Modeling the Fermi arc in underdoped cuprates

M. R. Norman, 1 A. Kanigel, 2 M. Randeria, 3 U. Chatterjee, 2 and J. C. Campuzano 2, 1
1Materials Science Division, Argonne National Laboratory, Argonne, IL 60439
2Department of Physics, University of Illinois at Chicago, Chicago, IL 60607
3Department of Physics, The Ohio State University, Columbus, OH 43210

(Dated: February 1, 2008)

Angle resolved photoemission data in the pseudogap phase of underdoped cuprates have revealed the presence of a truncated Fermi surface consisting of Fermi arcs. We compare a number of proposed models for the arcs, and find that the one that best models the data is a d-wave energy gap with a lifetime broadening whose temperature dependence is suggestive of fluctuating pairs.

PACS numbers: 74.25.Jb, 74.72.Hs, 79.60.Bm

I. INTRODUCTION

It is well established that cuprates possess a superconducting phase with an order parameter that has d-wave symmetry 1 and for hole-doped materials, this phase exists over a range of doping above 5%. It is also well established that at very low dopings, the material is an antiferromagnetic Mott insulator. Connecting these two states is an unusual phase known as the pseudogap, the nature of which is still being debated. It is felt by many that the proper identification of this phase will have a major impact on the ultimate ‘mechanism’ for pairing in cuprate superconductors.

Angle resolved photoemission spectroscopy (ARPES) reveals the presence of a truncated Fermi surface in the pseudogap phase. 2, 3 In a study of the pseudogap versus temperature, 4 this truncated Fermi surface was denoted as a ‘Fermi arc’. The arc was shown to be intermediate between the d-wave node of the superconductor and the complete Fermi surface of the normal state. Moreover, the arc appears to form by a closing of the energy gap of the superconducting state as the temperature is raised above Tc. Off the arc, in the ‘pseudogapped’ (antinodal) region of the Brillouin zone, the spectral gap appears instead to fill in with temperature. This filling in effect is also seen in c-axis conductivity data 5 and is consistent with the thermal evolution of the specific heat. 6 This ‘dual’ nature of the energy gap is suggestive of a ‘two gap’ scenario where a ‘superconducting’ gap resides on the arc and a ‘pseudogap’ resides off the arc. Such a two gap picture was proposed by Deutscher 7 and recent Raman 8, 9 ARPES 10, 11, 12 and STM 13 data have been offered in its support. On the other hand, even for underdoped samples, the gap function below Tc seems to be more or less d-wave like 14. This conundrum of having a single gap below Tc transforming into a dual gap above Tc was stressed sometime ago. 15

Recently, a very detailed temperature and doping study of the energy gap above Tc was done by Kanigel et al. 16 They found that the length of the arc scales as T/T*, where T*, the temperature at which the spectral gap ‘fills up’ in the antinodal region of the zone, strongly increases with underdoping. 17 As a consequence, the angular anisotropy of the pseudogap looks more and more like a d-wave gap as the temperature is lowered relative to T*. This finding is supported by thermal conductivity data, which indicates that the d-wave dispersion of the superconducting state at low temperatures survives when the doping is reduced into the pseudogap state. 18 It is also consistent with recent ARPES and STM data on the stripe ordered phase of La2/3Ba1/3CuO4, which indicates a d-wave like gap anisotropy at low temperatures (but above Tc). 19 More recently, the study of Kanigel et al. has been extended to below Tc 20 where it was found that the arc collapses to a node within the resistive width of the transition, with a simple d-wave like gap below Tc. These recent studies bring into question the ‘two gap’ picture.

