High Resolution Photoemission Experiments on Copper Oxide Superconductors

A Dissertation Presented

by

Jonathan David Rameau

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

May 2009

Stony Brook University

The Graduate School

Jonathan David Rameau

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Peter David Johnson – Dissertation Advisor
Professor, Department of Physics and Astronomy

Philip Allen – Chairperson of Defense
Professor, Department of Physics and Astronomy

Harold Metcalf
Professor, Department of Physics and Astronomy

John Tranquada
Physicist
CMPMSD, Brookhaven National Laboratory

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

High Resolution Photoemission Experiments on Copper Oxide Superconductors

by

Jonathan David Rameau

Doctor of Philosophy

in

Physics

Stony Brook University

2009

The mechanism for achieving high transition temperatures (T_c) in copper oxide superconductors and the nature of the mysterious ”pseudogap” phase from which this phenomenon arises are two of the most pressing issues in solid state physics. High resolution angle resolved photoemission spectroscopy (ARPES), which can directly probe the momentum and energy dependence of the electronic structure of a crystal, is considered one of the foremost tools for unraveling these mysteries. In this thesis we present work on both the further development of the ARPES technique itself and the results of two experiments on the high temperature super-

conductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ (BSCCO) - the drosophila of copper oxide superconductors - based upon these analytical and experimental advances. On the analytical side we have shown that the precision of any ARPES experiment can be radically enhanced by using the Lucy-Richardson method (LRM) of iterative deconvolution to remove the worst effects of experimental resolution broadening present in all ARPES spectra. On the experimental side we have constructed a deep ultraviolet laser system capable of increasing our data acquisition rate by more than an order of magnitude compared to what is possible using traditional synchrotron radiation sources at the same momentum and energy resolutions. Using the LRM, in conjunction with synchrotron radiation, spectroscopic evidence was found for the existence of incoherent Cooper pairs in underdoped BSCCO in the normal pseudogap state (above T_c). At the same time an asymmetry between the particle and hole states of BSCCO was found, implying that doped Mott insulators, of which BSCCO is a primordial example, are characterized by the presence of a Fermi-Luttinger surface, rather than a Fermi surface, as would be the case for a simple metal. This study provided the first spectroscopic evidence for either phenomenon. In our second experiment we were able to use the LRM on data acquired with the laser ARPES system to show the presence in optimally doped BSCCO, well below T_c, of a previously unobserved electron-boson interaction. The momentum dependence of this interaction, which appears as a "kink" in ARPES spectra, shows that the responsible boson is an Einstein phonon. The ARPES data is well reproduced by a simple theoretical model based on these observations. After more than two decades of study this result represents the first unambiguous enumeration of an electron-boson interaction by ARPES in BSCCO. The identification of this interaction will require a theoretical reevaluation of the nature of several similar features long seen in BSCCO, the origins of which remain unknown and highly controversial to this day.

To my mom, Marty, Sarah, Cowboy and all my friends and family who never doubted I could become a real physicist.

Contents

Acknowledgements . . . . **xxvii**

1 Introduction **1**

2 Copper Oxide High Temperature Superconductors **6**

2.1 High and Low Temperature Superconductivity 6

2.2 Crystal Structures and the Families of the Cuprates 9

2.3 The Cuprate and BSCCO Phase Diagrams 12

2.4 The "Ideal" Electronic Structure of Bi2212 15

2.5 Gaps and Arcs in the Hole Doped Cuprates 22

2.6 Electron-Boson Interactions 28

3 Photoemission Spectroscopy **34**

3.1 Kinematics of Photoemission 34

3.2 Quantum Theory of Photoemission 45

3.3 The Modern Practice of ARPES 54

3.4 Very Low Energy ARPES 67

3.5 Coherence in the ARPES Light Source 70

4 Harmonic Generation **77**

4.1 Nonlinear Mechanics of a Driven Harmonic Oscillator 79
4.2 The Wave Equation in Ponderable Media 83
4.3 Phase Matching and Crystal Properties 86
4.4 Fundamental Considerations 92
4.5 Simulation 97
4.6 Outlook 102

5 Experimental Apparatus 106
5.1 Laser System 107
5.2 Photon Spectroscopy 125
5.3 Vacuum System 129
5.4 Magnetic Shielding 132
5.5 Sample Handling 139
5.6 Auxiliary Apparatus 142

6 Lucy-Richardson Deconvolution and Fermi Normalization . 145
6.1 Photoemission Above the Fermi Level 145
6.2 The Resolution Problem 148
6.3 Lucy Richardson Deconvolution 155
6.4 Deconvolution of Simulated ARPES Spectra 164
6.5 Finer Points of Deconvolving ARPES Spectra 177
6.6 Deconvolution and Fermi Normalization of Real Data 189

7 Laser ARPES on Optimally Doped Bi2212 202
7.1 Introduction 202
7.2 Laser ARPES Data and Initial Findings 204

7.3 Extraction of Quantitative Information from the Data 213

7.4 The Identification of a Boson 219

7.5 Self Consistency and the Coupling Constant 224

7.6 Conclusion of the Phonon Story 232

8 Unoccupied Electronic States of Underdoped and Optimally Doped Bi2212 238

8.1 Methodology and Expectations 238

8.2 Result of the Examination of Unoccupied States in Optimally and Under Doped Bi2212 242

8.3 Interpreting the Results 250

  8.3.1 The Nodal Region 251

  8.3.2 The Antinodal Region 261

8.4 Tieing it All Together 263

Bibliography 266

List of Figures

2.1 Primitive unit cells of the Bismuth based cuprates. 10

2.2 Generic phase diagram of the electron and hole doped cuprates adapted from reference [1]. 13

2.3 Temperature vs. doping δ phase diagram of Bi2212 as it is currently understood. 14

2.4 Simplified, schematic representation of a) the AFM/Mott insulating state on the copper oxide square lattice. Copper atoms are orange; oxygen atoms are green. Arrows indicate the relative spin orientation. b) upon doping the AFM/Mott insulator nodal conductivity becomes preferred. Hopping along the bond direction is still suppressed by the on site Coulomb repulsion U. 17

2.5 Single LDA band approximating the dispersion of the Cu-O $pd\sigma$ band on the square lattice. 20

2.6 Electronic dispersion along the high symmetry lines of the tetragonal cuprate Brillouin zone. 21

2.7 LDA Fermi surface of Bi2212 in the single band approximation from Eq. 2.1. 22

2.8 a) Normal state Fermi surface of a heavily overdoped cuprate. b) and c) the Fermi arcs for progressively more underdoped cuprates in the normal psuedogap state. d) nodal points in the d-wave superconducting state. 23

2.9 Schematic representations of the s-wave, p-wave and d-wave order parameters. The + and - labels denote relative phases between lobes of p- and d-wave type superconductors. The distance from the origin in all three cases denotes relative gap magnitudes. 25

2.10 Ideal d-wave superconducting gap as a function of Fermi surface angle ϕ from Eq. 2.2. Here we take the node as the zero angle. 27

2.11 The classic kink in Bi2212. Panels a)-c) show the nodal spectrum of under doped, optimally doped and over doped Bi2212, respectively, with T_c indicated. Lower panels show the MDC derived dispersions above and below T_c. 30

2.12 Historical evolution of the nodal kinkology of Bi2212. In all panels the dashed line denotes the supposed unrenormalized, or bare, LDA-type band used to directly extract the change in the real part of the self energy induced by the interaction. a) the picture used in the analysis of Figure 2.11 b) the same picture modified after observations proposing the existence of a "big" kink at 350 meV. c) the same picture now including the 8 meV kink discussed in this thesis. d) Generic anatomy of a kink showing the deviation of dispersion from the bare band with the subsequent broadening of the electron lifetime above the kink energy. Note that the existence of the big kink remains a topic of debate. 31

3.1 Schematic of the original photoelectric effect experiment. Light with sufficient energy $h\nu$ ejects electrons from a cathode. The work function ϕ is determined from the stopping potential V applied between the anode and cathode. 36

3.2 Energy level diagram relating density of states in a sample to what is measured in a photoelectron spectrometer. 39

3.3 All angles parameterizing an ARPES experiment are measured relative to the sample normal $\hat{n}$. (α, β) and (θ, φ) are the azimuthal and polar angles of the incident photon and exiting photoelectron, resepctively. 42

3.4 Cutaways of generic hemispheric electron spectrometer used in ARPES showing a) kinetic energy sorting geometry and b) angle sorting geometry in hemisphere. Note that the entrance slit can only define energy resolution, not angular resolution. 57

3.5 An early sample of 6 eV laser ARPES data taken at 20 K on optimally doped Bi2212 in the nodal region. Panel a) shows the 2D spectrum in energy and angle space. The horizontal and vertical dashed lines mark E_F and k_F, respectively. Intensities are mapped to the false colorscale per the colorbar at the top of the panel. Panels b) and c) show EDC and MDC stack plots, respectively, as are often found in the literature. 60

3.6 Inelastic mean free path of electrons in a crystal, in nanometers, as a function of electron kinetic energy. Adapted from reference [2] 63

3.7 A typical 6 eV laser ARPES spectrum near the node of Bi2212. Panel a) shows the spectrum as it's acquired by the electron spectrometer. Panel b) shows the result of kinematic compression upon converting from θ to $k_\parallel$ space. Regions of the MCP detector beyond those where spectra are typically acquired are shown for clarity. The main effect of the conversion from angle to momentum space is the "tilting" and compression of the spectrum in the θ direction and the subsequent sharpening in energy of the observed dispersing band. 70

3.8 Lines of constant angle in k space. The red cut through (k_x, k_y, E), easily obtainable at high photon energies, must be constructed from many cuts though angle space in low energy ARPES. . . 71

4.1 Schematic of polarization properties of a uniaxial birefringent crystal. Gray lines indicate the manner in which the crystal structural symmetry is broken. The optic axis is defined by the wave propagation direction for which all polarizations experience the same index of refraction. Ordinary (o) polarized waves experience the same index of refraction n_o as if they traveled along the optic axis. Extraordinary (e) polarized waves experience a fraction of the extraordinary index of refraction n_e depending upon the tuning angle θ of their angle of incidence. 89

4.2 Initial parameters for the SNLO 2D short pulse mixing algorithm for second harmonic generation in LBO of a 410.5 nm beam from the 821 nm fundamental. 100

4.3 Initial parameters for the SNLO 2D short pulse mixing algorithm for second harmonic generation in BBO of a 205.25 nm beam from the 410.5 nm second harmonic of a Ti:Sapphire laser. 103

4.4 Plots of power and spectral output of SNLO simulations of second and fourth harmonic generation. a),b) are the normalized and absolute power, respectively, as a function of time and c) is the spectral distribution of the fundamental (Idler) and second harmonic (Pump) pulses after the harmonic conversion in LBO initialized with the parameters in Fig. 4.2. d)-f) same as a)-c) for the result of simulating the doubling of the second harmonic with parameters in Fig. 4.3, which were themselves obtained from the results of the previous simulation. 104

5.1 Schematic of the VUV ring experimental hall of the NSLS circa 2008. The laser hutches at U6 and U13A as well as the hutch enclosing U13UB are shown. The steel laser pipe connecting U6 to U13A is shown as the thick red line connecting the two hutches. 111

5.2 Arrangement of optics and lasers in the U6 hutch. The Ti:Sapphire oscillator output was first corrected for ellipticity and then expanded before being steered through the beam pipe to U13. The last mirror, residing at the pipe entrance, was moroized and controlled remotely as needed. 112

5.3 The turning mirror at the end of the pipe is removed during initial alignment of the 820 nm beam from U6 to U13. Alignment is established by monitoring the signal of a thermopile power meter. The mirror turning mirror is then replaced and the beam steered through the last two meters of pipe into the U13A laser enclosure. 114

5.4 Layout of optics for generation and characterization of second and third harmonics of the Ti:Sap fundamental beam. An enclosure (not shown) was installed over the optics bench to limit dust and air currents that induced noise in the second and fourth harmonics. Waveplates ensured that the fourth harmonic was s-polarized, and the second harmonic p-polarized, with respect to the $\mathrm{ArF_2}$ mirrors. This arrangement ensured their function as effective dichroic mirrors. 117

5.5 Image and real time x and y profiles of the fourth harmonic beam taken with laser beam profiler. The fourth harmonic beam has been separated from the second harmonic beam by the grating monochrometer. 120

5.6 Laser beamline. 123

5.7 Layout of home built imaging spectrometer. 126

5.8 Calibration plot for home built imaging spectrometer. The λ points are read from the Jobin Yvon spectrometer and the pixel value from stored camera images taken at each wavelength calibration point. 128

