model. The effective theory for these modes follows from the above effective potential (92) for $\vec{s}$. For non-constant fields it contains the kinetic energy term $\partial_\mu \vec{s} \cdot \partial_\mu \vec{s}$.

9. Conventional s-wave superconductivity in 3d for attractive interactions

It is well understood that a gap that spontaneously breaks the $U(1)$ symmetry in the presence of electromagnetic gauge potential automatically has the characteristic electromagnetic properties of a superconductor, i.e. Meissner effect, etc. This precursor to the Higgs mechanism can be understood in the original Ginzburg–Landau theory [31]. A clear explanation of this can be found in Weinberg's book [5].

We are primarily interested in positive coupling g since this corresponds to repulsive interactions. (In the following section we describe how this should correspond to the positive U Hubbard model.) However, as a check of our formalism thus far, let us consider attractive interactions, i.e. negative g, or negative U Hubbard, which should reveal the usual s-wave SC instability of the BCS theory. As our analysis shows, symplectic fermions give a proper field theoretic description of conventional s-wave superconductivity which has not been considered before.

As expected there is no s-wave superconductivity for positive g since the coupling g_q is negative and the gap equation of the last section has no solutions. For pure SC at negative g, equation (87) then implies $g_q = -g$ is positive. Specializing to $d = 3$ one should recover some of the basic features of the BCS theory. The gap equation reads

$$\frac{4}{g_q} = \log\left(1 + \Lambda_c^4/q^2\right). \tag{96}$$

As the cut-off goes to infinity, the divergence can be absorbed into the coupling by defining $1/g_q(\Lambda_c) = \log \Lambda_c$. This gives the 1-loop correction to the RG beta-function: $\beta(g_q) = -dg_q/\log \Lambda_c = g_q^2$ and is consistent with the results in [17, 18]. A more complete RG prescription will be described in section 13 where 1-loop diagrams are computed. Note also that positive g_q is marginally relevant, i.e. g_q increases as the energy is lowered. Since q has dimension 2, let us define a gap $\Delta = \sqrt{q}$ with units of energy. When the cut-off is large compared to Δ the solution to the gap equation is approximately

$$\Delta = \sqrt{q} = \Lambda_c \, \mathrm{e}^{-1/g_q} \tag{97}$$

which is characteristic of the BCS theory, i.e. Δ vanishes as g_q goes to zero, and saturates at the cut-off when g_q goes to infinity. The zero-temperature gap as a function of g_q behaves as in figure 6.

Let us now restore the infrared regulator mass m and confirm the idea that it can be viewed as proportional to the temperature, as long as the latter is close enough to zero. As described

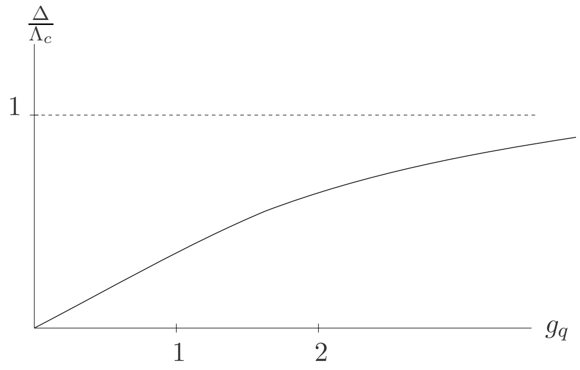


Figure 6. s-Wave gap as a function of the coupling.

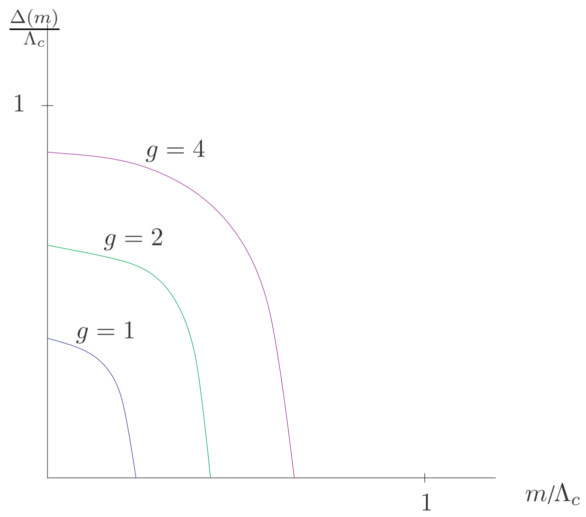


Figure 7. s-Wave gap as a function of the mass m.

in section 7, $m = \alpha T$ where $\alpha = \pi\sqrt{2/3}$ in $3d$. Restoring the mass, the gap equation now reads in $3d$

$$\frac{1}{g} = \frac{1}{4}\log\left(\frac{q^2 + \left(m^2 + \Lambda_c^2\right)^2}{q^2 + m^4}\right) + \frac{m^2}{2q}\left[\tan^{-1}(m^2/q) - \tan^{-1}\left(\left(m^2 + \Lambda_c^2\right)/q\right)\right]. \tag{98}$$

When m is too large, the rhs can go through zero and change sign, at which point the solution is lost. Numerical solutions of the gap Δ as a function of m is shown for several values of g in figure 7.

Let $m_c(g)$ denote the value of m where the gap disappears. Then, one expects the critical temperature to be given by a formula of the form

$$T_c = \frac{m_c}{\alpha} = \frac{c_g}{\alpha}\Delta_0, \tag{99}$$

Table 1. Critical mass versus zero temperature gap.

g	Δ_0/Λ_c	m_c/Λ_c	$c_g = m_c/\Delta_0$
0.5	0.135	0.083	0.61
1.0	0.370	0.235	0.64
2.0	0.629	0.440	0.70
4.0	0.873	0.660	0.76
10.0	1.19	0.99	0.83
20.0	1.46	1.27	0.87
100.0	2.22	2.09	0.94

where Δ_0 is the gap at $m = 0$ and c_g is a constant of order unity. This is similar to the relation in the BCS theory where $T_c \approx \Delta_0/1.76$. The constant $c_g = m_c/\Delta_0$ can be estimated numerically and shows a weak dependence on g, as table 1 shows.

10. Anti-ferromagnetic phase (Mott–Hubbard insulator)

10.1. Symplectic fermions from the Hubbard model at half-filling

As stated in the introduction, one cannot derive our model by taking a direct scaling limit of a lattice fermion model like the Hubbard model; this should be clear from the fact that they have different symmetries, i.e. $SO(4)$ versus $SO(5)$. The main reason for this is that one needs to take special care of the Fermi surface in taking such a continuum limit. Nevertheless, we can further motivate our model by comparison with lattice models as follows. Since our model has no explicit lattice, the meaning of the $\vec{s}\cdot\vec{s} \neq 0$ phase can only be understood by comparing with low-energy, continuum descriptions of magnetism. It is well known [32, 35] that excitations above the anti-ferromagnetic Néel state of the Heisenberg model are described by the nonlinear $O(3)$ sigma model. The basic facts about lattice models we need for this discussion are collected in appendix B. Since the $\vec{n}$ field corresponds to our $O(3)$ vector of order parameters $\vec{\phi}$ in equation (65), let us use that notation.

At half-filling and strong coupling, the Hubbard model can be mapped to the Heisenberg model and the order parameter $\vec{\phi}$ is constrained to have fixed length $\vec{\phi}\cdot\vec{\phi} =$ constant. In our model, the magnetic order parameter can be expressed as

$$\vec{\phi} = \frac{1}{\sqrt{2}}\chi^-\vec{\sigma}\chi^+, \tag{100}$$

where $\vec{\sigma}$ are the Pauli matrices. A fixed length constraint on $\vec{\phi}$ is equivalent to a similar constraint on the χ's. Using

$$\vec{\sigma}_{ij}\cdot\vec{\sigma}_{kl} = 2\delta_{il}\delta_{jk} - \delta_{ij}\delta_{kl}, \tag{101}$$

one can show

$$\vec{\phi}\cdot\vec{\phi} = -\tfrac{3}{2}(\chi^-\chi^+)^2, \tag{102}$$

where $\chi^-\chi^+ = \sum_{\alpha=\uparrow,\downarrow}\chi^-_\alpha\chi^+_\alpha$. Therefore, if one imposes

$$\chi^-_\uparrow\chi^+_\uparrow + \chi^-_\downarrow\chi^+_\downarrow = \mathrm{i}h\Lambda_c \tag{103}$$

for some constant h, then $\vec{\phi}\cdot\vec{\phi} = 3h^2\Lambda_c^2/2$. (The i is consistent with the propagators in Minkowski space.)

At half-filling there is a simple argument that leads to symplectic fermions. Imposing the above half-filling constraints in the Lagrangian:

$$\partial\vec{\phi}\cdot\partial\vec{\phi} = 4\mathrm{i}h\Lambda_c(\partial\chi^-\partial\chi^+) + \text{irrelevant operators}, \tag{104}$$

where the additional irrelevant operators can be expressed in terms of dimension 4 current–current interactions $J_\mu J^\mu$. Thus, due to the second-order nature of the nonlinear sigma model, one obtains a fermionic theory that is second order in space and time derivatives. Furthermore, let us recall now that we showed in sections 2 and 3 how symplectic fermions also arise when one expands around a circular Fermi surface. Since for lattice fermions the nearly circular Fermi surface occurs inside the half-filling diamond, this shows how symplectic fermions can actually extend below half-filling. Above half-filling (electron rather than hole doping) the same formalism applies where now one expands around the circular Fermi surface centered on the node $\mathbf{k} = (\pi, \pi)$.

Relaxing constraint (103) moves us away from half-filling. A meaningful measure of the degree of filling that we will utilize in the sequel is the following. Specializing formula (81) to $2d$ in Euclidean space and with cut-offs, one finds to zeroth order:

$$h \equiv -\frac{1}{\Lambda_c}\langle\chi^-\chi^+\rangle = -2\int_\Lambda^{\Lambda_c}\frac{\mathrm{d}^3 p}{(2\pi)^3}\frac{1}{p^2} = \frac{1}{\pi^2}\left(1 - \frac{\Lambda}{\Lambda_c}\right). \tag{105}$$

Thus, increasing Λ/Λ_c corresponds to lowering the density below half-filling, so that h is a measure of hole doping. We will return to this point when we analyze the phase diagram as a function of doping in section 14, and will also include 1-loop corrections.