Primarily motivated by the ARPES data, a wide range of models have been proposed to explain the Fermi arc. Basically, these models can be grouped into two categories. In the first, the pseudogap is associated with a q=0 instability. Most of the models in this category have the pseudogap as a precursor to the superconducting gap, and involve pair formation with the absence of long range order. Most of these models are associated with a q=0 instability. Most of the models in this category have the pseudogap as a precursor to the superconducting gap, and involve pair formation with the absence of long range order. Most of these models are associated with a q=0 instability. Most of the models in this category have the pseudogap as a precursor to the superconducting gap, and involve pair formation with the absence of long range order. Most of these models are associated with a q=0 instability. Most of the models in this category have the pseudogap as a precursor to the superconducting gap, and involve pair formation with the absence of long range order. Most of these models are associated with a q=0 instability. Most of the models in this category have the pseudogap as a precursor to the superconducting gap, and involve pair formation with the absence of long range order. Most of these models are associated with a q=0 instability. Most of the models in this category have the pseudogap as a precursor to the superconducting gap, and involve pair formation with the absence of long range order. Most of these models are associated with a q=0 instability.
category ranges from models based on a precursor spin density wave \cite{36,37} charge density wave \cite{38} stripes \cite{39} flux phases \cite{40,42,43} or orbital currents \cite{44,45}. In the case of fluctuating order \cite{32,41,42,43} the non-zero q vector is not as obvious in the excitation spectrum. Those scenarios involving a \((\pi, \pi)\) wavevector possess small hole pockets centered at \((\pi/2, \pi/2)\) where the intensity is reduced on one side of the pocket due to the amplitude factors which mix the states differing by \(q\). Related models are those where the Luttinger surface (surface of zeros of the single particle Greens function) differs from the Fermi surface.\cite{44,45,46,47}

In this case, the Fermi surface is truncated where it crosses the Luttinger surface. In a more general \(2K_F\) context, the flat parts of the Fermi surface which reside in the antinodal region of the zone can be eliminated by nesting\cite{48,49} leaving a residual arc.

In this paper, some of these scenarios will be addressed in the context of the ARPES data. In Section II, several non-zero q scenarios, where for simplicity long range order is assumed, will be analyzed. These scenarios typically lead to (a) Fermi arcs whose length is \(\pi\) independent, (b) deviations of the arcs from the underlying Fermi surface, (c) energy gaps which are not centered symmetrically about the Fermi energy, and (d) shadow bands. We argue that there is no evidence for these effects in ARPES and tunneling data, at least in the mildly underdoped region. In Section III, we turn to the \(q=0\) solutions. We find that the scenario most consistent with the data is one where the node remains along the zone diagonal and at the Fermi energy. The temperature evolution of the arc above \(T_c\) is consistent with lifetime broadening of the node, though the data also indicate a distortion of the d-wave gap anisotropy with temperature. In Section IV, we offer some conclusions, and suggest future ARPES experiments that could further differentiate between the various models for the Fermi arc.

II. NON ZERO Q SCENARIOS

A. Commensurate density wave

These scenarios assume a q vector of \((\pi, \pi)\) with an energy gap that is either isotropic, or has d-wave symmetry.\cite{40} The secular matrix is of 2 by 2 form, and the Greens function associated with the wavevector \(k\) in the presence of simple elastic broadening, \(\Gamma\), can be written as:

\[
G_k = \left( \frac{E_+ - \epsilon_{k+q}}{E_+ - E_-} \right) \frac{1}{\omega - E_+ + i\Gamma} - \left( \frac{E_- - \epsilon_{k+q}}{E_+ - E_-} \right) \frac{1}{\omega - E_- + i\Gamma}
\]

where

\[
E_{\pm} = \frac{\epsilon_k + \epsilon_{k+q}}{2} \pm \sqrt{\left( \frac{\epsilon_k - \epsilon_{k+q}}{2} \right)^2 + \Delta_k^2}
\]

We have looked at several cases, with various dispersions, \(\epsilon_k\), including some with bilayer splitting, and several different forms for \(\Delta_k\). For brevity, we present results using for \(\epsilon_k\) the tight binding dispersion of Norman et al.\cite{40} and a d-density wave gap\cite{40} \(\Delta_k = \frac{\Delta_0}{2k}(\cos(k_x) - \cos(k_y))\).