5.9 Finite difference calculation of electron trajectories normal to a magnetic field pointing out of the plane. Electrons originate from a point at (x,y)=(0,0) with initial velocities in the $\hat{y}$ direction. Trajectories of electrons with kinetic energies between .5 and 2 eV are shown for $B_z = 0.01$ Gauss (blue), 0.005 Gauss (red) and 0.001 Gauss (black). Electrons with kinetic energies between 21.5 and 23 eV in a 0.005 Gauss field – a condition similar to typical synchrotron experiments – are shown in green. The horizontal dashed line marks the distance from the sample to the entrance aperture of the SES-2002 electron spectrometer. The vertical dashed and dotted lines mark the half widths of a .5 mm and .1 mm analyzer entrance slit, respectively. Calculations were performed with time steps of 10^{-10} seconds. 134

5.10 Arrangement of magnetic field shim coils. The vertical field canceling coils are marked red and the horizontal coils blue. The sense of the currents driving the coils are marked by arrows. The entrance to the Scienta analyzer is green. The position of the sample, mounted on a cryostat, relative to the coils and the analyzer is also shown. 138

6.1 Fermi-Dirac distribution for several temperatures. For ARPES spectra with good counting statistics, dividing out the Fermi function reveals excitations up to $4k_BT$ above the Fermi level. The dashed lines indicate $4k_BT$ for the curves at 80 K, 140 K and 300 K; 27.6 meV, 48.3 meV and 103.4 meV. 148

6.2 Schematic of ARPES data acquisition modes. In ADC mode hits on the detector phosphor screen are converted to grayscale intensities by the CCD camera spread across several pixels. In pulse counting mode each pulse is registered as a single count at the energy and angle position of the pulse center. 152

6.3 Simulations of a linear band crossing E_F. a) Theoretical spectral function $A^0(k_\parallel, \omega)$ b) $\tilde{A}^0(k_\parallel, \omega) = A^0(k_\parallel, \omega)/f(\omega; T_{exp})$ c) $I_0(k', \omega') = \tilde{A}^0(k_\parallel, \omega) \bigotimes R(k, \omega | k', \omega')$, d) $\tilde{A}^7(k, \omega)$ 165

6.4 a) EDC's at k_F and b) MDC's taken at E_F. The thick blue lines are from the theoretical $\tilde{A}$, the thick black lines are from I_0, the thick red lines are from $\tilde{A}^7$ and the thin green lines are intermediate iterations. 167

6.5 a) $I_0/f(\omega; T_{exp})$, b) $I_0/f(\omega; T')$, c) $\tilde{A}^7/f(\omega; T_{exp})$ 168

6.6 a) Lorentzian FWHM $\Delta k(\omega)$, b) spectral weight Z_ω and c) MDC peak positions $k_0(\omega)$. In all panels, the magenta line is the fit to theory, Fig 6.3a), the blue line is the fit to Fig. 6.5a), the red line is the fit to 6.5b) and the black line is the fit to 6.5c). The dashed line marks $4k_BT$, the energy up to which one can usually recover spectra in a real experiment. 170

6.7 Simulations for a single band with a 7 meV BCS type superconducting gap. a) the whole spectral function, b) cut off by the Fermi distribution, c) blurred by 10 meV energy and .01 degree angular resolution and d) result of deconvolving c) with the LRM after seven iterations. 173

6.8 EDC's at k_F taken from the unbroadened thoery (blue curve), the broadened theory (black curve), the deconvolved spectrum (red curve) after seven iterations and intermediate results of the LRM (green curves). 175

6.9 FNS resulting from a) dividing the broadened theory by the experimental Fermi function, b) dividing the broadened theory by the resolution broadened Fermi function and c) dividing the deconvolved theory, after seven iterations, by the experimental Fermi function. 177

6.10 EDC's at k_F for the full theoretical spectral function (pink curve), the broadened theory divided by the experimental Fermi function and broadened Fermi function (black and red curves, respectively) and the deconvolved theory (after seven iterations) divided by the experimental Fermi function (blue curve). . . . 178

6.11 EDC's taken at k_F corresponding to the original full theoretical spectral function (pink curve) and spectra divided by the physical Fermi function after seven iterations (blue curve) and twenty iterations (green curve). The inset shows the FNS after twenty iterations. 181

6.12 Comparison of the occupied spectral function after broadening to the re-broadened, deconvolved spectral function used to calculate $\chi^2(r)$. 182

6.13 Log plot of $\chi^2(r)$. 183

6.14 Effective temperature $T^{\aleph}$ a) and FWHM Gaussian effective resolution $\Delta E^{\aleph}$ b) fit for a Fermi distribution initialized at 50 Kelvin with an initial energy broadening (FWHM) of 15 meV. The black line in panel a) denotes the "physical" temperature of the Fermi function. The fifth iteration, going too far, yields an imaginary result for the resolution calculated for panel b) (not shown). The inset of the first panel shows $\chi^2(r)$ for the LRM in this instance out to 10 iterations on a linear scale. . . 186

6.15 Cut in k space taken at location diagrammed in inset of panel e) on optimally doped Bi2212 at 60 K. Panel a) shows the original spectrum for which the experimental energy and angular resolution was set to 16.5 meV and .1^o, respectively. Panels b) through e) show the result of applying the LRM to the spectrum in panel a) after 4, 6, 8 and 10 iterations. The false color intensity scale for all panels is shown in the inset of panel a). The Fermi level is indicated by dotted lines. 192

6.16 Comparison of different Fermi normalization methods applied to the data in Fig. 6.15. Panel a) results from normalizing the original data (Fig. 6.15a) to a Fermi function at the experimental temperature. Panel b) shows the result of normalizing the same spectrum to a resolution broadened Fermi function. Panels c)-e) show the result of normalizing to the experimental Fermi function after 4, 8 and 10 iterations, respectively. The Fermi level is indicated by dotted lines. The dashed lines indicate $4k_BT$ and $5k_BT$, respectively. The false color intensity scale is the same as Fig. 6.15. 194

6.17 EDC stack plots generated from corresponding panels in Fig. 6.16. The Fermi level and $4k_BT$ are marked by the solid lines. 195

6.18 $\chi^2(r)$ for the deconvolution of Fig. 6.15. 197

6.19 (a) the raw data before analysis, (b) the same data as in (a) after dividing by the Fermi function, (c) the spectra after dividing the data in (a) by the Fermi function broadened with a function representing the experimental resolution and (d) the deconvoluted spectra (derived from (a) using the LRM) divided by the Fermi distribution function. The solid black lines represent the EDC corresponding to the Fermi wave-vector, k_F. 198

7.1 Laser ARPES spectra from the nodal region of optimally doped Bi2212 at 10 K. Panels a-c are the raw data. Panels d-f are the same spectra deconvolved using the LRM. The locations of the cuts in the Brillouin zone are diagrammed in the Figure 7.5a inset. 206

7.2 Near nodal spectrum of optimally doped Bi2212 at 8 K taken at the VUV ring of the NSLS with a photon energy of 16.7 eV. The orientation of the cut in the Brillouin zone is shown by the dashed line in panel c). b) MDC's at $\omega = 0$ meV (circles) and -10 meV (triangles). The peaks are fit by two Lorentzians. d) EDC at $k - k_F$= -.015 Å^1 with a single Lorentzian fit to the high binding energy (bonding band) peak. [3] 207

7.3 Panel a) EDC's and panel b) MDC's taken at the k_F and E_F, respectively. The blue curves are from the raw data of Fig. 7.1a and the red curves from the deconvolved data of Fig. 7.1d. All cuts are taken along a single column or row of channels demonstrating the superior signal to noise that can be achieved with a laser. Note that in addition to sharpening the peaks the peak positions themselves are corrected by application of the LRM. 209

7.4 Fits to the deconvolved EDC (panel a) and MDC (panel b) shown in Fig. 7.3. Both fits are performed using Lorentzians on a linear background. The EDC fit is cut off by the Fermi-Dirac distribution. Peak widths ΔE and Δk are shown in the figure. 210

7.5 Imaginary part of the self energy as a function of binding energy extracted from the deconvolved spectra of Figure 7.1d-f. The scattering rate $\Gamma_k(\omega) = 2\,\mathrm{Im}\,\Sigma_k(\omega)$. The dotted vertical lines denote the gap position corrected for kinematic and resolution effects. Double valued points in the vicinity of the gap result from the parabolic dispersion along the gap edge upon crossing k_F. 216

7.6 Important energy scales extracted from the self energies of Figure 7.5 as a function of Fermi surface angle ϕ. The red points are Ω_k, the blue points are Δ_k and the black points are $\delta_k = |\Omega_k - \Delta_k|$. Lines connecting points are a guide to the eye. . . 220

7.7 Confocal Raman spectrum of Bi2212 at room temperature using a 532 nm excitation source. 223

7.8 Analytical approximation of electron-phonon coupling to an Einstein mode; a) $\alpha^2F(\omega)$ as a delta function at Ω from Eq. 7.4, b) $2\,\mathrm{Im}\,\Sigma(\omega)$ calculated from Eq. 7.5, c) $\mathrm{Re}\,\Sigma(\omega)$ calculated from Eq. 7.6 226

7.9 Numerical calculation of electron-phonon coupling to an Einstein mode; a) $\alpha^2F(\omega)$ with $\alpha^2 = 1.5$ and a finite width as derived from the Raman spectrum of Figure 7.7, b) $2\,\mathrm{Im}\,\Sigma(\omega)$ calculated from Eq. 7.2 using the function in panel a), c) $\mathrm{Re}\,\Sigma(\omega)$ calculated from Eq. 7.7 with step cutoff at 70 meV. 227

7.10 Simulation of nodal BSCCO ARPES spectrum at 10 K using the experimentally derived Einstein phonon spectrum. 231

7.11 Schematic of the Bi2212 primitive unit cell atomic motions associated with the 8 meV A_{1g} optical phonon mode. The arrow lengths depict the relative magnitudes of displacement suffered by various atoms. Note that in the Bi2212 primitive unit cell the out of plane apical oxygens are located directly above (below) the copper atoms of the upper (lower) copper-oxygen planes. This figure is adapted from Reference [4]. 235

8.1 The d-wave weak coupling BCS gap Δ as a function of temperature T and Fermi surface angle ϕ for a superconductor with $T_c = 91$ K. Regions of zero height on the z axis denote Fermi surface angles for which unoccupied Bogoliubov quasiparticles are unobservable by the thermal population technique. As one raises the temperature from absolute zero unoccupied states of the d-wave gap further from the node can be observed. For optimal doping one can observe low lying unoccupied states across the whole Brillouin zone for $T \gtrsim 50$ K. 240

8.2 Spectra from optimally doped Bi2212 below T$_c$ treated by the LRM a) before and b) after Fermi normalization. Panels c) and d) show EDC's corresponding to panels a) and b), respectively. The location of the cut through the Brillouin zone is shown in the inset of panel a). 244

8.3 FNS of optimally doped (panels a) and b)) and T_c= 65 K underdoped (panels c) and d)) Bi2212 below T_c. Panels a) and c) show spectra acquired at the nodes. Panels b) and d) show spectra acquired away from the nodes at Fermi surface angles ϕ shown in the insets. 246

8.4 Normal state FNS for optimally doped (panels a-c) and underdoped (panels e-g) Bi2212 in the normal pseudogap state at T = 140 K. Panels d and h show the cut locations in k space for the optimal and underdoped samples, respectively. Panel h shows a small, closed Fermi pocket consistent with the underdoped FNS data. The area of this pocket is consistent with the nominal doping level of the underdoped sample, about .12. The underdoped sample displays a striking electron-hole asymmetry in the normal state. 247

8.5 FNS from the T_c = 65 K samples a) at 50 K, b,d) at 140 K and c) 40 K. The cut locations are diagrammed in panel f. Panel e) EDC's at k_F from panels a and b just off the Fermi arc. Panel g) EDC's at k_F from panels c and d near the antinode. Nodal states display electron-hole asymmetry in the normal state with symmetry restored in the superconducting state. Antinodal states display electron-hole symmetry in both the normal pseudogap and superconducting states. 249

8.6 Model calculation from Yang et al. (unpublished) showing how the band structure is renormalized away from the nodes by the opening of a gap above E_F. The thickness of the bands schematically denotes the relative spectral weights of the "rising" and "falling" parts of the band arising from coherence factors similar to the BCS u_k^2 and v_k^2. 257

Acknowledgements

Real physics happens outside of a vacuum, not inside of one. To that end I would like to thank and acknowledge the hard work and dedication of the Electron Spectroscopy group of Brookhaven National Laboratory in teaching, guiding and often lending to me two strong hands as I went through my graduate work. In particular, I would like to acknowledge our post docs Tim Kidd and Hongbo Yang, our Emeritus, Myron Strongin, Toni Valla, Chris Homes and Alexei Tsvelik. Tim taught me how to do ARPES, analyze ARPES data, mount samples and fix toys after I break them. Hongbo, who succeeded Tim, was an invaluable second pair of hands and often second brain as we did the heavy lifting constructing a safe and functional laser ARPES facility at the NSLS. He also did a magnificent job leading the Fermi normalization experiment and I'm sure neither of us will forget running alternating twelve hour shifts on and off for a few months at the NSLS to get the data for that experiment. Myron reminded us all how much basic physics had been explored decades ago in solid state, often by him, that we still didn't understand. Toni reminded us all of what was true and what wasn't and was not shy about putting a bad idea to rest as quickly as possible. Chris and Alexei provided valuable insight from the world beyond photoemission and were always cheerfully willing to discuss with and teach me new physics. I would also like to recognize our technicians – though the name hardly does their skill justice – Fran Loeb and Al Langhorn without whose abilities none of these experiments would have gotten off the ground. I would like to thank Larry Carr of the NSLS who taught me how to use really big lasers and that a synchrotron isn't just a big black box that spits out light when you push button. Last but certainly not least I would like to acknowledge my advisor, Peter Johnson, who taught me to be ambitious and not to settle when I could potentially do something really good, who funded and guided my efforts to learn how to be an experimentalist, encouraged me when things were rough not to return to life behind the counter of a Starbucks, whose advice was usually good even when I hated it and ultimately produced what I hope is a decent scientist whom can make some contributions to condensed matter physics in the years to come.