The coupling g should be proportional to the Hubbard coupling U (see appendix B). It can be estimated below half-filling near the circular Fermi surface. From the lattice kinetic energy equation (14), $\varepsilon(\mathbf{k}) \approx ta^2\mathbf{k}^2$ for small k, which gives a Fermi velocity $v_F = 2ta^2k_F$. Since g has inverse length units, $U = v_F g$ has units of energy. This gives

$$g = \frac{U}{2t(k_F a)}\frac{1}{a}, \tag{106}$$

where a is the lattice spacing.

10.2. Solutions of the gap equation

We now study the solutions of the pure AF order gap equation (95) at $m = 0$, which should correspond to zero temperature. Because of the two additional minus signs in the AF gap equation in comparison with the attractive SC one studied in the last section, the interpretation of its solutions is somewhat subtle, and perhaps the most delicate point in this whole paper. First of all, the gap equation definitely has solutions for g_s positive due to the two compensating minus signs. However, because of the pole at $p^4 = s^2$ we require that $s > \Lambda_c^2$ where Λ_c is the upper momentum cut-off. A proper understanding of the solutions requires the introduction of both UV and IR cut-offs and to implement the RG directly in $2d$. The precise description of our RG prescription is postponed until we calculate some explicit Feynman diagrams and is described in section 13.

When $g_q = 0$ equation (87) gives the relation between g and g_s and they are both positive. However, the resulting relation is purely classical and specific to the auxiliary field construction. Instead we fix the relation between g_s and g by requiring consistency with the perturbative RG. To lowest order this requires $g_s = 2g$, in order for equation (110) to be consistent with equation (141). The gap equation now reads

$$\frac{1}{g} = -8\int_0^{\Lambda_c}\mathrm{d}p\,\frac{p^2}{p^4 - s^2}. \tag{107}$$

Let us define

$$s = \delta_s^2\Lambda_c^2. \tag{108}$$

The result of doing the integral is

$$\frac{\Lambda_c}{g} = \frac{4}{\delta_s}\left(\frac{1}{2}\log\left(\frac{\delta_s+1}{\delta_s-1}\right) - \tan^{-1} 1/\delta_s\right). \tag{109}$$

Since the log-term is positive in the above equation, there are solutions for $\delta_s > 1$. In fact, when the coupling is lowered, the log-term dominates, and the solution approaches $\delta_s = 1^+$ and remains there for arbitrarily small g. This behavior is not physically sensible since the gap should vanish as the coupling g vanishes. To resolve this puzzle, first note that when $s = 0$ the gap equation has IR divergences that need to be regulated. Let us therefore introduce a low-energy cut-off Λ. Setting $s = 0$ in equation (107), the result can be written as

$$\frac{1}{g} + \frac{8}{\Lambda} = \frac{8}{\Lambda_c}. \tag{110}$$

The left-hand side is precisely an expansion of the running coupling $g(\Lambda)$, since by equation (137), $1/g(\Lambda) \approx (1 + 8g/\Lambda)/g$. Denoting the solution as g_{AF},

$$g_{\mathrm{AF}}(\Lambda) = \frac{\Lambda_c}{8}. \tag{111}$$

Our interpretation of this value of the coupling g_{AF} is that s should vanish for $g < g_{\mathrm{AF}}$ since $s = 0$ is a consistent solution at $g = g_{\mathrm{AF}}$.

The value of the coupling g_{AF} is closely related, but not the same as the low-energy fixed point value of g. (See in section 13.) Namely, at 1-loop the fixed point is at $g_* = \widehat{g}_*\Lambda$ where $\widehat{g}_* = 1/8$, thus the low-energy fixed point occurs at $g_* = \frac{\Lambda}{\Lambda_c} g_{\mathrm{AF}}$. As we will see, the low-energy quantum critical point at g_* is a second-order continuous phase transition that terminates the super-conducting phase on the over-doped side. This leads us to propose that the termination point g_{AF} of the AF phase is a first-order transition, i.e. the gap drops discontinuously to zero.

Let us now return to solutions of the gap equation with $s \neq 0$. The solution in the two asymptotic limits $g/\Lambda_c \to 0, \infty$ have simple expressions. It is useful to use the identity $\frac{1}{2}\log(\delta_s + 1)/(\delta_s - 1) = \tanh^{-1} 1/\delta_s$. When $g \to \infty, \delta_s$ is large. Using $\tanh^{-1} 1/\delta_s - \tan^{-1} 1/\delta \approx 2\big/\big(3\delta_s^3\big)$, one obtains

$$\delta_s \approx \left(\frac{8g}{3\Lambda_c}\right)^{1/4}, \qquad (g/\Lambda_c \to \infty). \tag{112}$$

As g decreases, δ_s saturates to 1 since the argument of the $\tanh^{-1}$ must be less than 1 otherwise the solution is complex. Where the solution starts to flatten out can be approximated by extrapolating equation (112) down to $\delta_s = 1$, i.e. around $g/\Lambda_c = 3/8$. Using $\tanh^{-1} 1/\delta_s \approx \frac{1}{2}\log(2/(\delta_s - 1))$ when $\delta_s \approx 1$, one obtains

$$\delta_s \approx 1 + 2\,\mathrm{e}^{-\Lambda_c/2g}, \qquad (g/\Lambda_c < 3/8). \tag{113}$$

This behavior is shown in figure 8 where we have incorporated the drop to zero at g_{AF}.

11. d-Wave gap equation in 2d

In the last section we derived gap equations in the approximation that the gap has no momentum dependence. This is essentially equivalent to assuming the lowest-order scattering of pairs is momentum independent. In appendix A we show how to incorporate momentum-dependent scattering and derive the following form of gap equation:

$$q(\mathbf{k}) = -\int \frac{\mathrm{d}\omega\,\mathrm{d}^d\mathbf{k}'}{(2\pi)^{d+1}} G(\mathbf{k},\mathbf{k}') \frac{q(\mathbf{k}')}{(\omega^2 + \mathbf{k}'^2)^2 + q(\mathbf{k}')^2}, \tag{114}$$

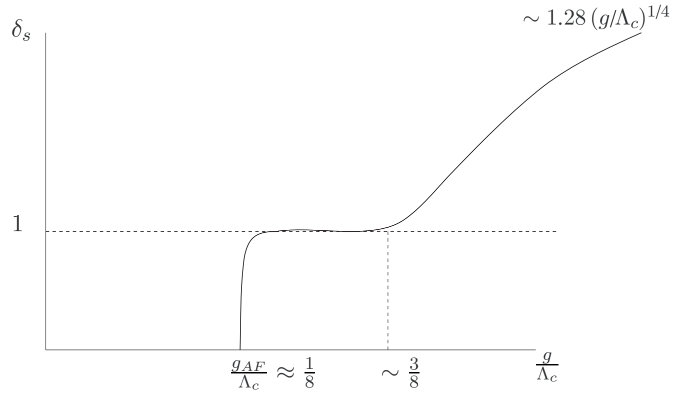


Figure 8. Solutions to the anti-ferromagnetic gap equation.

where the kernel G is given by a Green function related to the scattering of pairs with momenta $\pm\mathbf{k}$ and $\pm\mathbf{k}'$. We have set the mass $m = 0$; it can be restored by $\omega^2 \to \omega^2 + m^2$. (As explained in appendix A, $q(k)$ is not simply the Fourier transform of $q(x)$.)

For our model the kernel will be computed to 1-loop in the following section. In this section we analyze the orbital properties of the gap in $2d$ in a model-independent way based only on the structure of the gap equation and its symmetries. Similar arguments apply to a BCS type of gap equation.

Let us assume that G is symmetric, $G(\mathbf{k}, \mathbf{k}') = G(\mathbf{k}', \mathbf{k})$. If G is also rotationally invariant, then its angular dependence arises only through the dependence on $\mathbf{k} \cdot \mathbf{k}' = kk' \cos(\theta - \theta')$. The kernel and gap can thus be expanded as follows:

$$\begin{aligned} G(\mathbf{k}, \mathbf{k}') &= \sum_{\ell=0}^{\infty} G_\ell(k, k') \cos \ell(\theta - \theta') \\ q(\mathbf{k}) &= \sum_{\ell=0}^{\infty} q_\ell(k) \cos \ell\theta. \end{aligned} \tag{115}$$

Substituting the above expansions into the gap equation one sees that due to the nonlinearity different ℓ can mix. For simplicity consider a single channel, i.e. assume G has only one term in its expansion at fixed ℓ. The angular integral $\int \mathrm{d}\theta'$ can be performed and turns out to be independent of ℓ:

$$\int_0^{2\pi} \mathrm{d}\theta \, \frac{\cos^2 \ell\theta}{1 + a \cos^2 \ell\theta} = \frac{2\pi}{a}(1 - (1 + a)^{-1/2}). \tag{116}$$

The result is

$$q_\ell(k) = -\frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \mathrm{d}\omega \int_0^{\infty} \mathrm{d}k' \, k' G_\ell(k, k') \frac{1}{q_\ell(k')} \left(1 - \frac{\omega^2 + k'^2}{\sqrt{(\omega^2 + k'^2)^2 + q_\ell^2(k')}} \right). \tag{117}$$

It is important to note that although $G(\mathbf{k}, \mathbf{k}')$ varies in sign due to the oscillating cosine, the sign of $G_\ell(k, k')$ is meaningful and determines whether the ℓ channel is attractive or repulsive. Negative G_ℓ corresponds to an attractive channel. Furthermore, any $\ell = 0, 1, 2, \ldots$ is in principle allowed.

The channel $\ell = 2$ can arise rather naturally from a term in the kernel of the form $-2g_2(\mathbf{k}\cdot\mathbf{k}')^2 = -g_2k^2k'^2(1+\cos 2\theta)$ for g_2 a constant, which gives rise to both $\ell = 0, 2$ with the same sign. As we will show in the following section, for our model $\ell = 2$ is the first attractive channel. In particular, G_2 has the form

$$G_2(k, k') = -8\pi^2 g_2 k^2 k'^2 \tag{118}$$

with g_2 a positive constant which we will calculate in the following section. This leads to a solution of the pure d-wave gap equation of the form

$$q(\mathbf{k}) = \delta_q^2 k^2 \cos 2\theta = \delta_q^2\big(k_x^2 - k_y^2\big), \tag{119}$$

where δ_q is a constant satisfying the integral equation:

$$\delta_q^4 = 2g_2 \int_0^{\Lambda_c} \mathrm{d}\omega\,\mathrm{d}k^2 \left(1 - \frac{\omega^2 + k^2}{\sqrt{(\omega^2+k^2)^2 + \delta_q^4 k^4}}\right). \tag{120}$$

The dependence on $\mathbf{k}$ for $\ell = 2$ is of the same form as a particular linear combination of $\ell = 2$ spherical harmonics in $3d$, thus we refer to it as $d_{x^2-y^2}$, or simply d-wave, as in the literature; $\ell = 0, 1$ can be referred to as s- and p-wave.