In Fig. 1a, we present the intensity plot of the spectral function (imaginary part of \(G_k\)) in the 2D zone for \(\omega = 0\). At the simplest level, one indeed finds an arc. But there are several details worth pointing out. First, the ends of the arc turn away from the underlying Fermi surface of the normal state. This is due to the fact that the zero energy contour traces out a pocket centered at \((\pi/2, \pi/2)\), the back side of which is suppressed by the coherence factors (the prefactor of each term in Eq. 1). Second, there is a strong suppression of the intensity at the ‘hot spots’ - the points where the normal state Fermi surface \((\epsilon_k = 0)\) crosses its \((\pi, \pi)\) displaced image. This can be related to the ‘Luttinger surface’ effect mentioned in the Introduction. To see this, we note that the Greens function in this model can be rewritten as:

\[
G_k^{-1} = \omega - \epsilon_k + i\Gamma - \frac{\Delta_k^2}{\omega - \epsilon_{k+q} + i\Gamma}
\]

The ‘gap’ self-energy (the last term of this equation) diverges when \(\omega = \epsilon_{k+q}\) in the absence of broadening \((\Gamma = 0)\). Thus the \((\pi, \pi)\) translated image of the normal state Fermi surface \((\epsilon_k + q = 0)\) is the Luttinger surface, and therefore the zero energy intensity is suppressed when the normal state Fermi surface crosses this surface. Finally, there is weaker intensity centered around the \((\pi, 0)\) points which will be suppressed as \(\Delta_0\) increases. To investigate this further, in Fig. 1b, we show the spectral function for several \(k\) points along the \((\pi, 0) - (\pi, \pi)\) direction. One clearly sees that the spectral function has a minimum value that sits at negative energy. At \(k = (\pi, 0)\), it is obvious from Eq. 2 that this minimum value occurs at \(\omega = \epsilon_{\pi,0}\), which is \(-34\) meV for this dispersion. This asymmetry in energy is obviously enhanced for dispersions where \(\epsilon_{\pi,0}\) is deeper in energy.
In relation to the experimental data, we note the following issues with this model, which are generic to models based on a finite q order parameter. First, there is no natural way to generate an arc whose length is proportional to the temperature. Second, there is no obvious mechanism to obtain an arc proportional to the d density-wave model. The similarity of these two models can be seen in Fig. 3, where we show the zero energy intensity plot in the 2D zone, as well as the intensity versus ω for k along (π, 0) − (π, π). Again, note the pronounced suppression of the intensity at the ‘hot spots’ in Fig. 3a, which is not evident in the data (a plot of the experimental zero energy intensity around the Fermi energy is shown in Fig. 2b), as well as the pronounced asymmetry of the energy gap relative to the Fermi energy in Fig. 3b. And, as with the d density-wave model, there is no obvious mechanism to obtain an arc proportional to T.

C. Nesting density wave

These scenarios assume a q vector which nests the antinodal points of the 2D Fermi surface. Two approximations were analyzed. In the first, a single q vector along \( q_y, q = (0, -q) \), was used in the first octant (bounded by \( (0,0) - (\pi,0) - (\pi,\pi) - (0,0) \)) of the square lattice zone (a 2 by 2 secular equation), the result of which was then reflected to the other octant. The orientation of q was designed so as to connect the antinode at

FIG. 2: (Color online) (a) Experimental energy distribution curves (EDCs) for optimal doped \( \text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8 \) (Bi2212) around the underlying Fermi surface in the pseudogap phase (T=140K) divided by a resolution broadened Fermi function. The bottom curve is at the node and the top curve at the antinode. The data set is the same as in Fig. 1b of Ref. \cite{16}. (b) Zero energy intensity from (a) as a function of the Fermi surface angle, \( \phi \), where \( \phi = 0^\circ \) corresponds to the antinode and \( \phi = 45^\circ \) to the node. (c) Zero energy intensity versus \( k_x, k_y \) (the data were reflected relative to \( k_x = k_y \)). For (b), the intensities were obtained by subtracting a background EDC (obtained from an unoccupied k), then normalizing this subtracted intensity by its energy integrated weight. This was designed to minimize the effect of the photoemission matrix elements. This was not done in (c) in order to demonstrate that the raw data show no indication for any deviation of the arc from the underlying Fermi surface (black curve).