Chapter 1

Introduction

High temperature superconductivity in the copper-oxide based doped Mott insulators has presented the foremost challenge in solid state physics since the primordial compound, Lanthanum Barium Copper Oxide, (LBCO), was discovered by Bednorz and Muller in 1986. Angle resolved photoemission spectroscopy (ARPES) has been one of the foremost tools in the advancement of our understanding of this phenomenon.

It is the object of this introductory chapter to explain what the above statement means. In so doing we shall attempt to lay out the jargon common to the field and used often in the rest of this thesis. This is necessary in order to not only present the list of methods and results that constitute a dissertation but to do so in the context of a wider argument about one of the most pressing problems in modern physics. My intention is that this work can be read as a self consistent whole by either a seasoned researcher in the field or by someone approaching either ARPES or high temperature superconducivity for the first time. Hopefully both types of reader will find the work not entirely uninteresting. To that end, I have attempted to adopt a more colloquial language in the text, where appropriate, than is often found in volumes such as these. While perhaps unfit for the summary of research proper to a journal article or other brief communication it is my firm belief that not only does this respect for language not harm a work of this scope and size but may even render it somewhat enjoyable. Consider it an experiment in applied liberal arts. From now on I shall even bend to the scientific heresy of the royal we!

This thesis is organized as follows.

In the second chapter we shall briefly describe what high temperature superconductivity *is*. We shall then briefly describe the materials in which the phenomenon occurs and, most importantly, sketch the generic phase diagram of the materials that exhibit this property as it is currently understood. We will then connect this phase diagram to the two properties that decisively determine the fate of any crystalline material; the crystal structure of copper-oxide superconductors and the electronic structure that results from it. In the course of this discussion we shall use as our example a particular family high temperature superconductor, $Bi_2Sr_2CaCu_2O_{8+\delta}$, known in the field as either BSCCO (pronounced Bisco) or Bi2212, because it is the material family studied most often by ARPES. This nomenclature and the reasoning behind using this material as our primary object of study will become clear shortly.

In Chapter 3 we shall review the photoelectric effect, a.k.a. photoemission, and describe the theory of photoemission from crystals. It will be shown that the generalization of the photoelectric effect to account for the anisotropy of crystalline media results in a photoelectron distribution in emission angle that reflects the distribution of electrons in the momentum space of the crystal cathode. This effect allows the use of angle resolved photoemission spectroscopy, the study of the density of states of emitted photoelectrons as a function of emission direction and kinetic energy, as a direct probe of the electronic structure of crystals and through that structure and its details of the field theoretic propagators that describe the quantum electrodynamics of the system. Understanding this information gives important insight into the electronic properties, such as high temperature superconductivity, one expects a system to display and thus is of paramount importance in the study of condensed matter systems. We shall give particular emphasis in this chapter to the peculiar properties of very low energy photoemission pertinent to a laser based ARPES system. The construction of this system and its operation were the primary objective of the dissertation work presented herein.

In chapter 4 we shall briefly review the basic theory of second harmonic generation and present the results of numerical simulations showing that a two step harmonic generation process can produce a high brightness, narrow bandwidth deep ultraviolet (DUV) laser beam useful for photoemission. These simulations provided the theoretical guide for the construction of a high rep rate DUV laser system for ARPES. This was done for the following reason; ARPES requires a light source with a photon energy greater than both the binding energy of the electrons we wish to observe and the surface energy barrier, called the work function, of a sample. Successive harmonic generation steps can yield an ultraviolet laser of energy 6 eV that is both brighter and of narrower bandwidth than what can be achieved by more traditional methods such as helium discharge lamps and synchrotron light sources. The relatively low energy of the laser light translates into both higher electron momentum and electron energy resolution.

In Chapter 5 we describe in detail the physical arrangement and construction of the laser ARPES system and accompanying instrumentation. This includes, naturally, a description of the laser itself and the setup of optics used in the successive harmonic generation steps used to produce a 6 eV laser beam. It also includes details of the construction and operation of a home built imaging spectrometer for DUV beam characterization and the techniques used by our group to achieve and maintain ultra high vacuum (UHV), safe class four laser operation in a public space and unconventional magnetic field compensation techniques. Taken together the chapter lays out the key components for the construction of a high resolution laser ARPES station.

Chapter 6 describes a major development in our ability to analyze ARPES data that was totally unanticipated as we set out on the laser ARPES project. In the course of a very demanding experiment using synchrotron radiation it was found that the best energy resolution we could achieve for the exceptionally high signal to noise ratio required was not good enough to allow a reliable, physically meaningful analysis of the data. To save the day and render useful several months worth of round the clock data collection we adapted an algorithm often used in astronomy, Lucy-Richardson deconvolution, to remove most effects of the experimentally induced energy and momentum broadening in our spectra. The application of this technique to ARPES data of any kind has since become a mainstay of our analytical toolkit.

Lastly, in Chapters 7 and 8 we apply the results obtained in the preceding chapters to two very different experiments. In the first experiment we apply our new resolution deconvolution methods to laser ARPES data to reveal the presence of what is in many respects a surprisingly conventional signature of electron-phonon coupling in optimally doped Bi2212 at low temperatures. While many interactions have been previously observed to affect the electronic states of Bi2212, none has yet been enumerated without ambiguity and the rancor of disagreement within the community. The origin of the interaction revealed for the first time in this thesis on the other hand is entirely unambiguous and therefor provides a rare moment of clarity in the study of a system that is oftentimes baffling and impenetrable. The second experiment involves the observation of "unoccupied" states in both the normal and superconducting states of optimally and over doped Bi2212. The results of this experiment are found to have profound implications for our understanding of how and why the normal states of the copper oxide superconductors appears to be different from all other known materials.

Of course as the first sentence of this work suggests the copper oxide superconductors have been studied a very long time. As one of the most fertile topics for study in solid state physics this is unlikely to change any time soon. The field is changing rapidly and has been since it began over twenty years ago. The work presented herein is literally the state of the art in photoemission and high temperature superconductivity as it is now. But it is not uncommon to see a result that forces you to rethink your global view of these materials at least several times a year as well as reevaluate what can be achieved with the tools of study at our disposal. Nevertheless it is hoped the work contained within these pages will continue to provide a useful foundation of understanding for at least some time to come.

Chapter 2

Copper Oxide High Temperature Superconductors

2.1 High and Low Temperature Superconductivity

Superconductivity is first and foremost a macroscopic phenomenon. The term describes a state of matter in which a material has zero resistance to the flow of electrical currents and exhibits a perfect diamagnetic response to the application of an external magnetic field called the Meissner effect. These effects were first observed to occur in many metallic elements when they were cooled below a material specific critical transition temperature, T_c, so long as the magnetic field external to the sample H_c was not too high and the current flowing through a sample did not also exceed a critical value J_c. Because elemental metals have not been observed under normal circumstances to achieve a T_c greater than about 10 Kelvin they are often called low temperature superconductors.

Low temperature superonductivity is well described by the Bardeen-Cooper-Schrieffer (BCS) theory [5]. Its signal achievement is to explain how the macroscopic properties enumerated above emerge from the microscopic organization of quantum mechanical objects, electrons and phonons, in metals at low temperatures. At the core of the BCS theory of superconductivity is the idea that at very low temperatures the normal metallic order of states known as a Fermi liquid, which gives a good description of the properties of metals encountered in every day life, becomes unstable to a ground state of even lower energy. That ground state, the superconducting state, is formed when below T_c electrons are bound into *pairs* of a Bosonic character. These pairs are formed by electrons that in the normal metallic state would be characterized as having the complementary states $k_\uparrow$ and $-k_\downarrow$ where k is the crystal momentum and the subscript arrows denote spin up and spin down states, respectively. The macroscopic parameters enumerated above are essentially controlled by the strength of the pairing potential, the amount of energy required to break a pair into two normal electrons. That pairing potential thus produces a gap in the spectrum of *single* particles 2Δ near $T = 0$ that can be observed in the superconducting state. It must always be remembered that in the quantum theory of low temperature superconductivity Cooper pair states are formed from the correlation of pairs of normal electron states.

Until the discovery of high temperature superconductors in non-metallic systems [6] the highest T_c achieved was on the order of 20 K. In every case the mechanism that paired the electrons and yielded a superconducting state was the same: the BCS mechanism in which phonons overscreen the electron-electron interaction, which in turn causes a retarded net attractive ionic potential to arise between electrons of opposite spin and momentum. This is not thought to be the case for the high transition temperature (high T_c) materials discovered in the late 1980's.

High temperature superconductivity in the copper oxides is characterized by the same fundamental macroscopic phenomena – zero electrical resistance and the Meissner effect – as elemental BCS superconductivity. There are several macroscopic differences however. First, the copper oxide materials, or *cuprates*, are *not* metals over an appreciable portion of their doping-temperature phase diagram. Second, the magnitude of the supercurrent that will flow between an identical cuprates depends on the relative orientation of their crystal structures; certain orientations will allow no supercurrent at all to flow between them while others permit the maximum amount. Finally, the macroscopic critical values for temperature, current density and magnetic field can be up to two orders of magnitude greater in the cuprates than in any BCS superconductor. This last fact is the origin of great practical as well as academic interest in the high T_c cuprates. Finally, as the name implies, high temperature superconductivity was until last year unknown outside of materials with copper-oxide layers as their main determining feature. A discussion of the new materials, the iron pnictides, is beyond the scope of this work and too poorly understood at this point to even begin to paint a picture.

In what remains of this chapter we shall briefly review what is known about the microscopic origins of the above listed differences between BCS and cuprate superconductors. We shall first examine the chemical composition and crystal structure of the cuprates. This review is critical as many of the most important aspects of cuprate physics are tied directly to the complex crystal structure of the materials. We shall then introduce the temperature-doping ($T-\delta$) phase diagram of the cuprates and examine what is known about how the electronic structure of the cuprates evolves across it.

2.2 Crystal Structures and the Families of the Cuprates

The defining characteristic of the cuprates is the presence in the unit cell of at at least one and often two or more layers of copper-oxygen planes. The copper-oxygen planes are formed by lattices of squares made up of alternating copper and oxygen atoms with the copper atoms at the corners of the squares. In all cases the cuprates also are comprised of a number of structural components separating the planes from each other and forming a scaffolding in which they can reside easily in two dimensions. These scaffolding components are insulating and do not contribute to the valence bands of the cuprates as observed by photoemission. The crystal structures of the one, two and three layer Bismuth based cuprates [4], of which Bi2212 is the bilayer variety, are diagrammed in Figure 2.1.