The above gap equation has some interesting properties, in particular, $\delta_q = 0$ when g_2 is too small. Since this kind of gap equation must be regularized in the UV, we are led to define

$$g_2 = \frac{\widehat{g}_2}{\Lambda_c^3}, \tag{121}$$

where $\widehat{g}_2$ is dimensionless. To estimate the lowest value of $\widehat{g}_2$ with non-zero gap, the integrand in the above equation can be expanded in powers of δ_q. Keeping terms of order δ_q^8, one finds that for δ_q small it behaves as

$$\delta_q \approx \left[\frac{a}{a'}\left(1 - \frac{1}{a\widehat{g}_2}\right)\right]^{1/4}, \tag{122}$$

where

$$a = (8 + \pi - 8\log 2)/12, \qquad a' = 3(104 + 5\pi - 128\log 2)/384. \tag{123}$$

Thus,

$$\delta_q = 0 \qquad \text{for} \quad \widehat{g}_2 < \frac{1}{a} \approx 2.15. \tag{124}$$

For $\widehat{g}_2$ large, one finds

$$\delta_q \approx (2\widehat{g}_2)^{1/4}. \tag{125}$$

However, as we will explain in section 14, there will also be an upper threshold that comes about when the RG is properly implemented, and this leads to an SC dome.

12. 1-Loop scattering and the gap equation kernel

12.1. Feynman diagrams

In this section we derive the kernel for our model and show that it has an attractive d-wave channel. Let $G^{(4)}(p_1, p_2, p_3, p_4)$ denote the 4-particle vertex function with the overall $(2\pi)^D\delta^{(D)}(p)$ removed. (Apart from this overall factor, $G^{(4)}$ differs from $\Gamma^{(4)}$ in appendix A by an overall sign since the Feynman rule vertex is defined to be negative in Euclidean space.) The arrows in the figure indicate how the flavor indices $\alpha, \beta = 1, \ldots, N$ are contracted. They

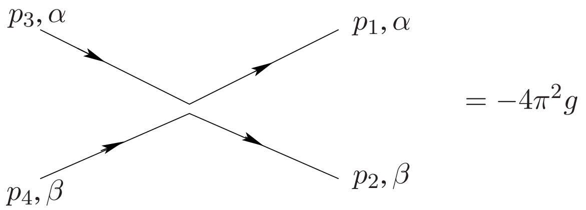


Figure 9. Interaction vertex.

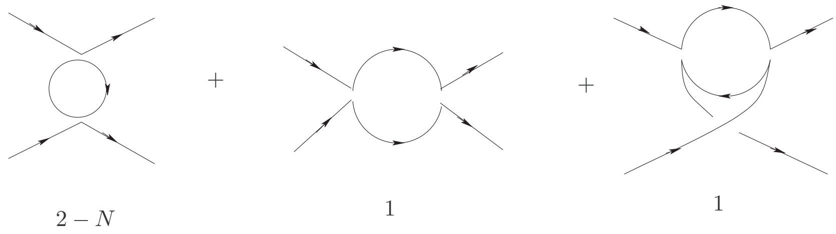


Figure 10. 1-Loop Feynman diagrams with group theory factors.

also indicate the flow of charge since the interaction is proportional to $\left(\sum_\alpha \chi_\alpha^- \chi_\alpha^+\right)^2$. This structure is shown in figure 9, which indicates the interaction of the Euclidean space Feynman rules.

The 1-loop contributions to $G^{(4)}$ are displayed in figure 10 and consist of three separate channels which differ in how the α, β flavor indices are contracted and also in their momentum dependence. It is interesting to carry out this part of the calculation for arbitrary N. The $Sp(2N)$ group theory factors are $2 - N$, 1, 1, respectively, for the three diagrams, where the $-N$ dependence comes about from the closed loop and a fermionic minus sign.

To this order, one thus has

$$G^{(4)}(p_1, p_2, p_3, p_4) = 4\pi^2 g - 2(4\pi^2 g)^2\left[(2 - N) f\left(p_{13}^2\right) + f\left(p_{12}^2\right) + f\left(p_{14}^2\right)\right], \tag{126}$$

where $p_{ij} = p_i + p_j$ and f is the function:

$$\begin{aligned} f(p^2) &= \int \frac{\mathrm{d}^D \ell}{(2\pi)^D} \frac{1}{[\ell^2 + m^2][(\ell + p)^2 + m^2]} \\ &= \int \frac{\mathrm{d}^D \ell}{(2\pi)^D} \int_0^1 \mathrm{d}x \frac{1}{[\ell^2 + x(1 - x)p^2 + m^2]^2}. \end{aligned} \tag{127}$$

Since we are interested in low energies, it is meaningful to expand the integrand in powers of p^2/ℓ^2. At zero temperature, we can also set $m = 0$. One finds

$$f(p^2) = \int \frac{\mathrm{d}^D \ell}{(2\pi)^D} \frac{1}{\ell^4} \left(1 - \frac{1}{3}\frac{p^2}{\ell^2} + \frac{1}{10}\frac{p^4}{\ell^4} + \cdots\right). \tag{128}$$

As explained in appendix A, the gap equation kernel is obtained by specializing to pairs of opposite momenta. We thus fix

$$p_1 = -p_2 = p, \qquad p_3 = -p_4 = p'. \tag{129}$$

There are also two inequivalent ways in which to contract the vertex with external legs which leads us to define

$$G(p, p') = G^{(4)}(p, -p, p', -p') + (p' \to -p'). \tag{130}$$

The momentum-independent part is the following:

$$G(p, p')|_{p,p'=0} = 8\pi^2 g - 64\pi^4 g^2 (4 - N) \left[\frac{d^D\ell}{(2\pi)^D} \frac{1}{\ell^4} \right]. \tag{131}$$

The second term in the above equation can be absorbed into a redefinition of the coupling g. It in fact is just the 1-loop contribution to the RG beta-function and we will return to it in section 13 where our RG prescriptions will be fixed.

The order p^2 term is

$$G(p, p')|_{p^2} = \frac{64\pi^4(3 - N)}{3}(p^2 + p'^2) \left[\frac{d^D\ell}{(2\pi)^D} \frac{1}{\ell^6} \right]. \tag{132}$$

According to the derivation in appendix A, for static gaps, the kernel $G(\mathbf{k}, \mathbf{k}')$ in the gap equation is simply $G(p, p')$ with the time components of p disregarded, i.e. $p = (0, \mathbf{k})$. This implies that though $G(\mathbf{k}, \mathbf{k}')$ is closely related to the S-matrix, it is not identical since the external momenta are not on-shell. The above term is then a repulsive correction to the s-wave contribution, which is already repulsive at tree level.

Finally, we come to the p^4 term:

$$G(\mathbf{k}, \mathbf{k}')|_{k^4} = -\frac{32\pi^4(3 - N)}{5}((\mathbf{k}^2 + \mathbf{k}'^2)^2 + 4(\mathbf{k}\cdot\mathbf{k}')^2) \left[\frac{d^D\ell}{(2\pi)^D} \frac{1}{\ell^8} \right]. \tag{133}$$

Using

$$(\mathbf{k}\cdot\mathbf{k}')^2 = k^2 k'^2 \cos^2(\theta - \theta') = k^2 k'^2 (\cos 2(\theta - \theta') + 1)/2, \tag{134}$$

one sees that this term gives both s- and d-wave contributions. Whether the channel is attractive depends on the sign of the factor $3 - N$. For $N = 2$, the above 1-loop s-wave correction is attractive but since the leading tree-level term is repulsive, it is unlikely that it could lead to s-wave pairing. The leading contribution to the d-wave term is attractive for $N < 3$. This is rather interesting since this d-wave instability is invisible to large N methods. In the following we will make the approximation of the last section and not consider possible mixing of the s- and d-wave gaps.

The d-wave term $G_2(k, k')$ is of the form in equation (118) with

$$g_2 = \frac{(3 - N)8\pi^2 g^2}{5} \int \frac{d^3\ell}{(2\pi)^3} \frac{1}{\ell^8}. \tag{135}$$

The gap thus has the characteristic d-wave form in equation (119) where the constant δ_q is a solution to equation (120).

13. Renormalization group specifically in 2*d*

In the present context it is most appropriate to adopt the Wilsonian effective action view of the RG. In this approach there are two energy scales to consider, a fixed cut-off scale Λ_c and a running scale $\Lambda < \Lambda_c$. Integrating out high-energy modes leads to couplings that depend on the running scale Λ. Wherever possible, one lets the cut-off Λ_c go to infinity.

For illustration, consider first the contribution of loop modes with $\Lambda < \ell < \Lambda_c$ to the momentum-independent part of G in 3*d*:

$$G(p, p') = 8\pi^2 \left(g - (4 - N)g^2 \int_\Lambda^{\Lambda_c} \frac{d\ell}{\ell} \right). \tag{136}$$

The effect of these high-energy modes is to modify the coupling according to $g(\Lambda) = g + g^2(4 - N)\log \Lambda/\Lambda_c$. The beta-function is then $-dg/d\log\Lambda = -(4 - N)g^2$, in

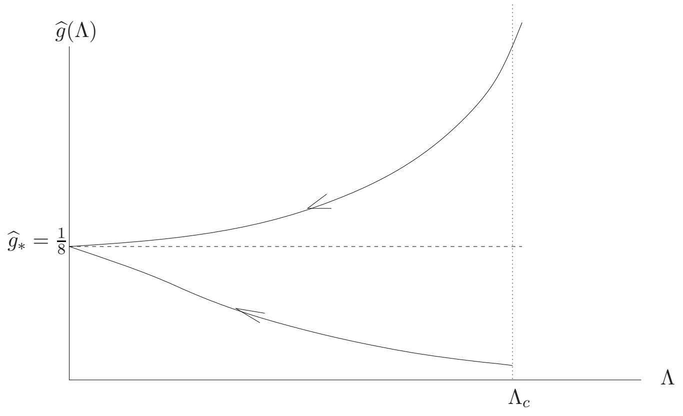


Figure 11. RG flow to $\frac{1}{8}$ at low energies from either strong or weak coupling at short distances.

agreement with results in [17, 18]. An equivalent manner to describe this RG prescription is to let $\int \mathrm{d}\ell \to \int_{\Lambda}^{\Lambda_c} \mathrm{d}\ell$ and then set the cut-off $\Lambda_c \to \infty$ and keep the dependence on the renormalization energy scale Λ.