FIG. 3: (Color online) (a) Spectral intensity at zero energy versus \( k_x, k_y \), and (b) versus energy and \( k_y \) for \( k_x = 1 \), for the model of Yang et al.\cite{44}.

dent. This is consistent with tunneling data as well\cite{45}, where the minimum in the tunneling conductance is at zero bias, even in the pseudogap phase.

B. Differing Luttinger surface

These scenarios\cite{44,45,46,47} are related to the ones just discussed. For discussion purposes, we look at the recently proposed model of Yang, Rice, and Zhang\cite{44}. In this case, the Greens function is

\[
G_{k}^{-1} = \omega - \epsilon_{k} + i\Gamma - \frac{\Delta_{k}^{2}}{\omega + \epsilon_{k}^{NN} + i\Gamma}
\]

where \( \epsilon_{k}^{NN} \) is just the near neighbor term of the tight binding dispersion (\( \Delta_{k} \) has the same form as the d density-wave case). Note that if \( \epsilon_{k} \) only had a near neighbor term, then at half filling, this model would be equivalent to the d density-wave model. The similarity of these two models can be seen in Fig. 3, where we show the zero energy intensity plot in the 2D zone, as well as the intensity versus \( \omega \) for k along (\( \pi,0 \) − (\( \pi,\pi \)). Again, note the pronounced suppression of the intensity at the ‘hot spots’ in Fig. 3a, which is not evident in the data (a plot of the experimental zero energy intensity around the Fermi energy is shown in Fig. 2b), as well as the pronounced asymmetry of the energy gap relative to the Fermi energy in Fig. 3b. And, as with the d density-wave model, there is no obvious mechanism to obtain an arc proportional to T.
\[(\pi, q/2)\] with the one at \((\pi, -q/2)\). The equation for the Greens function is the same as in Eqs. 1 and 2, except that \(\Delta_k\) in this case was taken to be isotropic.

In the second approximation, a 3 by 3 secular equation is separately solved for \(q\) vectors oriented respectively along \(q_x\), \(q = (q, 0)\) and \(q = (-q, 0)\), and along \(q_y\), \(q = (0, q)\) and \(q = (0, -q)\), in the first quadrant of the zone, and then the two results are averaged (representing averaging over two domains). The unaveraged \(G_k\) is given by

\[
G_k = \sum_{i=1}^{3} \frac{(E_i - \epsilon_{k+q})(E_i - \epsilon_{-q})}{(E_i - E_{i+1})(E_i - E_{i+2})} \frac{1}{\omega - E_i + i\Gamma} \tag{5}
\]

where by \(i + 1\) and \(i + 2\) we mean modulo 3. The \(E_i\) are given by solving the appropriate cubic equation and can be written as

\[
E_i = -2\sqrt{d}\cos((\theta + 2\pi i)/3) - a/3 \tag{6}
\]

where

\[
a = -(\epsilon_k + \epsilon_{k+q} + \epsilon_{k-q})
\]

\[
b = \epsilon_k\epsilon_{k+q} + \epsilon_k\epsilon_{k-q} + \epsilon_{k+q}\epsilon_{k-q} - 2\Delta_k^2
\]

\[
c = -\epsilon_k\epsilon_{k+q}\epsilon_{k-q} + 2\Delta_k^2(\epsilon_k + \epsilon_{k+q} + \epsilon_{k-q})
\]

\[
d = (a^2 - 3b)/9
\]

\[
r = (2a^4 - 9ab + 27c)/54
\]

\[
\theta = \cos^{-1}(r/d^{3/2}) \tag{7}
\]

In Fig. 4, we show the zero energy intensity in the 2D zone for the two approximations. Again, a clear arc is seen, with extra structure that can be attributed to the reduced intensity (due again to the coherence factors) of the ‘shadow’ bands. This is particularly true in 4b where more shadow bands occur. A similar situation would occur if one had ‘checkerboard’ order (this would be obtained by solving a 5 by 5 secular matrix associated with a ‘double q’ structure).