As is well known the undoped cuprates that these structures represent do not superconduct. They are in fact antiferromagnetic charge transfer insulators[1]. In the literature they are often referred to as Mott insulators for short and we shall often do the same in this text. That is because while a

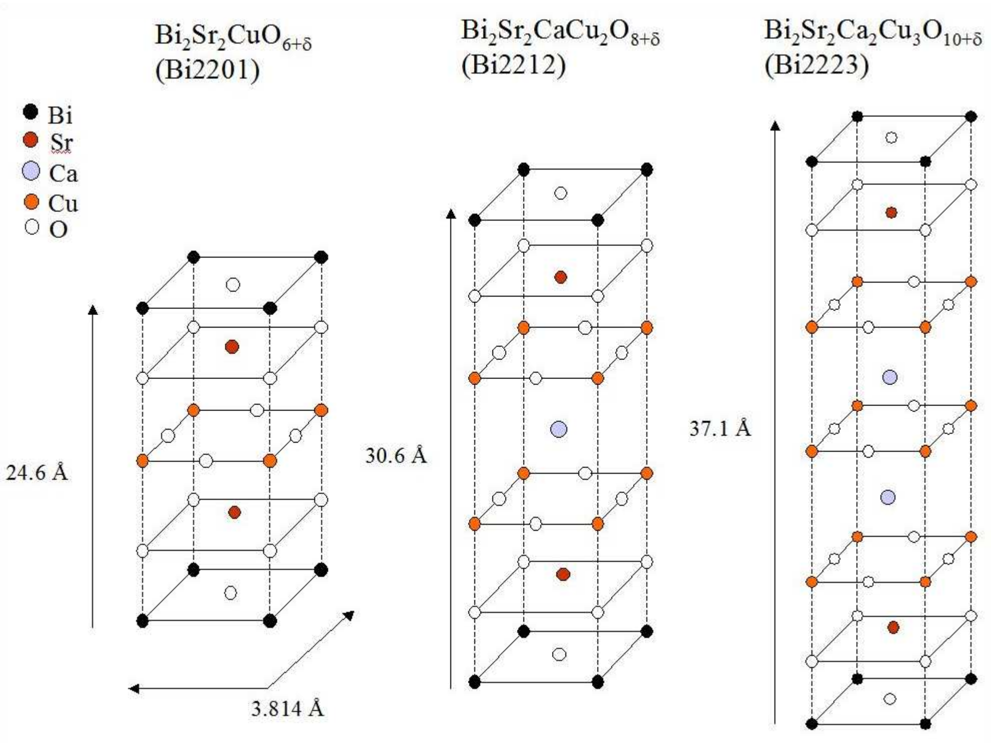


Figure 2.1: Primitive unit cells of the Bismuth based cuprates.

simple counting argument would show these materials to be metals with half filled bands, corresponding to one free electron per copper-oxygen plane per unit cell, strong on site Coulomb repulsion causes extreme electron localization and thus induces an insulating character. One must add additional holes or electrons to the half filled band before these materials do anything useful. Thus the doped materials are often referred to as doped Mott insulators. Doping is accomplished through chemical alteration of the crystal. In several families this doping is accomplished by chemical substitution. In $\mathrm{La}_\delta(\mathrm{Ba,Sr})_{2-\delta}\mathrm{CuO}_4$ (LBCO and LSCO) for example, where δ is the doping fraction, the full range of hole doping levels can be achieved by substituting a fraction of Ba or Sr atoms for La atoms in the undoped compound. In other families of hole doped cuprates such as $\mathrm{YBa}_2\mathrm{Cu}_3\mathrm{O}_{6+\delta}$ (YBCO) and $\mathrm{Bi}_2\mathrm{Sr}_2\mathrm{CaCu}_2\mathrm{O}_{8+\delta}$ (BSCCO or Bi2212), the hole doping is accomplished by adding extra oxygen atoms to the crystal. The doping of YBCO from the undoped parent compound, and the overdoping of the as-grown optimally doped BSCCO is performed by enclosing the samples in high pressure cells of liquid oxygen. Underdoping of BSCCO, the removal of extra oxygen, is accomplished by annealing a sample to several hundred degrees Celsius in a vacuum for a day or two. (For the record, optimal doping refers to the doping concentration with the highest T_c.) In BSCCO the dopant oxygen ions reside above and below the planes interstitial in the lattice. YBCO is more complicated.

The astute reader will no doubt have noticed the lack of any mention so far of the electron doped cuprates. That topic will be touched on below in the context of the cuprate phase diagram but they are for the most part beyond the scope of this work. There is no consensus at this time how they are related to the hole doped cuprates beyond the presence of doped copper oxygen planes.

2.3 The Cuprate and BSCCO Phase Diagrams

There is a phase diagram generic to the cuprates as a whole and there are more detailed phase diagrams for individual families of cuprates [1]. This breakdown is useful because the several families often exhibit behavior with respect to doping pathological to themselves with explanations one might call *local* rather than global. The presence of a layer of CuO chains in YBCO in addition to two planes introduces its own peculiar physics into the problem. A less clear distinction can be made, for example, regarding the static stripe order seen in cuprates at the famous 1/8 doping level as in LBCO and to a lesser extent LSCO. Bi2212 is marked as being a particularly anisotropic material with resistivity ratios between the ab plane and the c axis ρ_{ab}/ρ_c on the order of 1000 suggesting it is a highly two dimensional material. After the global phase diagram we will be most concerned with the Bi2212 phase diagram.

A global phase diagram is shown in Figure 2.2. The salient feature on either side of the zero doping line is the presence of the charge transfer/antiferromagnetic (CT/AFM) insulating state. To get a sense of the energies involved, the temperature scale of this magnetic order at zero doping is on the order of several hundred degrees Celsius. On the hole doped side the CT/AFM is suppressed rapidly by doping. At even very low dopings, up to a few percent, it yields to the ”normal pseudogap” state below a temperature T^*, above the Néel temperature T_N and a “bad metal” or “marginal Fermi liquid” state above T^*. The

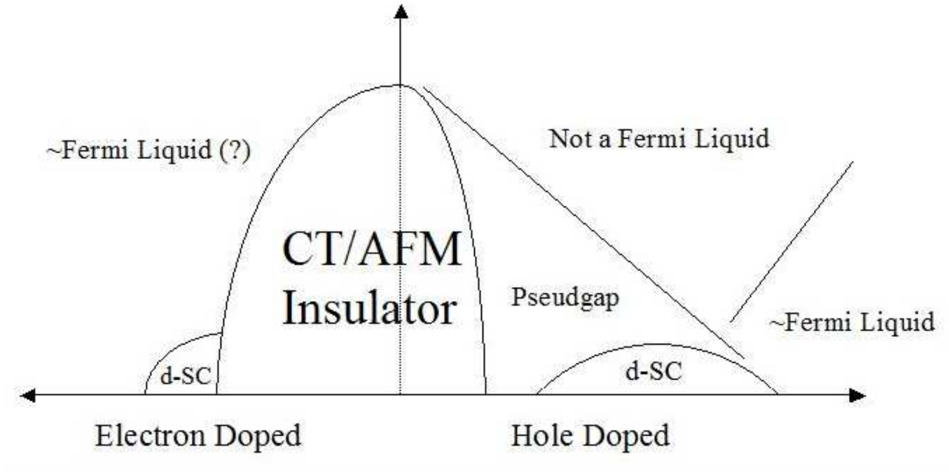


Figure 2.2: Generic phase diagram of the electron and hole doped cuprates adapted from reference [1].

quotation marks are emblematic of our understanding of either of these phases. There is no known counterpart to the pseudogap phase for the electron doped cuprates. On both sides, with sufficient doping, a T_c "dome" appears. On the hole doped side this occurs around 5 percent doping (δ=.05). The doping at which the maximum T_c occurs is said to be "optimal", about 16 percent for the hole doped cuprates. Dopings closer to the zero of the doping axis are said to be underdoped, dopings further away are said to be overdoped. For both hole and electron doped systems the overdoped ends of the phase diagram are characterized by normal state properties approaching those of a normal metallic Fermi liquid with the T^* line either zero or coinciding with the T_c line as it goes to zero on the hole doped side. In addition to the apparent lack of a pseudogap and the extended AFM phase, the electron doped cuprates tend to have much lower maximum T_c.

The current phase diagram of Bi2212 is shown in Figure 2.3 [7]. It rep-

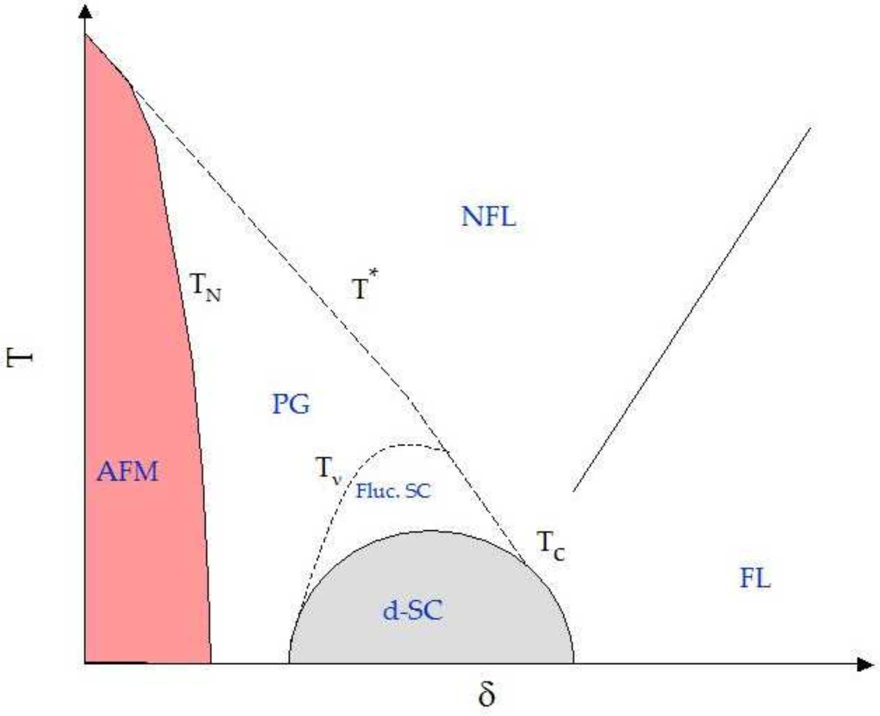


Figure 2.3: Temperature vs. doping δ phase diagram of Bi2212 as it is currently understood.

resents a point of view based on what is known at this time. While it looks complex, Bi2212 actually has a relatively straightforward phase diagram compared to many other cuprates. T_c^{max} of Bi2212 is 91 K, at which point T^* is about 140 K. The line T_ν represents the temperature below which the Nernst effect detects the presence of magnetic flux flow indicating the presence of diamagnetic superconducting fluctuations in the normal pseudogap state [8]. It is not a true phase but nevertheless represents an experimentally important energy scale. It is not certain if T^*, the temperature at which the psuedogap is observed to close, represents a real phase transition or some other phenomenon. It is often referred to as a “crossover” temperature.

2.4 The “Ideal” Electronic Structure of Bi2212

A Mott insulator is a material that is predicted on the basis of band theory to be a metal with the valence band half filled but is in fact an insulator. Insulating behavior arises on the square lattice of the cuprates because when there is one electron per copper atom the repulsive electron-electron Coulomb potential U is poorly screened. The repulsive potential U introduces a large energy barrier for an electron to hop from one site on the square lattice to a neighboring site. This barrier corresponds to an insulating gap in the electronic structure equal to U. Antiferromagnetism sets in because the ground state of the system of localized electrons on the square lattice is is reached by letting the z component of nearest neighbor spins align antiparallel with respect to nearest neighbors. The electrons in this case are said to be strongly correlated. This correlation represents a failure of the single electron picture of metals.

This order is shown on schematically in Figure 2.4a. The cuprates from this point of view are considered doped Mott insulators. By adding extra holes to the square lattice previously localized electrons gain mobility. Because of the underlying antiferromagnetic order, electron hopping to next-nearest neighbor sites is preferred at low to moderate dopings as depicted in Figure 2.4b. In this regime the system has been likened to a disordered spin liquid in which conduction is better described by a spin moving though a sea of spins than a charge moving through a sea of like charges as in a traditional metal [9]. At high enough hole concentrations the underlying AFM order is broken up enough, and the Coulomb repulsion U is screened well enough for the system to attain a more metal-like state. Such a state is observed in heavily over doped cuprates. The exact nature of the ground state of the underdoped Mott insulator is in this sense the main theoretical problem of the cuprates. Curiously enough, as more and more holes are doped into the system, the electronic states that arise inside this U gap [10] begin to resemble more closely the naive LDA band structure described below and so its usefulness as a tool is not completely lost.