Let us now specialize this to $2d$. The momentum-independent part of G is now the following:

$$G = 8\pi^2(g - 4(4-N)g^2/\Lambda) \equiv 8\pi^2 g(\Lambda), \tag{137}$$

where we have sent the cut-off Λ_c to ∞. This gives $-\mathrm{d}g/\mathrm{dlog}\,\Lambda = -4(4-N)g^2/\Lambda$, which is consistent with the fact that g has units of energy. The latter is an important feature of our model, and in order to deal with it, we define a dimensionless coupling $\widehat{g}$:

$$g(\Lambda) = \Lambda\widehat{g}(\Lambda). \tag{138}$$

The beta-function for $\widehat{g}$ is now

$$-\Lambda\frac{\mathrm{d}\widehat{g}}{\mathrm{d}\Lambda} = \widehat{g} - 4(4-N)\widehat{g}^2, \tag{139}$$

where the linear term just reflects the classical dimension 1 of g. There is a fixed point at

$$\widehat{g}_* = \frac{1}{4(4-N)}, \tag{140}$$

where the above is only approximate due to corrections beyond 1-loop.

The solution to the RG flow equation is then

$$\widehat{g}(\Lambda) = \frac{\Lambda_c\widehat{g}_0}{\Lambda + 4(4-N)(\Lambda_c - \Lambda)\widehat{g}_0}. \tag{141}$$

The fixed point value $\widehat{g}_*$ is reached irregardless of whether the initial coupling $\widehat{g}_0$ at the cut-off Λ_c is large or small, as long as it is positive. This behavior is sketched in figure 11.

The coupling g_2 that enters the d-wave gap equation can now be expressed as

$$g_2 = \frac{4}{25}\frac{\widehat{g}^2}{\Lambda^3}, \tag{142}$$

where we have safely taken Λ_c to ∞ in equation (135) and set $N = 2$.

14. Scaling and global features of the phase diagram

14.1. RG scaling

For the purposes of comparison with real compounds, our model has only two independent parameters. First is the strength of the coupling $\widehat{g}_0$ at the short distance scale Λ_c. Second is the overall scale Λ_c. The complete phase diagram falls into place when one takes into account the RG scaling. Throughout this section it is implicit that we are only working to 1-loop and all equalities should be taken as approximations.

The structure of the phase diagram is most transparent and the universal features more clearly revealed if one works with the dimensionless inverse couplings:

$$x = \frac{1}{\widehat{g}}, \qquad x_0 = \frac{1}{\widehat{g}_0}, \qquad x_* = \frac{1}{\widehat{g}_*}. \tag{143}$$

Throughout this section we fix $N = 2$ where $x_* = 8$ is the fixed point value. The scale Λ of $g = \Lambda\widehat{g}$ can now be expressed in a simple way in terms of x using equation (141):

$$\frac{\Lambda}{\Lambda_c} = \frac{x_* - x}{x_* - x_0}. \tag{144}$$

The variable x should not be confused with the conventional doping variable in the literature, although it is closely related; the hole doping is the variable h below. The above linear relation is specific to $2d$ and is crucial to understanding the overall structure of the phase diagram. Since the line Λ/Λ_c represents the scale of the coupling, below this line energies are comparable to the coupling and the non-Fermi liquid properties begin to reveal themselves. We have labeled this region the pseudogap as in the literature and will return to it in section 16. Note that when $x = x_0$, $\Lambda/\Lambda_c = 1$.

As we defined them, both the AF and SC gaps δ_s and δ_q are in units of the cut-off Λ_c. Since we have performed an RG transformation of the coupling from Λ_c to Λ, we must also scale the gaps as follows:

$$\delta'_{q,s} = \frac{\Lambda}{\Lambda_c}\delta_{q,s} = \left(\frac{x_* - x}{x_* - x_0}\right)\delta_{q,s}. \tag{145}$$

It will also be convenient to define the parameter γ which measures the distance to the fixed point at short distances:

$$\gamma \equiv \frac{|x_* - x_0|}{x_*}, \tag{146}$$

where $0 < \gamma < 1$. The AF gap equation (109) should now be solved for $\delta'_s(x)$ where g is replaced by

$$\frac{g}{\Lambda_c} = \frac{1}{x}\left(\frac{x_* - x}{x_* - x_0}\right). \tag{147}$$

There are two cases to consider depending on whether the coupling $\widehat{g}$ is strong or weak at short distances, i.e. whether $\widehat{g}_0$ is above or below the fixed point value $\widehat{g}_*$, corresponding to the two curves in figure 11, which we will refer to as Type A and B.

14.2. Type A: strong coupling at short distances

Here we assume $\widehat{g}_0 > \widehat{g}_* = 1/8$, which implies $x_0 = (1 - \gamma)x_*$. The scale as a function of x now has negative slope

$$\frac{\Lambda}{\Lambda_c} = -\frac{1}{\gamma}\left(\frac{x}{x_*} - 1\right) \tag{148}$$

and since the above ratio is by definition less than 1, we have $x_0 < x < x_*$. From equation (111) the first-order transition that terminates the AF phase corresponds to $\widehat{g}_{\mathrm{AF}} = \widehat{g}_* \Lambda_c/\Lambda$ which translates to

$$x_{\mathrm{AF}} = \frac{x_*}{1+\gamma}. \tag{149}$$

The value x_{AF} is always to the right of x_0 since $\gamma^2 > 0$.

Let us now turn to the SC phase. The d-wave gap equation (120) depends on g_2, which by equations (121) and (142) can be expressed as

$$\widehat{g}_2 = \frac{4\gamma^3}{25x^2(1 - x/x_*)^3}. \tag{150}$$

The d-wave gap equation only has solutions if $\widehat{g}_2$ is positive, which requires $x < x_*$. Furthermore, $\widehat{g}_2$ must be above the threshold $\widehat{g}_2 > 1/a$ by equation (124). Thus the d-wave SC gap is non-zero in the range

$$\delta_q \neq 0 \qquad \text{for} \quad x_1 < x < x_*, \tag{151}$$

where the lower threshold is a solution to

$$4a\gamma^3 = 25x_1^2(1 - x_1/x_*)^3. \tag{152}$$

(The constant $a \approx 0.466$ is defined in equation (123).) It should be emphasized that, unlike the values of x_0, x_{AF} and x_* which are determined by the low-energy fixed point, the value x_1 is not universal, i.e. not predicted by the RG itself. Since the solution to the above equation is close to x_*, one can estimate

$$x_1 \approx x_* \left(1 - \gamma \left(\frac{4a}{25x_*^2}\right)^{1/3}\right) \approx (1 - 0.1\gamma)x_*. \tag{153}$$

In principle, it is mathematically possible to have a non-zero solution to the d-wave gap for x sufficiently small, but this will tend to be small and inside the AF phase.

The transition point x_{AF} is outside the SC dome in the approximation we have made, since $x_{\mathrm{AF}} > x_1$ would require $\gamma > 1$. These features, and the geometrical relationships between the various transition points, are shown in figure 12, which is not drawn to scale since the latter depends on γ.

In figure 13 we display numerical solutions to the AF and d-wave SC gap equations for $\gamma = 1$. Rough comparison with experimental results suggests the SC dome as we have calculated it is too narrow, which would imply that we overestimated x_1. However, as stated above, the value of x_1 is less universal than x_0, x_{AF} and x_*, and could easily change by improving our approximations, for example taking into account s- and d-wave mixing, or incorporating other effects we have neglected, such as interplane coupling and disorder. The important point is that there is both an onset and termination of the SC phase, and the termination point on the over-doped side is what is universal since it corresponds precisely to a quantum critical point. Furthermore, since $x_{\mathrm{AF}} < x_1$, AF and d-wave SC do not appear to compete so that the SC is robust.

14.3. Type B: weak coupling at short distances

In this case $\widehat{g}_0 < \widehat{g}_*$ at short distances, and $\gamma = (x_0 - x_*)/x_*$ where in principle $0 < \gamma < \infty$. Now the slope of Λ versus x is positive:

$$\frac{\Lambda}{\Lambda_c} = \frac{1}{\gamma}\left(\frac{x}{x_*} - 1\right) \tag{154}$$

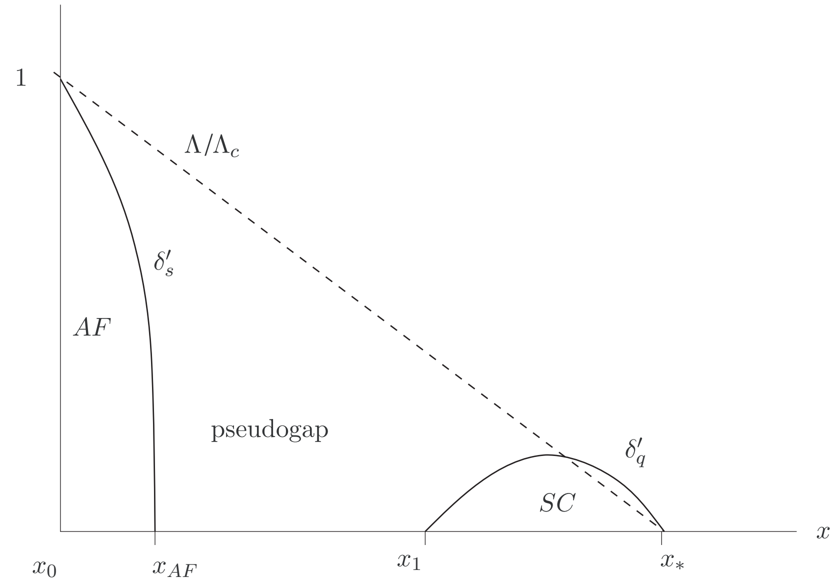


Figure 12. Global phase diagram for Type A. The slope of the dashed line is $-\frac{1}{\gamma x_*}$. The various transition points are related geometrically by $x_0 = (1-\gamma)x_*$, $x_{\rm AF} = x_*/(1+\gamma)$ and $x_1 \approx (1-0.1\gamma)x_*$, where $x_* = 8$. The dashed line represents equation (144).