A significant difference from the previous cases is the origin of the arc. In the previous cases, the arc is due to the Fermi energy cutting across the lower of the two energy bands. In essence, the energy gap is centered above the Fermi energy for \(k\) vectors from the node to the ‘hot spots’, and it is centered below the Fermi energy for \(k\) vectors from the ‘hot spots’ to the antinode. But in this ‘antinodal’ nesting case, it is the reverse situation. In the 2 by 2 approximation, the arc is formed from the Fermi energy cutting across the upper of the two bands. This is particularly evident near the arc tip, as shown in Fig. 5. In essence, the energy gap is centered below the Fermi energy for \(k\) along the arc. For momenta near the arc tip, one would find a minimum in the spectral function at a negative energy (as in Fig. 1b). This effect is not evident, though, in the ARPES data. And again, as the arc tip is associated with where the underlying normal state Fermi surface intersects the density wave zone boundary (in the first octant, this would correspond to \(k_y = q/2\)), there is no natural mechanism for an arc proportional to temperature. As discussed by McElroy, this would require a ‘two gap’ scenario, where the density wave gap would wipe out the antinodal parts of the Fermi surface, and then a second gap would wipe out the remaining arc with reducing temperature. Despite the attractiveness of such scenarios in regards to some experimental data, a definitive signature of this density wave gap would be to observe the shadow bands evident in Fig. 4 and the asymmetry of the gap relative to the Fermi energy evident in Fig. 5. So far, we have found no evidence for either of these effects.

### III. ZERO Q SCENARIOS

#### A. Energy displaced node

The RVB model of Wen and Lee is based on incorporating both the effect of a d-wave gap in the particle-particle channel and a staggered flux phase gap in the particle-hole channel. An ansatz for the Greens function in this model that makes it of the same form as the earlier cases we studied is

\[
G_k = \frac{\omega - \epsilon_k + i\Gamma - \Delta_k^2}{\omega - \epsilon_k + \mu_{sh} + i\Gamma} \tag{8}
\]
The effect of $\mu_{sh}$ is to move the d-wave node off the Fermi energy, and the sign of $\mu_{sh}$ is chosen to be negative so that the node is above the Fermi energy. The result is an arc at zero energy that is tied to the underlying Fermi surface (Fig. 6a). This is a positive feature of this model. If $|\mu_{sh}|$ were proportional to $T$ for some (unknown) reason, this could also account for the temperature evolution of the arc. The major problem with this model, though, is that could also account for the temperature evolution of the arc. The upturn of the arc length at larger $\Gamma$ is the consequence of the linear variation of $\Delta_k$ with $\phi$ around the node.

The theory of Varma and Zhu is similar except that $\Delta_k$ is taken to be the square of the d-wave gap. As they point out, this fits the angular anisotropy of the parameter $\Delta_k$ in the pseudogap phase better than the simple d-wave model, as can be seen in Fig. 9b. On the other hand, the arc length variation with $T$ is more consistent with the simple d-wave model, as shown in Fig. 9a, though we remark that Varma and Zhu were able to obtain a much better fit to the arc length by allowing a self-energy with a more sophisticated frequency and temperature dependence.

In Fig. 9a, we show the variation of the arc length with $\Gamma$. This variation is consistent with experiment, as shown in Fig. 9a, if one assumes that $\Gamma \propto T$ and $\Delta_0$ is a constant in $T$ (similar plots are shown in Refs. 43,54). This linear variation of the arc length with $\Gamma$ is a natural consequence of the linear variation of $\Delta_k$ with $\phi$ around the node. The upturn of the arc length at larger $\Gamma$ is due to the quadratic dependence of the energy gap with $\phi$ about the antinode.