LDA picture is still spoiled by the presence of, among other things, the pseudogap state. This violation is most grossly manifested by the absence of a Fermi surface in the system as defined by the zero energy surface traced out by the bands in k space. Thus Luttinger's theorem, which relates the volume (or in two dimensions the area) of the Brillouin zone enclosed by the Fermi surface to the number of free carriers in the system, also fails. Very overdoped cuprates, upon approaching a Fermi liquid like state, are thought to be well described by LDA but we shall not delve too deeply into this question as it

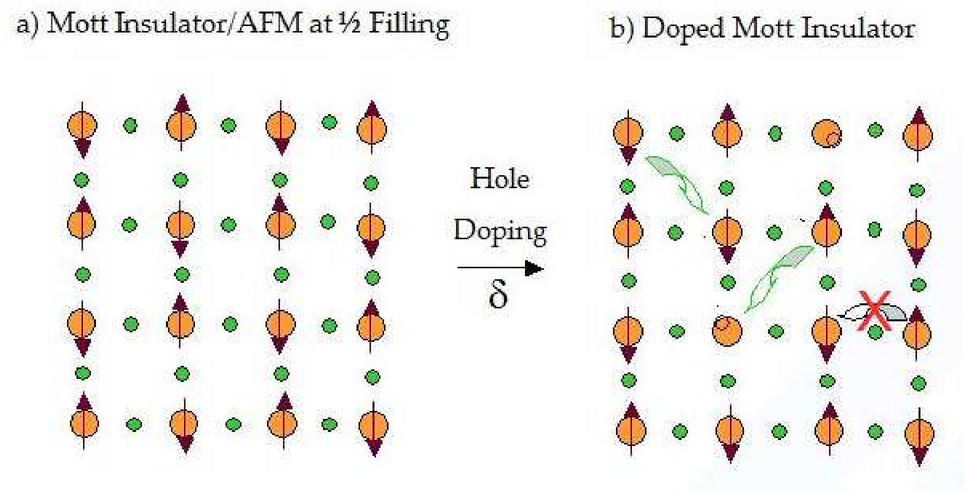


Figure 2.4: Simplified, schematic representation of a) the AFM/Mott insulating state on the copper oxide square lattice. Copper atoms are orange; oxygen atoms are green. Arrows indicate the relative spin orientation. b) upon doping the AFM/Mott insulator nodal conductivity becomes preferred. Hopping along the bond direction is still suppressed by the on site Coulomb repulsion U.

is not settled and will not be addressed much further in this work. While we shall provide some insight into these problems near the end of this thesis it suffices for now to explain the terminology and phenomenology, as currently known, that is used to describe the electronic structure of the cuprates. This is especially important here as ARPES is above all else a tool for the study of electronic structure.

While the tight binding band structure calculations are, strictly speaking, wrong they are "close enough" around optimal doping to provide a road map by which to orient one's self in the Brillouin zone. Throughout this text we shall adopt the language most commonly used in the ARPES literature to describe the electronic states of the cuprates. A common parametrization of the 2D cuprate band structure found in the copper oxygen planes is given by the dispersion relation [11]

$$\varepsilon_k = -2t(\cos(k_xa)+\cos(k_yb))+4t'\cos(k_xa)cos(k_yb)-2t''(\cos(2k_xa)+\cos(2k_yb))-\mu \tag{2.1}$$

where μ is the chemical potential, a and b are the in plane lattice constants and $t = .34$, $t' = .22t$ and $t'' = .1t$ are the nearest, next nearest and next-next nearest neighbor hopping energies. k_x and k_y are the crystal momenta along the a and b axes of the copper-oxygen planes. A term proportional to $\pm t_\perp$, the interlayer coupling, can be added to account for the bilayer splitting in bilayer and trilayer cuprates. The single band picture is often used diagramatically because over most of the Brillouin zone the bilayer split bands are nearly degenerate. For optimal doping ($\delta \sim .16$) $\mu \sim .24$ eV for this parametrization of the band structure. The band structure generated by this dispersion is shown in Figure 2.5. In the cuprates it is common to normalize momentum axes to π so the points $(k_x a, 0)$ and $(0, k_y b)$ are called the $(\pi, 0)$ and $(0, \pi)$ points, respectively. In crystallographic notation they are also called the M points, being the point half way along each square face of the Brillouin zone. The corner point $(k_x a, k_y b)$is called X or Y and denoted (π, π). X and Y are identical in tetragonal cuprates and different in orthorhombic cuprates. The salient features of the gross band structure are the saddle points, often below the chemical potential, at the M points and the high velocity dispersion along the zone diagonals. Cuts through the full dispersion along the high symmetry lines $\Gamma M - MX - X\Gamma$ are plotted in 2.6. This model is qualitatively useful as it captures, for example, the saddle point around M and the nearly linear node crossing the chemical potential just off of $(\pi/2, \pi/2)$. Clearly however all the physics related to electron dressing interactions, superconductivity and the pseudogap, described below, are left out.

The Fermi surface, the locus of electronic momentum states at zero energy in a crystal, is often a far more useful tool for visualizing what part of the band structure is under discussion than the full 3D band structure plotted above. The metallic LDA Fermi surface is shown in Figure 2.7. The Fermi surface crossing on the $(0, 0)$ to (π, π) high symmetry line is called the nodal point. The $(\pi, 0)$ and $(0, \pi)$ points are called the antinodes. The states in the vicinity of these points are generally grouped into the nodal and antinodal *regions* of the Brillouin zone, respectively. Taken altogether the band structure described here gives the impression that the cuprates are technically metals, which they are not. Moving the chemical potential in the above model to a level consistent with half filling of the band will give no indication of the insulating state that

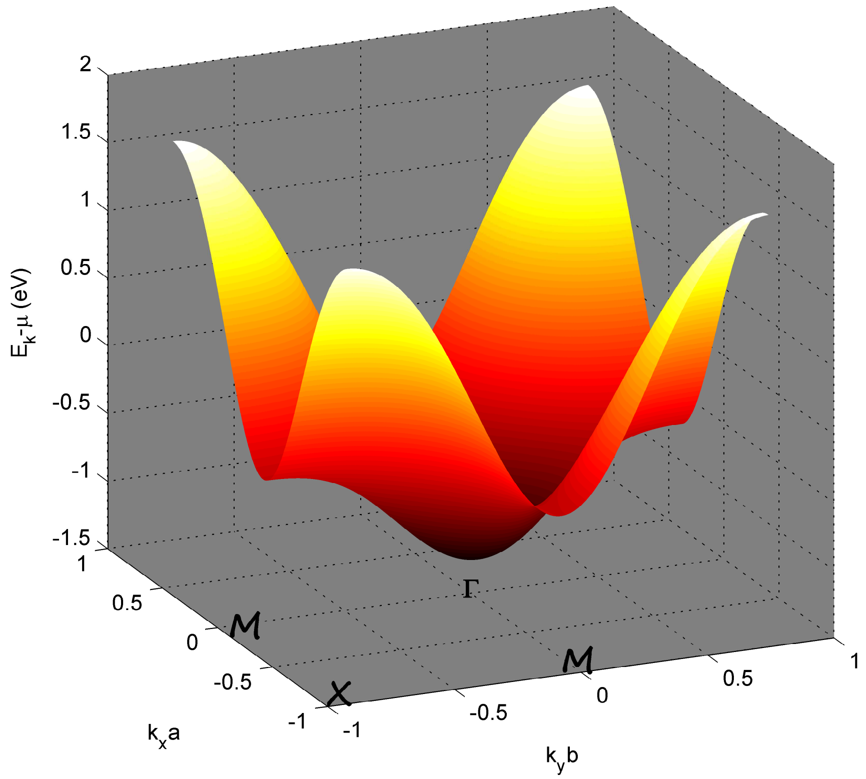


Figure 2.5: Single LDA band approximating the dispersion of the Cu-O $pd\sigma$ band on the square lattice.

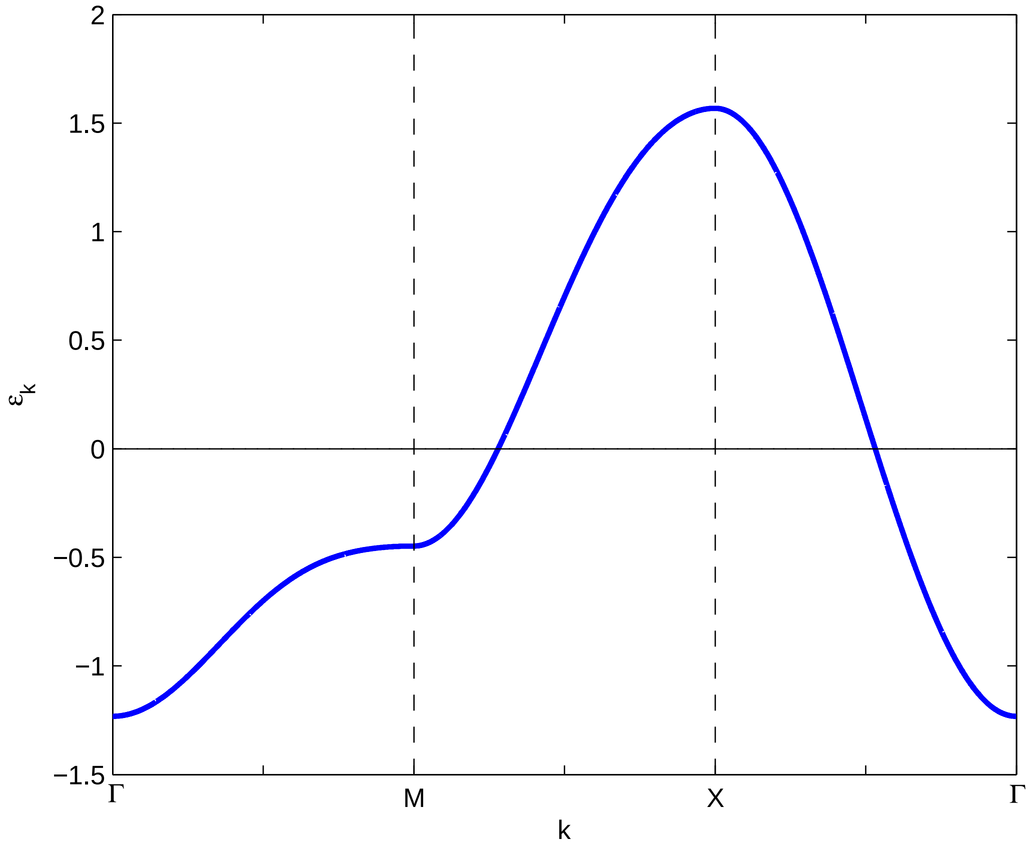


Figure 2.6: Electronic dispersion along the high symmetry lines of the tetragonal cuprate Brillouin zone.

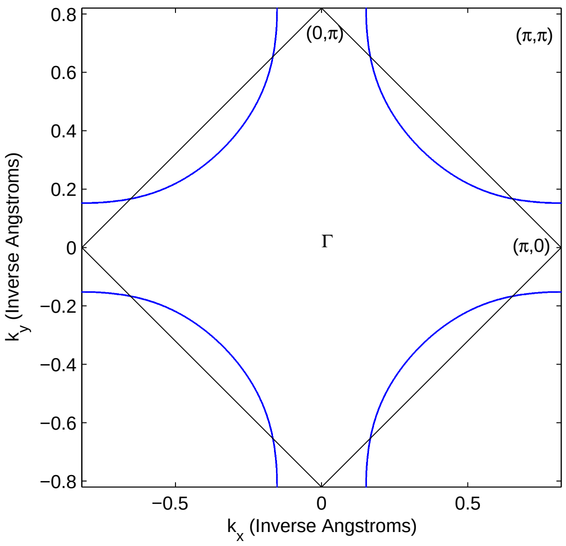


Figure 2.7: LDA Fermi surface of Bi2212 in the single band approximation from Eq. 2.1.

in fact results from electron correlations.

2.5 Gaps and Arcs in the Hole Doped Cuprates

There are many reasons the cuprates are special. We shall focus on three: the presence of the normal pseudogap state below T^*, almost metallic conduction in the absence of a true Fermi surface in this state and high temperature superconductivity with a d-wave order parameter. We shall briefly review how these phenomena manifest themselves in the band structure as determined by ARPES. Our aim is to be more phenomenological than explanatory at this stage.