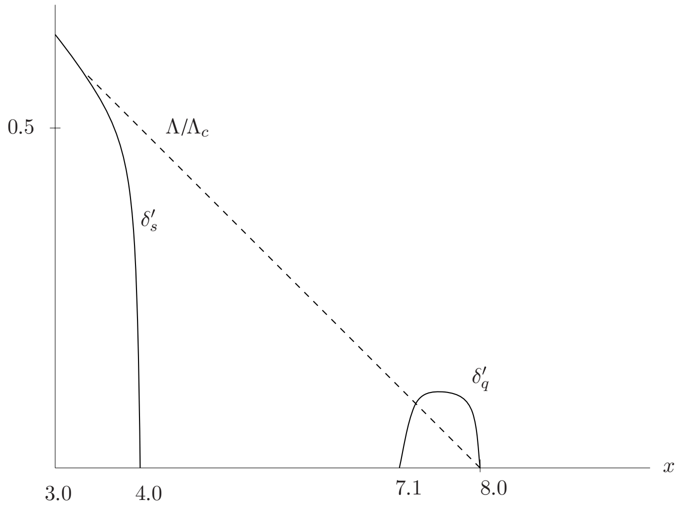


Figure 13. Numerical solutions of the AF and d-wave SC gap equations for $\gamma = 1$ (Type A).

and $x_* < x < x_0$. The AF transition point is at

$$x_{\rm AF} = \frac{x_*}{1-\gamma}. \tag{155}$$

Thus, $0 < \gamma < 1$ otherwise $x_{\rm AF}$ goes from ∞ to $-\infty$.

There are non-zero solutions to the d-wave gap equation in the range:

$$\delta_q \neq 0 \qquad \text{for} \quad x_* < x < x_2, \tag{156}$$

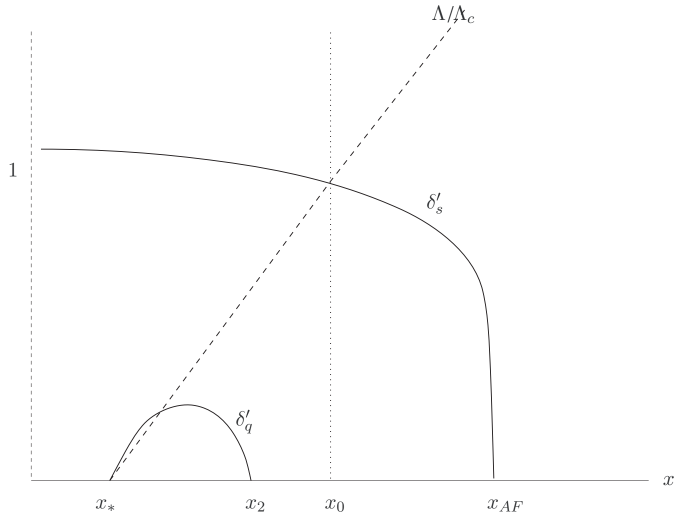


Figure 14. Global phase diagram for Type B. The values shown are related as follows: $x_0 = (1+\gamma)x_*$, $x_{\mathrm{AF}} = x_*/(1-\gamma)$ and $x_2 \approx (1+0.1\gamma)x_*$, where $x_* = 8$. The dashed line has slope $1/(\gamma x_*)$.

where x_2 is a solution to equation (152) with $x_1 \to x_2$ and $\gamma \to -\gamma$. An approximation of the form (153) also holds with $\gamma \to -\gamma$. With the range of parameters of our model, x_{AF} lies outside the SC dome, so that the SC dome is completely inside the AF phase. These features are shown in figure 14.

14.4. Doping: 1-loop corrections and optimal doping

The hole-doping variable h defined in section 10 takes the simple form in terms of the x variables:

$$h(x) = \frac{1}{\pi^2}\left(\frac{x - x_0}{x_* - x_0}\right). \tag{157}$$

Again, this linear dependence on x is characteristic to $2d$. As argued in section 10, since increasing Λ/Λ_c decreases the density to below half-filling, h is a measure of doping, where $x = x_0$ should correspond to half-filling. Thus apart from an overall scale of $1/\pi^2$ and a shift of the origin, our previous phase diagrams in terms of x are effectively in terms of the doping h.

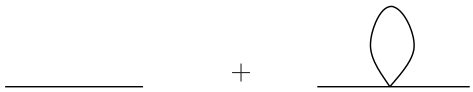
Since both the AF and SC termination points x_{AF} and x_* are closely related to the RG fixed point, the ratio of doping at these two values should be universal. Let $h_{\mathrm{AF}} = h(x_{\mathrm{AF}})$ and $h_* = h(x_*)$. Expressed in terms of γ one has

$$h_{\mathrm{AF}} = \frac{1}{\pi^2}\frac{\gamma}{1+\gamma}, \qquad h_* = \frac{1}{\pi^2}, \tag{158}$$

which gives the ratio $h_{\mathrm{AF}}/h_* = \gamma/(1+\gamma)$.

The above formula for h is from equation (105) where only the zeroth-order contribution was calculated. We can easily include the first-order correction in g, which is the 1-loop self-energy diagram shown in figure 15. This leads to the following correction:

$$h = \frac{2}{\Lambda_c}\int_{\Lambda}^{\Lambda_c} \frac{\mathrm{d}^3 p}{(2\pi)^3}\left(\frac{1}{p^2} + \frac{4g(\Lambda_c - \Lambda)}{p^4}\right). \tag{159}$$



$$= \frac{1}{p^2} + \frac{8\pi^2 g}{p^4} \int_{\Lambda}^{\Lambda_c} \frac{d^3\ell}{(2\pi)^3} \frac{1}{\ell^2}$$

Figure 15. 1-Loop correction to the propagator.

There is an important fermionic minus sign in the above expression coming from the fermion loop. Expressing the result of doing the above integral in terms of x using equation (144) one obtains

$$h(x) = \frac{1}{\pi^2}\left(\frac{x - x_0}{x_* - x_0}\right)\left[1 + \frac{4}{x}\left(\frac{x - x_0}{x_* - x_0}\right)\right]. \tag{160}$$

The corrections to formula (158) are then

$$h_{\mathrm{AF}} = \frac{1}{\pi^2}\frac{\gamma(2+\gamma)}{2(1+\gamma)}, \qquad h_* = \frac{3}{2\pi^2}. \tag{161}$$

The optimal doping fraction can be estimated as follows. Recall that for Type A, the SC phase occurs for $x_1 < x < x_*$. Evaluating h at the lower limit x_1 one finds a weak dependence on γ:

$$h(x_1) \approx \frac{1}{\pi^2}\left(\frac{1.3 - 0.09\gamma}{1 - 0.1\gamma}\right). \tag{162}$$

One thus concludes that optimal doping occurs in the tight range:

$$\frac{1.3}{\pi^2} = 0.13 < h_{\mathrm{optimal}} < 0.15 = \frac{3}{2\pi^2}. \tag{163}$$

This leads to

$$\frac{h_{\mathrm{AF}}}{h_{\mathrm{optimal}}} \approx \frac{\gamma(2+\gamma)}{3(1+\gamma)}. \tag{164}$$

For $\gamma = 1/2$ this gives $h_{\mathrm{AF}} \approx 0.04$.

15. Details on the AF and d-wave SC gaps and critical temperatures

In this section we provide detailed numerical solutions to the gap equations for a variety of γ and also re-introduce the mass to incorporate a temperature.

15.1. AF gap

Numerical solutions to the AF gap equation (109) expressed in terms of x for $\gamma = 1/2$ are shown in figure 16.

To study finite temperature, we introduce a mass $m = \alpha T$ where $\alpha = \pi^{5/4}/\sqrt{6} \approx 1.7$ was introduced in section 7. The gap equation becomes

$$\frac{\Lambda_c}{g} = \frac{4}{\delta_s^2}\left(\sqrt{\delta_s^2 - \widehat{m}^2}\tanh^{-1}\frac{1}{\sqrt{\delta_s^2 - \widehat{m}^2}} - \sqrt{\delta_s^2 + \widehat{m}^2}\tan^{-1}\frac{1}{\sqrt{\delta_s^2 + \widehat{m}^2}}\right), \tag{165}$$

where $\widehat{m} = m/\Lambda_c$. Since we have RG scaled the gaps, in order to make comparisons on the same scale we need to also define a scaled mass:

$$m' = \frac{\Lambda}{\Lambda_c} m. \tag{166}$$

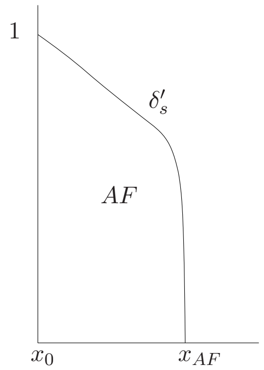


Figure 16. Numerical solutions to the AF gap for $\gamma = 1/2$, $x_0 = 4$ and $x_{\rm AF} = 5.3$.

Numerical study of equation (165) shows that the solution δ_s' vanishes when m' is too large, consistent with identification of m with temperature, as in conventional superconductivity discussed in section 9. Let m_N' denote the value of m' where the solution δ_s' vanishes. This leads us to define a Néel temperature T_N where the gap vanishes. As in the BCS theory, T_N is expected to be proportional to the zero-temperature gap. Let us define then

$$T_N = \frac{m_N'}{\alpha} = c_{\rm AF}(\gamma, x)\frac{\delta_s'\Lambda_c}{\alpha}, \tag{167}$$

where

$$c_{\rm AF}(\gamma, x) \equiv \frac{m_N'(\gamma, x)}{\Lambda_c \delta_s'} \tag{168}$$

and in these equations δ_s' is the zero-temperature gap. Inspection of the finite-temperature gap equation (165) shows that the solution should disappear when the argument of the square-root is negative, i.e. when $\widehat{m} \approx \delta_s$, and $c_{\rm AF} \approx 1$. This is a delicate limit, however we have verified that this is approximately correct numerically.

15.2. SC gap

Numerical solutions to the d-wave gap equation (120) expressed in terms of x are shown in figure 17 for $\gamma = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$. One sees that the peak value of the gap $\delta_q' \approx 0.11$ is not very sensitive to γ.

Let us now introduce a temperature T by letting $\omega^2 \to \omega^2 + m^2$ in the gap equation (120) where as before $m = \alpha T$. The behavior of the gap as a function of T is shown in figure 18 for $\gamma = \frac{1}{4}, \frac{1}{2}, 1$. Let m_c' denote the value of m where the gap vanishes. One sees from the figure that here there is a somewhat stronger dependence of m_c on the inverse coupling x than for the AF gap.