The theory of Varma and Zhu is similar except that $\Delta_k$ is taken to be the square of the d-wave gap. As they point out, this fits the angular anisotropy of the parameter $\Delta_k$ in the pseudogap phase better than the simple d-wave model, as can be seen in Fig. 9b. On the other hand, the arc length variation with $T$ is more consistent with the simple d-wave form, as shown in Fig. 9a, though we remark that Varma and Zhu were able to obtain a much better fit to the arc length by allowing a self-energy with a more sophisticated frequency and temperature dependence.

The simplest model in this class is equivalent to the one just described with $\mu_{sh} = 0$:

$$G^{-1}_k = \omega - \epsilon_k - i\Gamma - \frac{\Delta_k^2}{\omega + \epsilon_k + i\Gamma}.$$  \hspace{1cm} (9)

The spectral function for finite $\Gamma$ traces out an ‘arc’, as shown in Fig. 7a. And the energy gap is centered at the Fermi energy, as shown in Fig. 7b. That is, the gap is tied to the Fermi energy and the Fermi surface, consistent with experiment. In Fig. 8a, we plot the evolution of the spectral function on the Fermi surface ($\epsilon_k = 0$) for this model, and in Fig. 8b the angular anisotropy of the spectral gap (half the peak to peak separation).

Gapped and ungapped spectra on the Fermi surface (Fig. 8a) are obviously controlled by the sign of the second derivative of the spectral function with respect to $\omega$ at $\omega = 0$. The condition that this second derivative is zero is $\Gamma = \sqrt{3}\Delta_k$. Assuming a simple d-wave gap of the form $\Delta_k = \Delta_0 \cos(2\phi)$ where $\phi$ is the Fermi surface angle measured relative to the antinode, one then obtains for the position of the arc tip $\phi_0 = 0.5 \cos^{-1}(\Gamma/\sqrt{3}\Delta_0)$. $T^*$ would then be the condition that $\Gamma(T) = \sqrt{3}\Delta_0(T)$.

IV. CONCLUSIONS

In regards to the ‘non zero q’ scenarios, there are several ways that experiment could address this question.
Definitive evidence would be finding a departure of the arc from the underlying normal state Fermi surface, evidence for an energy gap which is asymmetric in energy relative to the chemical potential, or the existence of shadow bands (i.e., bands displaced from the main band by the wavevector \( q \)). Other evidence would be the existence of intensity suppression at ‘hot spots’ (where the Fermi surface would cross the Luttinger surface), as has been observed by ARPES in electron doped cuprates.\(^5\)\(^5\)

In regards to the ‘zero \( q \)’ scenarios, the simplest theory consistent with the data appears to be a d-wave gap with an inverse lifetime that is proportional to \( T \). There are, though, some limitations of this model. The data of Ref. 16 were actually fit with a form more general than that of Eq. 9:

\[
G_k^{-1} = \omega - \epsilon_k + i\Gamma_1 - \frac{\Delta_k^2}{\omega + \epsilon_k + i\Gamma_0} \quad (10)
\]

This ‘two lifetime’ model has the advantage of being able to describe a broad spectral function (\( \Gamma_1 \)) but with a sharp leading edge gap (\( \Gamma_0 \)) as indicated by ARPES data in the pseudogap phase.\(^4\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)\(^5\)\(^6\)

This work was supported by the U.S. DOE, Office of Science, under Contract No. DE-AC02-06CH11357 and by NSF DMR-0606255.

Acknowledgments

This work was supported by the U.S. DOE, Office of Science, under Contract No. DE-AC02-06CH11357 and by NSF DMR-0606255.
58 G. M. Eliashberg, J. Exp. Theor. Phys. 11, 696 (1960).
59 K. Maki and H. Won, Physica C 282-287, 1839 (1997).
60 E. Abrahams, A. V. Chubukov, A. J. Millis and M. R. Norman, unpublished.