Bi2212 samples with T_c's that are underdoped, optimally doped and even slightly overdoped exhibit a pseudogap in the normal state below the T^* scale

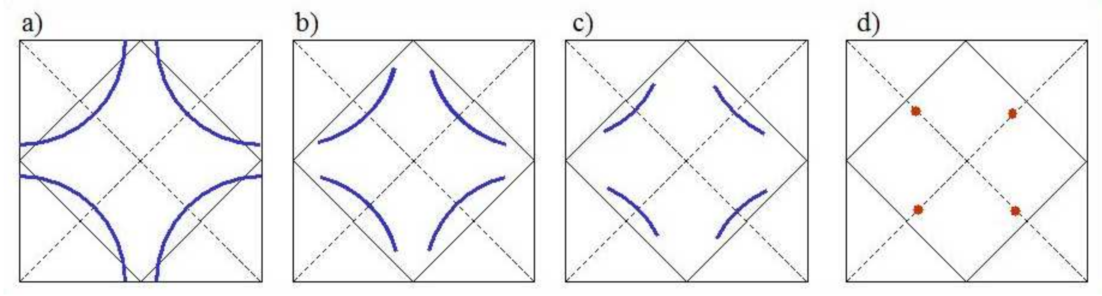


Figure 2.8: a) Normal state Fermi surface of a heavily overdoped cuprate. b) and c) the Fermi arcs for progressively more underdoped cuprates in the normal psuedogap state. d) nodal points in the d-wave superconducting state.

[12]. The pseudogap is observed in ARPES as a depletion in the electronic density of states around the Fermi level in the antinodal region of the Brillouin zone. This depletion, or gap, has an energy scale on the order of T^*, 30 meV at $(\pi, 0)$ in optimally doped Bi2212. The magnitude of this gap at the antinode and its extent in the Brillouin zone *increase* with *decreasing* doping. This behavior runs counter to the trend expected for a superconducting gap which should increase with increasing doping up to T_c^{max}. At the moment it is not clear if the pseudogap and superconducting gap scales cross or merge on the overdoped side of the Bi2212 phase diagram. This is not a question we shall address in this work; we shall be mostly concerned with the phenomenology of optimally doped and underdoped Bi2212. The pseudogap is postulated to result either from the formation of tightly bound, localized, phase incoherent Cooper pairs above T_c, called pre-formed pairs, or from an alternative ordered ground state in competition with or that occurs in some complicated relationship to high temperature superconductivity. We shall present evidence at the end of this thesis in favor of the former scenario which is also supported by

the presence of a strong diamagnetism observed in the normal pseudogap state by Nernst effect measurements on underdoped cuprates. The extent of this diamagnetic signal is illustrated by the T_ν line in the Bi2212 phase diagram in Figure 2.3. The pseudogap has a profound impact on the normal state band structure of the cuprates in that it is responsible for the lack of a Fermi surface in the normal state. The increase in extent as well as magnitude of the pseudogap with underdoping on the Fermi surface is illustrated schematically in Figure 2.8, panels a through c. The residual pieces of Fermi surface in the nodal region of the zone are referred to as Fermi arcs. An extension of the Fermi arc phenomenology and a subsequent interpretation of these arcs as the visible part of a unique kind of "Fermi pocket" will be presented in Chapter 8. For now it suffices to point out that these arcs are responsible for the quasi metallic conductivity of the underdoped and optimally doped cuprates. Both these phenomena, the pseudogap and the Fermi arcs, are unique characteristics of the hole doped cuprates.

The last phenomenological point to be covered here is the presence of high temperature d-wave superconductivity in the cuprates. The origin of d-wave superconductivity at elevated temperatures in the cuprates remains in dispute and is one of the great unsolved problems in solid state physics. Here d-wave refers to the symmetry of the superconducting order parameter and describes how both the magnitude of the superconducting gap and the phase of the Cooper pair condensate evolves in reciprocal and real space. For comparison we sketch generic s-wave, p-wave and d-wave order parameters in Figure 2.9. BCS superconductors such as elemental metals all exhibit s-wave superconductivity. Recently it has been shown that the Fe pnictide superconductors likely

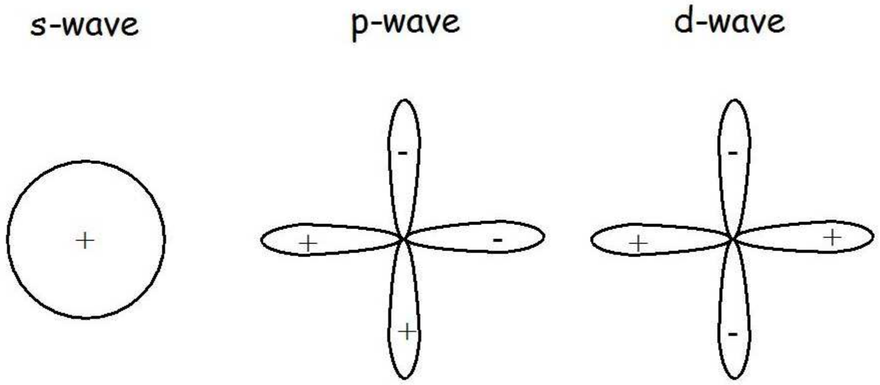


Figure 2.9: Schematic representations of the s-wave, p-wave and d-wave order parameters. The + and - labels denote relative phases between lobes of p- and d-wave type superconductors. The distance from the origin in all three cases denotes relative gap magnitudes.

also exhibit a peculiar form of s-wave symmetry, though the pairing mechanism is more likely magnetic in origin than due to the BCS phonon interaction. The only known p-wave superconductor is $\mathrm{SrRuO_4}$ which with a T_c^{max} of about 1.5 K for the undoped compound is difficult to study with ARPES below T_c. The known d-wave superconductors include the heavy fermion systems and the cuprates. Both p- and d-wave superconductors exhibit points of zero gap magnitude called *nodes*. In cuprates these nodes occur at the nodal point of the ideal Fermi surface along the $(0, 0)$ to (π, π) symmetry lines which in real space are aligned along the diagonals of the Copper-Oxide plane plaquettes. This is illustrated in panel d of Figure 2.8. The gap maximum occurs in the anti-nodal direction and implies the pairing in the cuprates is strongest along the copper-oxygen bonding direction in real space. The main difference between the p-wave and d-wave order parameters is the relative phases exhibited by the different lobes of the order parameter. This phase pattern causes p-wave superconductors to break time reversal invariance whereas d-wave superconductors are thought to preserve this symmetry. There exist other, more exotic types of order parameter but no firm evidence of their existence in a real material has been decisively confirmed. Note that photoemission can probe the magnitude of the superconducting gap but not the phase of the Cooper pair condensate.

Because the ideal cuprate Fermi surface is in most cases well approximated by a circle centered at (π, π) the location of a point on this surface is often parameterized by an angle, the Fermi surface angle, ϕ. ϕ is set to zero either at the node or the antinode depending upon which parametrization is more convenient for a given context. In terms of ϕ the magnitude of the ideal d-wave gap is takes the form

$$\Delta(\phi) = \Delta_0 |\sin(2\phi)| \tag{2.2}$$

where ϕ, which lies in the range $-\pi/4 < \phi < \pi/4$, is taken to be zero at the node and Δ_0 is the gap maximum at the antinode. This form gives the gap over one quarter of the Brillouin zone. While there is accumulating evidence[13] that the superconducting condensate that carries a supercurrent is formed by the pairing of electrons appearing only on the Fermi arc in the normal pseudogap state the order parameter is well settled and this physics is easily captured, if true, by a trivial redefinition of Δ_0 and the range of ϕ. In the simplest case where we assume the order parameter extends to the antinodes of the Brillouin zone this gap function appears as plotted in Figure 2.10 for a Δ_0 of 25 meV. Much debate in recent years has centered on whether

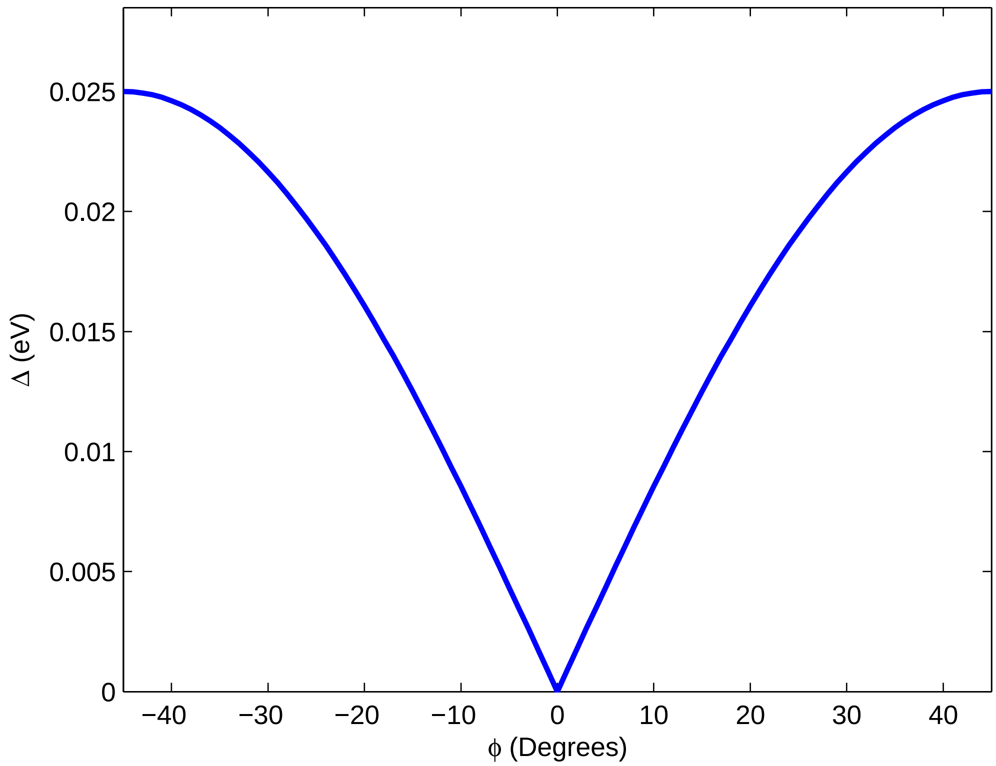


Figure 2.10: Ideal d-wave superconducting gap as a function of Fermi surface angle ϕ from Eq. 2.2. Here we take the node as the zero angle.

or not the gap near the node takes the "v" shape characteristic of the pure d-wave shown in the figure or contains additonal angular harmonics that would indicate the presence of, for example, more than one pairing channel. Such additional physics is thought to give more of a "u" shape to the gap function near the node.

2.6 Electron-Boson Interactions

No discussion of the phenomenology of the electronic structure of the cuprates would be complete without reference to the deviations from the LDA dispersion observed in the band structure of Bi2212. These deviations from simple linear or parabolic dispersion appear as "kinks" in the band structure. Where the study of Fermi surfaces and related phenomena is colloquially called "Fermiology" the study of these band structure anomalies can be called "kinkology". Since the inception of high momentum resolution photoemission in the mid 1990's a veritable zoo of kinks have been observed in the band structure of Bi2212. These kinks exhibit a variety of temperature, doping and momentum dependences. The causes and occasionally the very existence of these kinks are often disputed. The identification and examination of a previously unobserved dispersion kink in optimally doped Bi2212 will form the subject of much of Chapter 7. What, then, is a kink and why do we care about whether and how kinks manifest themselves in the electronic structure of Bi2212?

The mass of quasielectrons in a given band of a crystal's electronic structure is associated with the curvature of that band. A sudden change in that dispersion over some limited energy range thus corresponds to a modification of that curvature and thus the effective mass m^* of the carriers associated with that band [14]. There are a variety of possible physical causes for such an effect, the primary one being the onset at a specific energy and possibly momentum of a coupling between band electrons and a bosonic mode, e.g. phonons and magnons, of a well defined energy. Physically, one pictures the bosons participating in such an electron-boson interaction as "dressing" the electron by introducing new scattering channels excitable above the threshold energy for the interaction. Mathematically the interaction introduces new Feynman diagrams into the calculation of the electron self energy. The appearance of a new scattering channel decreases the lifetime of electrons of sufficient energy to excite the bosonic mode causing a sudden change in the scattering rate. The scattering rate, which is twice the imaginary part of the self energy $\Sigma_k(\omega)$, is causally related to the real part of the self energy which describes the dispersion ε_k. Thus a change in the scattering rate also renormalizes the dispersion, producing a "kink" in the band at the mode energy. The most famous example of this phenomenon in Bi2212 is the kink around 70 meV [15] [16] binding energy in the nodal region of the Brillouin zone. An early accounting of the temperature and momentum dependence of this features is given by Figure 2.11 as adapted from reference [17].

In addition to the 70 meV nodal kink, dubbed here the "classic" kink, there are thought to be at least one [18] [19] and possibly more [20] kinks at higher binding energy in the nodal region as well as a kink around 40 meV in the antinodal region. This is in addition to the newly discovered 8 meV kink in the nodal region discussed in this thesis. We shall argue that while the assignment of a cause for these various kinks is not clear the 8 meV assignment to a particular optical phonon mode for the first time is. The overall situation of nodal kinks is diagrammed in Figure 2.12.