The critical temperature T_c can now be defined as

$$k_B T_c = \frac{m_c'}{\alpha} = \frac{c_{\rm SC}(\gamma, x)}{\alpha} v_F \hbar \delta_q' \Lambda_c, \tag{169}$$

where

$$c_{\rm SC}(\gamma, x) = \frac{m_c'}{\Lambda_c \delta_q'} \tag{170}$$

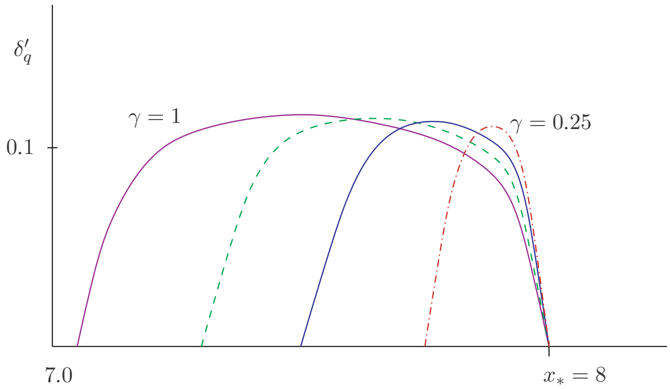


Figure 17. Numerical solutions of the d-wave gap equation for $\gamma = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$.

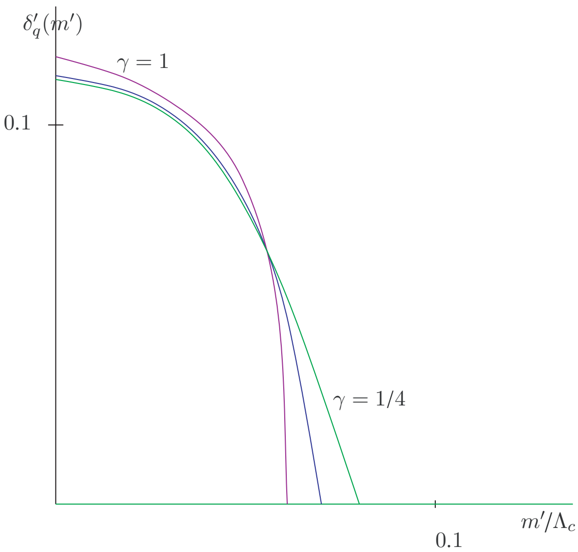


Figure 18. d-Wave gap as a function of m for $\gamma = 1/4, 1/2, 1$.

is a dimensionless constant. In this final formula, we have restored fundamental constants and the Fermi velocity v_F. Above, δ'_q is the zero-temperature gap, which depends on γ, x as does m'_c.

Let $x_{\rm opt}$ denote the value of x with the largest gap. The constant $c_{\rm SC}$ for various γ at the corresponding $x_{\rm opt}$ are shown in table 2. One sees that $c_{\rm SC}$ is of order 1 and has only a weak dependence on the couplings.

We can estimate now the critical temperature at optimal doping. The above table shows that $c_{\rm SC}$ depends relatively weakly on γ. Furthermore, δ'_q at optimal doping is also relatively constant $\approx$0.11. Thus the scale of the critical temperature (169) is set primarily by v_F and Λ_c, i.e. it is only weakly dependent on the coupling g or the Hubbard couplings by equation (106); this is the beauty of having a low-energy fixed point. The scale $\Lambda_c = 1/a$ is an inverse length a, which should be on the order of the lattice spacing.

Table 2. Critical mass versus zero temperature gap for various γ.

γ	x_{opt}	$\delta'_q(x_{\text{opt}})$	m'_c/Λ_c	$c_{\text{SC}} = m'_c/\delta'_q\Lambda_c$
1	7.5	0.118	0.06	0.51
3/4	7.6	0.116	0.06	0.52
1/2	7.8	0.114	0.07	0.61
1/4	7.9	0.113	0.08	0.71
1/8	7.94	0.113	0.06	0.53
1/16	7.97	0.113	0.06	0.53

To give a rough estimate of T_c, let us take the lattice spacing to be that of the CuO_2 square lattice $a = 3.8$Å. For v_F we use the universal nodal Fermi velocity from [20], which we estimated to be $v_F \approx 210\,\text{km s}^{-1}$ in section 3 for LSCO. The most unknown quantity is α that sets the relation between temperature and mass; let us use the estimate of $\alpha = \pi^{5/4}/\sqrt{6}$ from section 7. With the average values $\delta'_q = 0.11, c_{\text{SC}} = 0.6$, from the above table, this gives $T_c \approx 140\,\text{K}$, which is quite reasonable considering all of the order 1 constants we have approximated. Note that this result relies on the relatively small value of δ'_q we found from numerical solutions of the d-wave gap equation; a value δ'_q of order 1 would give T_c higher by an order of magnitude.

A useful form of the above equation for T_c is

$$T_c = c_{\text{SC}} \frac{v_F}{a} \cdot 650\,\text{K}, \tag{171}$$

where we have set α to its estimated value $\alpha = 1.7$, a is the lattice spacing in angstroms, and v_F is in ev-angstroms. For the range of c_{SC} shown in the above table, $120\,\text{K} < T_c < 160\,\text{K}$. The maximum T_c occurs around $\gamma = 1/4$. Since this T_c is on the high side, this is perhaps because we underestimated α in our simplified inclusion of temperature. Another possibility is that Λ_c should instead be set by the average separation of holes. At doping $h = 0.15$ this increases a by a factor of about 2.6 leading to $46\,\text{K} < T_c < 62\,\text{K}$.

The above formula gives some hints on how to increase T_c: shorten the lattice spacing, increase the Fermi velocity by somehow modifying the effective electron mass m_*, or tuning the material to $\gamma = 1/4$, which would require screening the Coulomb potential at short distances. In particular, T_c should increase with pressure if the main effect of higher pressure is to reduce the lattice spacing.

16. The pseudogap region

It is often suggested in the literature that the pseudogap is the key to understanding high T_c since the AF and SC order condense out of it, and there has been much speculation about its nature. (For a review of the experimental data see [33].) In this section we briefly discuss the insights on the pseudogap furnished by our model thus far.

As stressed throughout this paper, the essential ingredient is the non-Fermi liquid at higher temperatures before condensation. The pseudogap is the region where the energy scales are such that this non-Fermi liquid behavior is visible, i.e. there are strong correlations of the electrons described by our fields χ, but the energy is not low enough for them to condense into an AF or SC ordered state. The relevant energy scale is Λ of the coupling $g = \Lambda\widehat{g}$, thus it is natural to define a pseudogap temperature T_{pg}:

$$T_{\text{pg}} = \Lambda = xg(\Lambda) = \frac{1}{\gamma}\left(1 - \frac{x}{x_*}\right)\Lambda_c, \tag{172}$$

where we have used equation (147). Thus, as defined, $T_{\rm pg}$ is proportional to the coupling g and below this scale the order g corrections are relatively large. $T_{\rm pg}$ corresponds to the dashed line in the figures of section 14, and this line is seen in measurements of various properties on the under-doped side. Our model predicts that the $T_{\rm pg}$ line runs between the top of the AF gap to precisely the quantum critical point which terminates the SC phase on the over-doped side, and the experiments support this [33].

Since the dashed line in the figures simply represents the relevant energy scale where the interactions become strong, it does not represent a phase transition. There are thus no spontaneously broken $SU(2)$ or $U(1)$ symmetries just below $T_{\rm pg}$. The field with non-zero expectation value is the $SU(2) \otimes U(1)$ invariant $\langle \chi^- \chi^+ \rangle$ which by equation (105) is related to Λ/Λ_c. The spectrum still consists of particles and holes with charge $\pm e$, just strongly coupled. Any experiment performed at energies comparable to the scale of the coupling g is sensitive to the quantum corrections in our theory, thus there are a wide variety of probes of $T_{\rm pg}$.

A possible concrete manifestation of the above pseudogap scale is a dynamically generated mass m at zero temperature, since such a mass is necessarily proportional to Λ. At finite temperature this mass will typically lead to thermal suppression factors $\mathrm{e}^{-m/T}$ at low temperatures. This idea will be investigated in forthcoming work.

17. Concluding remarks

To summarize, we have constructed a $2d$ relativistic model with 4-fermion interactions and analyzed in detail many of its properties which revealed a striking resemblance with the known features of high-T_c superconductivity in the cuprates. Apart from the overall scale Λ_c, the phase diagram can be calculated as a function of a single parameter γ, and gives a very good first draft of what is observed in experiments. Since these features of the model were outlined in the introduction, we conclude by summarizing the main theoretical aspects that are new and responsible for the model's properties.

- In the approach we have taken for expanding around a circular Fermi surface, there is essentially a unique non-Fermi liquid in $2d$ for spin-$\frac{1}{2}$ electrons. The requirements of a local, rotationally invariant quantum field theory in a sense make this theory inevitable. The rotational invariance is not in conflict with the existence of a lattice if the wavelengths at low energies are long compared with the lattice spacing, in fact it is already known that the nonlinear sigma model description of the Heisenberg magnet is rotationally invariant.
- Our model has a low-energy fixed point for the same reason that the $O(M)$ vector models have the Wilson–Fisher fixed point, since the perturbative expansion differs from the latter only by some fermionic minus signs. The model is perturbatively tractable since the low-energy value of the coupling is relatively small $\approx 1/8$, even for arbitrarily strong interactions at short distances.
- The kind of theory considered here, namely a Lorentz scalar fermion with interactions, has only recently been considered in this context [17, 18] because it was previously thought to be inconsistent with the spin-statistics theorem and unitarity. In fact, the free version arises as Faddeev–Popov ghosts in gauge theory. As explained in [18], since the theory turns out to be pseudo-Hermitian, $H^\dagger = CHC$, it still defines a unitary time evolution and has real eigenvalues. The non-interacting theory is actually a perfectly Hermitian description of particles and holes near the Fermi surface. In this paper we further elaborated on this issue, noting that the operator C simply distinguishes between particles and holes. Furthermore, the kinematic constraints coming from the expansion

around the Fermi surface require eigenstates of H that are also eigenstates of C, and for such states H is actually Hermitian. Finally, the model is consistent with spin-statistics since spin is a flavor symmetry and we are still quantizing spin-$\frac{1}{2}$ particles as fermions.

- A small non-zero temperature can be introduced as a relativistic mass term in the Lagrangian. Thus, in Euclidean space, the theory is three-dimensional and spontaneous symmetry breaking is possible, in accordance with the Mermin–Wagner theorem.
- Our model gives a new quantum field theoretic treatment of conventional s-wave superconductivity for attractive interactions in $3d$ and can be extended down to $2d$.
- The same model can be motivated from the Hubbard model at half-filling. Thus, the model can interpolate between the nearly circular Fermi surface just below and up to half-filling. This analysis shows how to vary the hole doping by varying the cut-offs, or equivalently the inverse coupling, and leads to a calculable phase diagram. The phase diagram has some universal geometrical features that depend on the strength of the coupling at short distances. In order to properly understand the phase diagram, it was essential to recognize that hole doping was proportional to the inverse coupling and to implement certain RG scaling relations.
- It was necessary to derive a new kind of momentum-dependent gap equation that incorporates the scattering of Cooper pairs near the Fermi surface. Only then can one see that there is an attractive d-wave channel even if the original interactions that led to AF order were repulsive. This d-wave channel is only attractive for $N < 3$, so it cannot be understood using large N methods.
- Although the $SO(5)$ symmetry helps to explain the existence of both AF and SC order since order parameters for both are present, since the AF gap is s-wave and the SC gap is d-wave, they are not simply related by symmetry. In particular, the d-wave gap equation is second order in the coupling g whereas the AF one is first order. What is related by $SO(5)$ symmetry to the AF phase is a conventional s-wave SC obtained when one flips the sign of the interaction to make it attractive.
- The pseudogap is the region where the energy scales are comparable to the scale of the dimensionfull coupling g, but not low enough for the particles to condense.

Our theory certainly reproduces all of the qualitative features of high-T_c materials and we believe also explains the universal nodal Fermi velocity observed in [20]. On the quantitative side, the 1-loop calculation of the optimal doping fraction is in good agreement with experimental results, as is our estimate of T_c. It is beyond the scope of this paper to present additional detailed comparisons with the many impressive experimental results obtained over the past two decades.

Acknowledgments

The authors wish to thank Seamus Davis, Jim Sethna, Kyle Shen and Henry Tye for discussions. This work is supported in part by the National Science Foundation.

Appendix A. Derivation of the momentum-dependent gap equation

In order to consistently incorporate higher order scattering, one can start with the so-called 1PI effective action S_{eff} with the usual definition as the sum of 1-particle irreducible vertices [5]. Throughout this section we work in Euclidean space. Since we are only interested in gaps that depend on the spatial component $\mathbf{k}$ of $p = (\omega, \mathbf{k})$, one should start with a 4-particle term of the form

$$S_{\mathrm{eff}}|_{4\text{-particle}} = \int \mathrm{d}t \int \mathrm{d}^d\mathbf{x}_1 \cdot s\, \mathrm{d}^d\mathbf{x}_4\, \Gamma^{(4)}(x_1, x_2, x_3, x_4)\chi_\uparrow^-(x_1)\chi_\downarrow^-(x_2)\chi_\downarrow^+(x_3)\chi_\uparrow^+(x_4), \qquad \text{(A.1)}$$

where all x_i are at the same time: $x_i = (t, \mathbf{x}_i)$. However, in order to streamline the derivation, and also to more clearly express the result in terms of the usual momentum-space Green functions, we treat space and time on equal footing and make the above reduction at the end. The 4-particle term in our $N = 2$ model then has the form

$$\begin{aligned} S_{\mathrm{eff}}|_{4\text{-particle}} &= \int \mathrm{d}^D x_1 \cdots \mathrm{d}^D x_4\, \Gamma^{(4)}(x_1, x_2, x_3, x_4)\chi_\uparrow^-(x_1)\chi_\downarrow^-(x_2)\chi_\downarrow^+(x_3)\chi_\uparrow^+(x_4) \\ &= \int (\mathrm{d}p_1)\cdots(\mathrm{d}p_4)\Gamma^{(4)}(p_1, p_2, p_3, p_4)\chi_\uparrow^-(p_1)\chi_\downarrow^-(p_2)\chi_\downarrow^+(p_3)\chi_\uparrow^+(p_4), \end{aligned}$$

where $(dp) \equiv d^D p/(2\pi)^D$ and $\chi(p)$ is the Fourier transform of $\chi(x)$. The function $\Gamma^{(4)}$ is a 4-point correlation function of χ's and by translational invariance has an overall δ-function

$$\Gamma^{(4)}(p_1, p_2, p_3, p_4) = (2\pi)^D\delta^{(D)}(p_1 + p_2 + p_3 + p_4)(-8\pi^2 g + \cdots), \qquad \text{(A.2)}$$

where we have included the tree-level contribution and '$\cdots$' represents loop corrections.

To derive the gap equation we follow the auxiliary field method of section 8 and introduce momentum dependence in the manner described by Weinberg [4, 5]. Introduce pair fields $q^\pm(p_1, p_2)$ and the auxiliary action

$$\begin{aligned} S_{\mathrm{aux}} = \int (\mathrm{d}p_1)\cdots(\mathrm{d}p_4)\Gamma^{(4)}(p_1, p_2, p_3, p_4)\big[&-q^+(p_1, p_2)q^-(p_3, p_4) \\ &+ q^-(p_1, p_2)\chi_\uparrow^+(p_3)\chi_\downarrow^+(p_4) + q^+(p_1, p_2)\chi_\downarrow^-(p_3)\chi_\uparrow^-(p_4)\big]. \end{aligned} \qquad \text{(A.3)}$$

The equations of motion for q give $q^- = \chi_\downarrow^-\chi_\uparrow^-$ and $q^+ = \chi_\uparrow^+\chi_\downarrow^+$ and substituting back into S_{aux} recovers the correct quartic interaction of χ's.

We now specialize to Cooper pairs with total momentum zero:

$$q(p_1, p_2) = (2\pi)^D\delta^{(D)}(p_1 + p_2)q(p_1). \qquad \text{(A.4)}$$

Define the kernel G as follows:

$$\Gamma^{(4)}(p_1, -p_1, p_3, p_4) = (2\pi)^D\delta^{(D)}(p_3 + p_4)G(p_1, p_3). \qquad \text{(A.5)}$$

Then the auxiliary action becomes

$$\begin{aligned} S_{\mathrm{aux}} = \int (\mathrm{d}p)(\mathrm{d}p')G(p, p')\big[&-V^{(D)}q^+(p)q^-(p') + q^-(p)\chi_\uparrow^+(p')\chi_\downarrow^+(-p') \\ &+ q^+(p)\chi_\downarrow^-(p')\chi_\uparrow^-(-p')\big], \end{aligned} \qquad \text{(A.6)}$$

where $V^{(D)} = (2\pi)^D\delta^{(D)}(0)$ is the D-dimensional volume.

The free-field kinetic term is the following:

$$S_{\mathrm{free}} = \int (\mathrm{d}p) \sum_{\alpha=\uparrow,\downarrow} \chi_\alpha^-(p)(p^2 + m^2)\chi_\alpha^+(-p). \qquad \text{(A.7)}$$

The Gaussian functional integral over the χ fields can now be performed to give an effective potential $V_{\mathrm{eff}} = S_{\mathrm{eff}}/V^{(D)}$. To describe V_{eff} in a compact form it is convenient to define

$$\widehat{q}^\pm(p) = \int (\mathrm{d}p')G(p, p')q^\pm(p'). \qquad \text{(A.8)}$$

One then finds

$$V_{\mathrm{eff}} = -\int (\mathrm{d}p)(\mathrm{d}p')G^{-1}(p, p')\widehat{q}^+(p')\widehat{q}^-(p) - \frac{1}{2}\int (\mathrm{d}p)\, \mathrm{Tr}\, \log A(p), \qquad \text{(A.9)}$$

where $A(p)$ is of the same form as in equation (91) with $q^{\pm} \to -\widehat{q}^{\pm}(p)$ and $s = 0$. Above, G^{-1} is the inverse of G as an integral operator, and not simply $1/G$.

In order to understand how one recovers the constant gap equation, let us pass to position space where

$$q(x_1, x_2) = \int (\mathrm{d}p)\, \mathrm{e}^{\mathrm{i}p\cdot(x_1-x_2)} q(p). \tag{A.10}$$

For a constant kernel, it can be redefined to $G(p, p') = 1$, and one sees that

$$q(x, x) = \int (\mathrm{d}p) q(p) = \widehat{q} \tag{A.11}$$

thus a constant $\widehat{q}(p)$ corresponds to a constant $q = q(x, x)$ in position space. Thus, where there is no time dependence in the gap, one can let $p = (\omega, \mathbf{k})$ and define

$$\widehat{q}(\mathbf{k}) = \int (\mathrm{d}\mathbf{k}') G(\mathbf{k}, \mathbf{k}') q(\mathbf{k}'), \tag{A.12}$$

where now $G(\mathbf{k}, \mathbf{k}')$ is the same kernel but with the time Fourier transform disregarded. In other words, $G(\mathbf{k}, \mathbf{k}')$ is simply $G(p, p')$ with $p = (0, \mathbf{k})$ and $p' = (0, \mathbf{k}')$.

Finally, $\delta_q V_{\text{eff}} = 0$ gives the gap equation

$$\widehat{q}(\mathbf{k}) = -\int \frac{\mathrm{d}\omega\, \mathrm{d}^d\mathbf{k}'}{(2\pi)^{d+1}} G(\mathbf{k}, \mathbf{k}') \frac{\widehat{q}(\mathbf{k}')}{(\omega^2 + \mathbf{k}'^2)^2 + \widehat{q}(\mathbf{k}')^2}, \tag{A.13}$$

where $\widehat{q} = \widehat{q}^{\pm}$. After relabeling $\widehat{q} \to q$ this is equation (114).

Appendix B. Relevant aspects of lattice fermions

In this appendix we collect some known features of lattice fermion models that are referred to in the paper. All are contained in the reviews [34, 35].