There exist a few methods for examining kink physics. The "direct" method as illustrated above involves measuring the imaginary part of the self energy, i.e. the scattering rate, while simultaneously extracting the real part

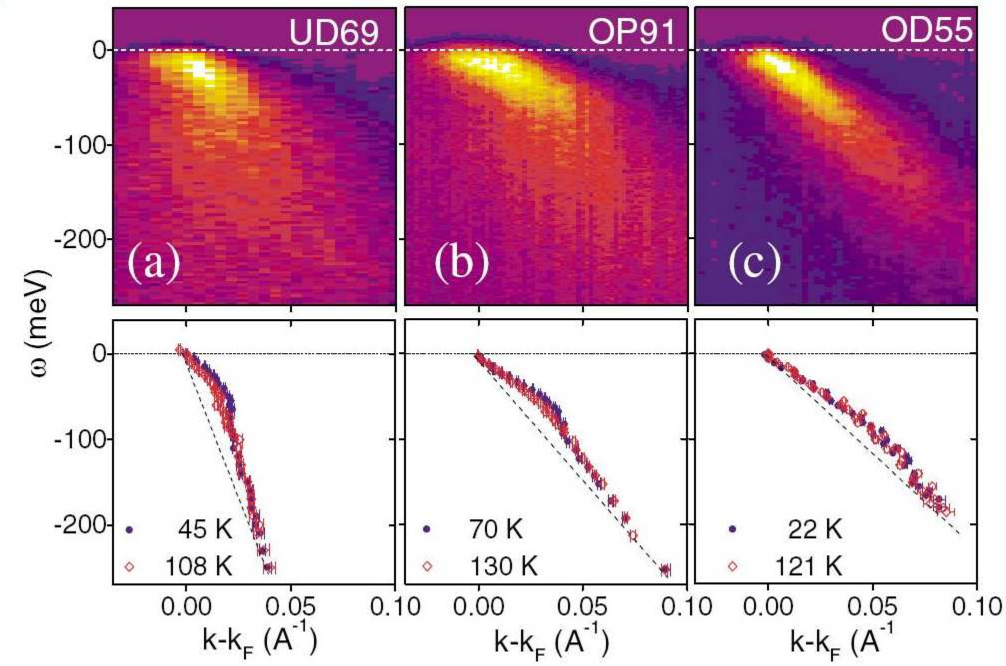


Figure 2.11: The classic kink in Bi2212. Panels a)-c) show the nodal spectrum of under doped, optimally doped and over doped Bi2212, respectively, with T_c indicated. Lower panels show the MDC derived dispersions above and below T_c.

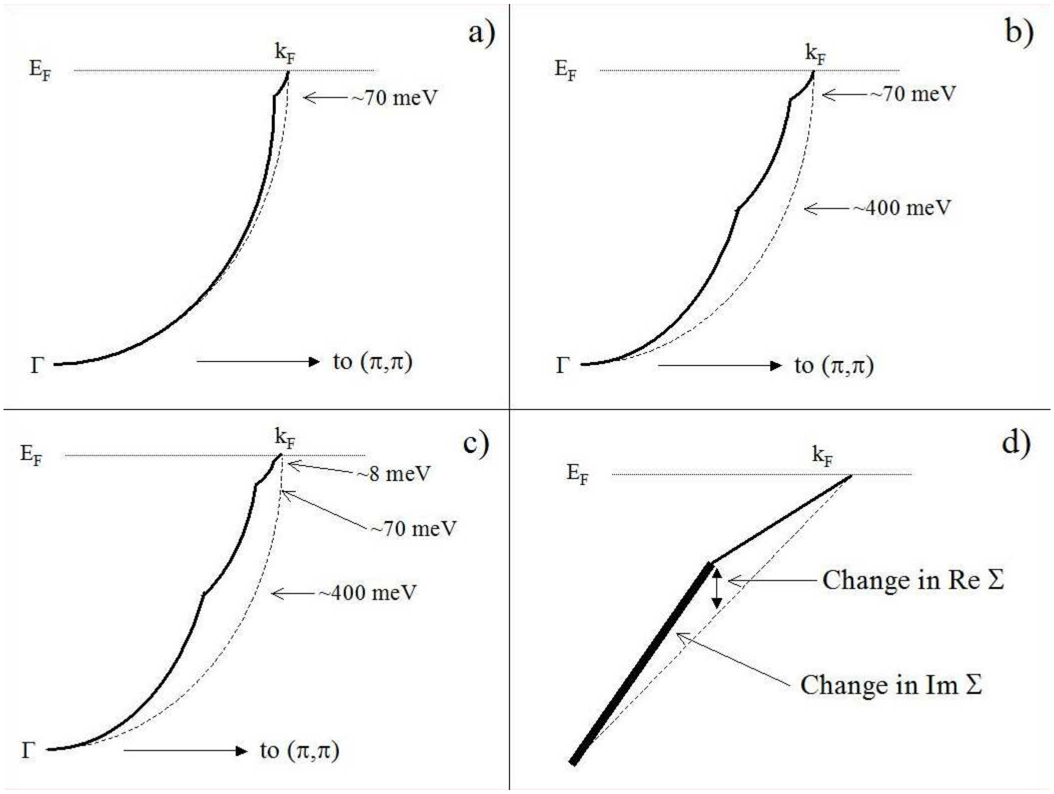


Figure 2.12: Historical evolution of the nodal kinkology of Bi2212. In all panels the dashed line denotes the supposed unrenormalized, or bare, LDA-type band used to directly extract the change in the real part of the self energy induced by the interaction. a) the picture used in the analysis of Figure 2.11 b) the same picture modified after observations proposing the existence of a "big" kink at 350 meV. c) the same picture now including the 8 meV kink discussed in this thesis. d) Generic anatomy of a kink showing the deviation of dispersion from the bare band with the subsequent broadening of the electron lifetime above the kink energy. Note that the existence of the big kink remains a topic of debate.

by calculating the change in band dispersion observed relative to a theoretically derived bare band dispersion. In practice if there is only one kink and the band appears unperturbed far from the kink energy one simply draws a line of the type shown in Figures 2.12 and 2.11 (in the lower panels) as dashed lines for the bare band and extracts $\mathrm{Re}\,\Sigma$ as the difference in binding energies between the measured and bare bands. If this procedure is justified the real and imaginary parts of the self energy thus extracted should be causally related by Kramers-Kronig (KK) relations [21]. Because the bare band is something of a fiction in the best case and for correlated electron systems like Bi2212 probably a total fabrication (in the sense there is no such thing as a bare electron) it is often preferable to only measure the imaginary part of the self energy and then use the KK relations to calculate the real part. This method suffers somewhat from the ambiguous location of the ultraviolet cutoff of this integral but this ambiguity is often less dire than that involving a postulated bare band. Either way the reason for extracting the real part of the self energy is that it can be directly related to the electron-boson coupling constant embodied by the mass enhancement factor λ such that [15]

$$-\frac{\partial\,\mathrm{Re}\,\Sigma}{\partial\omega} = \lambda \tag{2.3}$$

close to the Fermi level. This interpretation of λ in the high energy, multiple kink picture presented in Figure 2.12c clearly has some conceptual problems. Alternatively λ can be calculated in the Migdal-Eliashberg formalism if one knows both the boson density of states $F(\omega)$ and the electron-boson coupling α^2. (Migdal-Eliashberg equations involve integrals over $\alpha^2F(\omega)$.) While in principle this information can be garnered from an inversion of the ARPES derived scattering rates ($\mathrm{Im}\,\Sigma$) this is in general a very difficult problem in itself. Because the 8 meV kink discussed later can be reasonably assigned to single mode it is a straightforward matter to derive the relevant part of $\alpha^2F(\omega)$ and thus λ and $\mathrm{Re}\,\Sigma$ from the measurement.

The main imperative for identifying the origin of these interactions in Bi2212 lies in their supposed relevance to high temperature superconductivity. The reasoning goes that if a strong electron-phonon interaction in BCS superconductors is the mechanism responsible for electron pairing in those materials then an analogous electron-boson interaction may produce high T_c's in the cuprates. The universal appearance of the 70 meV in all hole doped cuprates, for example, may result from that interaction and, so the hypethesis goes, the study of this or other such kinks may eventually yield the boson, if any, and thus the pairing mechanism responsible for high temperature superconductivity. Thus if superconductivity in the cuprates is driven by an electron-boson interaction analogous to the electron-phonon interaction of BCS theory finding and then identifying electron-boson interactions in the cuprates is potentially of paramount importance in the problem of the high T_c's. Though the possibility of such a mechanism for high T_c superconductivity in the cuprates is thought likely by only a minority of the community it has yet to be ruled out entirely and thus anything further that can be learned about these processes provides a valuable addition to our ability to ultimately discern the correct theory of high T_c.

Chapter 3

Photoemission Spectroscopy

This chapter is devoted to a description of the kinematic and quantum theories of ARPES as it is currently understood. A discussion of the particular experimental methods required to perform a laser ARPES experiment will be presented in a later chapter. While the treatment here is essentially canonical, being thoroughly covered in several well known texts [22][1], special attention will be paid to the ramifications of using “low energy” light, especially lasers, as photoemission excitation sources. As we shall see, while the use of lasers offers many potential benefits for ARPES on certain condensed matter systems of great interest, even under the best of circumstances the interpretation of the resulting data must be treated with great care.

3.1 Kinematics of Photoemission

ARPES is the modern application of the photoelectric effect famous to all students of physics as the subject of Einstein’s Nobel Prize winning use of the then new quantum theory of matter and energy. It was discovered at the end of the 19th century that a clean metal surface in a vacuum vessel emitted electrons when light was shone upon it. The energy of electrons escaping the surface of the metal depended not upon the intensity of the light (which only governed the number of electrons emitted) but upon the wavelength or frequency of the light. The number of electrons was determined by measuring the (photo)current between the photocathode and the anode target and the kinetic energy of the emitted electrons was measured by finding the reverse bias voltage applied between the anode and cathode required to drive the photocurrent to zero. A schematic of this early type of photoemission experiment is shown in Fig. 3.1. The relationship between this stopping potential (and therefore the electron kinetic energy) and photon wavelength forms the fundamental kinematic relation for photoemission,

$$eV = h\nu - \phi = \frac{1}{2}m\boldsymbol{v}^2 \tag{3.1}$$

where e is the electric charge, h is Planck's constant, ν is the photon frequency and ϕ is the work function of the sample surface, about which we shall have more to say shortly. Historically this experiment was meant to yield only a measurement of Planck's constant, h, and show that it is a universal constant.

While this version of photoemission implied a great deal about the quantum nature of electrons and their interaction with light it does not tell us much about the metal used as the photocathode other than to say each metal has a unique work function, ϕ. Eventually it was realized that one could use

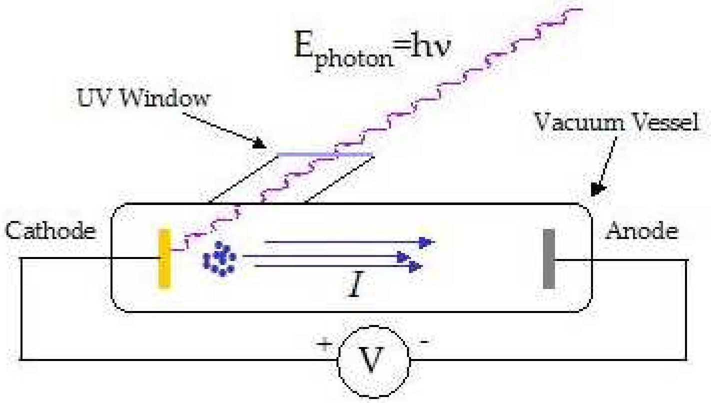


Figure 3.1: Schematic of the original photoelectric effect experiment. Light with sufficient energy $h\nu$ ejects electrons from a cathode. The work function ϕ is determined from the stopping potential V applied between the anode and cathode.

the photoelectric effect as a tool to probe the electronic structure of the photocathode. This early version of the technique, PhotoEmission Spectroscopy (PES), was designed in most cases to measure the photocurrent for a fixed photon energy as a function electron kinetic energy. This measurement yields a number proportional to the density of occupied electronic states in the sample as a function of binding energy. In a modern notation this is expressed as,

$$I(KE) \propto N(E_B)f(E_B) \tag{3.2}$$

where $I(KE)$ is the photocurrent as a function of the kinetic energy KE of the photoelectrons, $f(E_B)$ is the Fermi-Dirac distribution function and $N(E_B)$ is the joint density of electronic states in the sample as a function binding energy E_B measured relative to the chemical potential of the sample (μ). This quantity is similar to the joint density of states probed in tunneling experiments such as STM/STS and optical conductivity. The essential difference lies in the nature of the final states probed in the two types of experiment. Because in solid state physics the chemical potential is almost always equal to the Fermi energy level E_F, and because the Fermi levels of all systems in electrical contact must align, it is the natural zero reference for the binding energy scale of solids. Thus the fundamental kinematic relation of photoemission is usually expressed as

$$KE = h\nu - \phi - E_F + E_B \tag{3.3}$$

where, following standard practice in the literature, we have chosen E_B to be negative for state below the Fermi level (the occupied side) and positive for states above the Fermi level (unoccupied side) so that the Fermi factor is written,

$$f(E_B) = \frac{1}{1 + e^{\frac{E_B}{k_B T}}} \tag{3.4}$$

where k_B is Boltzmann's constant and T is the sample temperature in Kelvin. Note that while the Fermi function enforces the fact that ARPES can only probe occupied electronic states, at finite temperatures this cutoff between states above and below the Fermi level is not necessarily sharp, a fact we shall exploit later to apply photoemission to the measurement of low lying "unoccupied" states.