Let $\mathbf{r}_i$ denote the positions of sites of a two-dimensional square lattice with lattice spacing a. Nearest neighbors $\mathbf{r}_i$ and $\mathbf{r}_j$ are related by $\mathbf{r}_j = \mathbf{r}_i + \mathbf{a}$ with $\mathbf{a} \in \{\mathbf{a}_1, \ldots, \mathbf{a}_4\}$, where $\mathbf{a}_1 = -\mathbf{a}_3 = (a, 0), \mathbf{a}_2 = -\mathbf{a}_4 = (0, a)$. Introduce fermion operators $c_{\mathbf{r}_i,\alpha}$ where $\alpha = \uparrow, \downarrow$ represents spin and define the Hamiltonian

$$H = -t \sum_{\mathbf{r},\mathbf{a},\alpha} \left(c^{\dagger}_{\mathbf{r},\alpha} c_{\mathbf{r}+\mathbf{a},\alpha} \right), \tag{B.1}$$

where

$$\left\{ c_{\mathbf{r},\alpha}, c^{\dagger}_{\mathbf{r}',\alpha'} \right\} = \delta_{\mathbf{r},\mathbf{r}'} \delta_{\alpha,\alpha'}. \tag{B.2}$$

The Hamiltonian is Hermitian due to $\sum_{\mathbf{a}} = \sum_{-\mathbf{a}}$.

Introduce the momentum space expansion

$$c_{\mathbf{r},\alpha} = \sum_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} c_{\mathbf{k},\alpha}. \tag{B.3}$$

Using $\sum_{\mathbf{r}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} = \delta_{\mathbf{k},0}$, the Hamiltonian becomes

$$H = \sum_{\mathbf{k},\alpha} \varepsilon_{\mathbf{k}} c^{\dagger}_{\mathbf{k},\alpha} c_{\mathbf{k},\alpha}, \tag{B.4}$$

where

$$\varepsilon_{\mathbf{k}} = -2t(\cos k_x a + \cos k_y a). \tag{B.5}$$

Define the local spin operators:

$$S^{+}_{\mathbf{r}} = \frac{1}{\sqrt{2}} c^{\dagger}_{\mathbf{r}\uparrow} c_{\mathbf{r}\downarrow}, \qquad S^{-}_{\mathbf{r}} = \frac{1}{\sqrt{2}} c^{\dagger}_{\mathbf{r}\downarrow} c_{\mathbf{r}\uparrow}, \qquad S^{z}_{\mathbf{r}} = \frac{1}{2}\left(c^{\dagger}_{\mathbf{r}\uparrow} c_{\mathbf{r}\uparrow} - c^{\dagger}_{\mathbf{r}\downarrow} c_{\mathbf{r}\downarrow} \right). \tag{B.6}$$

The conserved spin $SU(2)$ charges are $\vec{Q} = \sum_{\mathbf{r}} \vec{S}_{\mathbf{r}}$ and satisfy the $SU(2)$ algebra with the convention in equation (58).

The Hubbard interaction is essentially unique up to shifts of the chemical potential as a consequence of the Fermi statistics:

$$H_{\text{int}} = U \sum_{\mathbf{r}} \left(n_{\mathbf{r}\uparrow} - \frac{1}{2}\right)\left(n_{\mathbf{r}\downarrow} - \frac{1}{2}\right), \tag{B.7}$$

where $n_{\mathbf{r}\alpha} = c^\dagger_{\mathbf{r}\alpha} c_{\mathbf{r}\alpha}$ are local number operators. The interaction can also be written as a spin–spin interaction since

$$\vec{S}_{\mathbf{r}} \cdot \vec{S}_{\mathbf{r}} = -\tfrac{3}{2} n_{\mathbf{r}\uparrow} n_{\mathbf{r}\downarrow} + \tfrac{3}{4}(n_{\mathbf{r}\uparrow} + n_{\mathbf{r}\downarrow}). \tag{B.8}$$

In addition to the above spin $SU(2)$ symmetry, there is another commuting $SU(2)$ symmetry so that the largest symmetry of the Hubbard model is $SU(2) \otimes SU(2) = SO(4)$ [36]. The latter $SU(2)$ is intrinsic to the lattice and does not obviously have a continuum limit.

At half-filling there is one fermion per site, which implies $n_{\mathbf{r}\uparrow} n_{\mathbf{r}\downarrow} = 0$ and $n_{\mathbf{r}\uparrow} + n_{\mathbf{r}\downarrow} = 1$. Thus at half-filling one sees from equation (B.8)

$$\vec{S}_{\mathbf{r}} \cdot \vec{S}_{\mathbf{r}} = \tfrac{3}{4}. \tag{B.9}$$

Equating the above with $j(j+1)$, $j = 1/2$ one sees that at half-filling the local spin operators $\vec{S}$ form the two-dimensional spin-$\frac{1}{2}$ representation of $SU(2)$. Away from half-filling the constraint $(\vec{S}_{\mathbf{r}})^2 =$ constant needs to be relaxed.

At large U and half-filling, the Hubbard model can be formulated as an effective spin-$\frac{1}{2}$ Heisenberg model [7, 37]:

$$H_{\text{eff}} = J \sum_{\langle i,j \rangle} (\vec{S}_{\mathbf{r}_i} \cdot \vec{S}_{\mathbf{r}_j} - 1/4), \tag{B.10}$$

where $J = 4t^2/U$.

In the continuum limit an effective theory for fluctuations above the AF state is the nonlinear $O(3)$ sigma model [32]. One starts from a staggered configuration

$$\vec{S}_{\mathbf{r}} = \pm \vec{n}_{\mathbf{r}} + \vec{\ell}_{\mathbf{r}}, \tag{B.11}$$

where $\pm$ is for even/odd sublattices. In the continuum limit the effective theory for the $\vec{n}$ field is second order in space and time derivatives with Lagrangian density

$$\mathcal{L} = \tfrac{1}{2}(\partial_t \vec{n} \cdot \partial_t \vec{n} - \vec{\nabla}\vec{n} \cdot \vec{\nabla}\vec{n}) + \mathcal{L}_{\text{top}}, \tag{B.12}$$

where $\vec{n}^2$ is constrained due to equation (B.9)

$$\vec{n}^2 = \text{constant}. \tag{B.13}$$

The topological term has the form $\mathcal{L}_{\text{top}} \sim \vec{n} \cdot \partial \vec{n} \times \partial \vec{n}$. In $1d$ the topological term is known to serve an important role in determining the low-energy fixed point [32]. In $2d$ the topological term is not known to play an analogous role; in fact it is an RG irrelevant operator of dimension $7/2$ and we ignore it.

References

[1] Benfatto G and Gallavotti G 1990 *J. Stat. Phys.* **59** 541
Benfatto G and Gallavotti G 1990 *Phys. Rev.* B **42** 9967

[2] Shankar R 1991 *Physica* A **177** 530
Shankar R 1994 *Rev. Mod. Phys.* **66** 129 (arXiv:cond-mat/9307009)

[3] Polchinski J 1992 *Effective Field Theory and the Fermi Surface* (*TASI 1992 lecture*) (arXiv:hep-th/9210046)

[4] Weinberg S 1994 *Nucl. Phys.* B **413** 567

[5] Weinberg S 1996 *The Quantum Theory of Fields* vols 1 and 2 (Cambridge: Cambridge University press)
[6] Anderson P W 1992 *Science* **256** 1526
[7] Baskaran G, Zou Z and Anderson P W 1987 *Solid State Commun.* **63** 973
[8] Scalapino D J, Loh J E and Hirsch J E 1986 *Phys. Rev.* B **34** 8190
[9] Anderson P W 1997 *The Theory of Superconductivity in the High T_c Cuprates* (Princeton, NJ: Princeton University Press)
[10] Zhang S C 1997 *Science* **275** 1089
[11] Demler E, Hanke W and Zhang S C 2004 (arXiv:cond-mat/0405038)
[12] Kapit E and LeClair A 2008 (arXiv:0805.2951)
[13] Carlson E W, Emery V J, Kivelson S A and Orgad D 2002 (arXiv:cond-mat/0206217)
[14] Annett J, Goldenfeld N and Leggett A 1996 (arXiv:cond-mat/9601060)
[15] Norman M R, Pines D and Kallin C 2005 *Adv. Phys.* **54** 715 (arXiv:cond-mat/0507031)
[16] Lee P A, Nagaosa N and Wen X-G 2004 (arXiv:cond-mat/0410445)
[17] LeClair A 2006 (arXiv:cond-mat/0610639)
LeClair A 2006 (arXiv:cond-mat/0610816)
[18] LeClair A and Neubert M 2007 *J. High Energy Phys.* JHEP10(2007)027 (arXiv:0705.4657)
[19] Vojta M and Sachdev S 1999 *Phys. Rev. Lett.* **83** 3916
[20] Zhou X J *et al* 2003 *Nature* **423** 398
[21] Semenoff G W 1984 *Phys. Rev. Lett.* **53** 2449
[22] Pauli W 1943 *Rev. Mod. Phys.* **15** 175
[23] Bender C M 2007 *Rev. Mod. Phys.* **70** 947 (arXiv:hep-th/0703096)
[24] Mostafazadeh A 2002 *J. Math. Phys.* **43** 205
[25] Georgi H 1982 *Lie Algebras in Particle Physics* (*Frontiers in Physics* vol 54) (New York: Benjamin-Cummings)
[26] Peskin M E and Schroeder D V 1995 *An Introduction to Quantum Field Theory* (Reading, MA: Addison-Wesley)
[27] Wilson K G and Fisher M E 1972 *Phys. Rev. Lett.* **28** 240
Wilson K G and Kogut J 1974 *Phys. Rep.* **12** 75
[28] Nienhuis B 1984 *J. Stat. Phys.* **34** 731
[29] Mermin N D and Wagner H 1966 *Phys. Rev. Lett.* **17** 1133
Hohenberg P C 1967 *Phys. Rev.* **158** 383
Coleman S 1973 *Commun. Math. Phys.* **31** 259
[30] Landau L D and Lifshitz E M 1980 *Statistical Physics* (Oxford: Pergamon)
[31] Ginzburg V L and Landau L D 1950 *JETP (USSR)* **20** 1064
[32] Haldane F D M 1983 *Phys. Lett.* **93A** 464
Haldane F D M 1983 *Phys. Rev. Lett.* **50** 1153
[33] Hüfner S, Hossain M A, Damascelli A and Sawatzky G A 2008 *Rep. Prog. Phys.* **71** 062501 (arXiv:0706.4282)
[34] Fradkin E 1991 *Field Theories of Condensed Matter Systems* (*Frontiers in Physics* vol 82) (Reading, MA: Addison-Wesley)
[35] Affleck I 1988 *Field Theory Methods and Quantum Critical Phenomena* (*Les Houches Lectures*) (Amsterdam: North-Holland)
[36] Yang C N and Zhang S C 1990 *Mod. Phys. Lett.* B **4** 759
[37] Anderson P W 1963 *Solid State Physics* vol 14 ed F Seitz and D Turnbull (New York: Academic) p 99