A measurement of the density of states as a function of binding energy is called an Energy Distribution Curve, or EDC. The kinematics of the simple photoemission process are diagrammed in Fig. 3.2. In the modern formalism of quantum mechanics it is clear that the work function ϕ is the minimum energy an electron requires to escape the influence of the sample into the continuum of final states in the vacuum. This vacuum level, denoted E_v, is a material specific property. Because ϕ is an immutable property of materials themselves a bias voltage applied between the sample and the detector adds a constant energy shift to the entire spectrum, in which case the difference in energy between the shifted E_F and the low energy cutoff in the photoemission spectrum is equal to its work function. Notice too that the analyzer, as a material different from the sample, has a different work function, ϕ_A, that must be accounted for in a real measurement. ϕ_A shows up in PES spectra as a constant energy shift in the spectrum due to a higher or lower barrier between the vacuum and the analyzer than between the vacuum and the sample. It can only be accounted for by knowing the photon energy used and the work function of the sample in question. Thus

$$KE_{meas} = h\nu - (\phi - \phi_A) + E_B - E_F \tag{3.5}$$

E_F itself is easily measured by fitting a Fermi distribution function, Eq. 3.4 to the part of the photoemission spectrum with the highest kinetic energy (e.g. binding energy equal to zero at zero temperature). In the figure we distinguish between three types of electronic states: unoccupied valence states, occupied valence states and occupied core level states. Electrons that make up core level states, which do not generally participate in chemical bonding, are roughly speaking the solid state equivalent of free atomic electron levels. Because in high resolution photoemission we are primarily concerned with the interactions and dispersions affecting the valence states we note the presence of core levels here only in passing. Clearly access to these electrons require

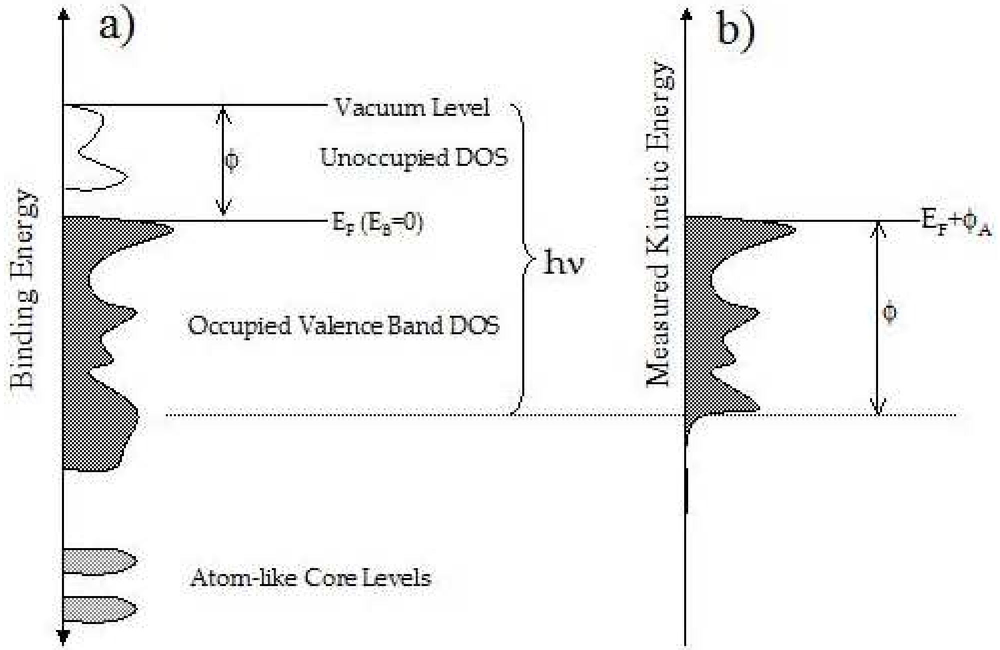


Figure 3.2: Energy level diagram relating density of states in a sample to what is measured in a photoelectron spectrometer.

a somewhat higher photon energy to excite out of a sample than do valence electrons and so are not normally accessible (or interesting) in a low energy ARPES experiment.

While PES gives invaluable information about the total electronic density of states of a system it was ultimately recognized that much more valuable information could be gleaned from photoemission were one able to resolve the photocurrent not only as a function of kinetic energy but also as a function angle of emission from the surface. The technique of angle resolved PES, ARPES from here on out, rests on the fact that electrons inhabiting the dispersing bands of a single crystal have a well defined momentum within that crystal and therefore, due to conservation of linear momentum, should retain that information by escaping the crystal surface in preferred directions. By measuring EDC's at many well defined angles of emission one could effectively reconstruct the band structure of the crystal by observing the peak photocurrent's dispersion as a function of kinetic energy and emission angle, so long as one could relate that emission angle to the crystal momentum $\hbar\boldsymbol{k}$ the electron had prior to photoexcitation.

In general this relation of emission angle to crystal momentum is a very difficult problem. The reason is that electrons, having a wavelike nature, refract (and reflect) at the crystal surface/vacuum interface in a manner analogous to the refraction and reflection of light upon traversing an interface between two regions of differing index of refraction. In the case of a solid the interface is effectively defined by the presence of the energy barrier at the surface which has both an energy component, the work function, and a spatial component defining the distance from the surface over which this potential operates. Physically this barrier derives from the atomic orbitals exposed at a crystal surface. As for light, the wavelike nature of the electron implies that it's wavefunction must satisfy certain boundary conditions at the crystal/vacuum interface. In terms of the electron momentum this condition is,

$$\hbar\boldsymbol{k}_{in} \cdot \hat{\boldsymbol{n}} = \hbar\boldsymbol{k}_{out} \cdot \hat{\boldsymbol{n}} \tag{3.6}$$

where we ignore the reflected wave because it is unobservable in photoemission and $\hat{n}$ is taken to be evaluated at the crystal surface/vacuum interface. In ARPES it is particularly convenient to break the wavevector down into components parallel and perpendicular to the surface, $\boldsymbol{k} = \boldsymbol{k}_{\parallel} + k_{\perp}$. As in electrodynamics this results in the important rule for the parallel component of the photoelectron momentum,

$$\boldsymbol{k}_{\parallel,in} = \boldsymbol{k}_{\parallel,out} \tag{3.7}$$

which is to say that the plane parallel momentum of the photoelectron is strictly conserved in the photoemission process. For photoelectrons originating from inherently two dimensional Bloch states this equation is exact, a result that is the basis and reason for the understandability of high resolution ARPES experiments performed on essentially 2D, layered electronic systems like the cuprates. The conservation of parallel momentum leads to a form of Snell's law for ARPES [23],

$$\hbar\boldsymbol{k}_{\parallel} = \boldsymbol{p}_{\parallel} = \sqrt{2mE_{kin}}\{\sin(\theta)\sin(\varphi)\hat{\boldsymbol{k}}_{\boldsymbol{x}} + \sin(\theta)\cos(\varphi)\hat{\boldsymbol{k}}_{\boldsymbol{y}}\} \tag{3.8}$$

where the polar and azimuthal angles of emission θ and φ are defined relative to the sample normal $\hat{\boldsymbol{n}}$ as in Fig. 3.3 and m is the rest mass of the electron. This relation is often simplified by noting that the quantity $\sqrt{\frac{2m}{\hbar^2}}$ is a "universal constant of photoemission" equal to .5124 $[\mathring{A}^{-1}/\sqrt{eV}]$. From now on we'll just call this C. Also, because φ is often small we can take $\cos(\varphi) \approx 1$ and so Eq. 3.8 is usually written in the scalar form

$$k_{\parallel} = C\sqrt{h\nu - \phi - E_F + E_B}\sin(\theta) \tag{3.9}$$

where we have written the photoelectron kinetic energy explicitly using Eq. 3.3. Note that we here on out omit reference to ϕ_A as it is extrinsic to the

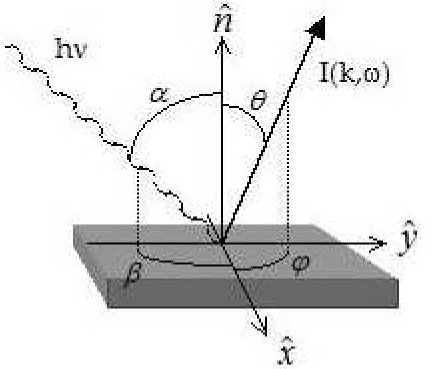


Figure 3.3: All angles parameterizing an ARPES experiment are measured relative to the sample normal $\hat{n}$. (α, β) and (θ, φ) are the azimuthal and polar angles of the incident photon and exiting photoelectron, resepctively.

photoemission process as such. Here the role of the “index of refraction” is played by the prefactor $C\sqrt{KineticEnergy}$. Practically speaking this is the most important equation in ARPES because it tells us how to relate the things we can measure, i.e. the various energies of the problem and the emission angle defined in our detector by, say, a slit or pinhole, to that electron’s origin in the reciprocal space of the crystal. It also yields a simple definition for the momentum resolution in ARPES,

$$\Delta k_{\parallel} = C\sqrt{h\nu - \phi - E_F + E_B}\cos(\theta)\Delta\theta. \tag{3.10}$$

There are three more contributions to the momentum of photoelectrons in ARPES. The first, the contribution to the momentum from the exciting photon, is easily dealt with in low energy ARPES. In the UV and VUV regimes, the photons themselves carry negligible total momentum $p = E/c$. This is not the case as one scales up in photon energy to x-ray wavelengths where, in the XPS regime (x-ray photoemission spectroscopy) one must take into consideration finite momentum q transitions in the spectrum. Such effects are of no concern to us here. An effect that may be of more concern as the quantitative precision of photoemission is pushed to higher scales and better energy resolutions is that of the photorecoil effect. In its simplest incarnation this effect is understood as a consequence of the conservation of linear momentum between a photoemitted electron and the ionic core from which it was previously bound. The recoil of the ionic core induces a shift in the kinetic energy of the electron that can be estimated as

$$\Delta E = KE\frac{m}{M} \tag{3.11}$$

where ΔE is the recoil energy imparted to an atom at rest, KE is the photoelectron kinetic energy, m is the electron mass and M is the atomic mass [24]. ΔE is observed as an apparent shift to higher binding energy of the emitted electron. While this effect is obviously most noticeable at high kinetic energies, it has recently been observed experimentally as a shift in the Fermi level of gold and aluminum on copper. For low energy ARPES this shift is negligible, on the order of .1 to .01 meV depending on the photon energy and atomic mass, and so is only mentioned here for the sake of completeness. Nevertheless, as the technique pushes the bounds of precision this effect may one day need to be taken account of.

A more important problem comes from considering the effects of having a component of crystal momentum perpendicular to the surface, $k_\perp$. Because

$k_\perp$ can itself take on an imaginary component, $k_\perp = k_1 + \imath k_2$, where k_1 and k_2 are real scalar quantities, there is a damping component associated with the propagation of an electron wave originating from a three dimensional state through the crystal surface. The magnitude of this damping is often very difficult to determine precisely in experiments. The perpendicular component of the momentum measured in ARPES is therefore not conserved and as such is broadened significantly in most measurements. This can be seen schematically by approximating the electron wavefunction as a plane wave at a crystal surface so that

$$\psi_\perp \; e^{\imath k_\perp r} = e^{\imath k_1 r} e^{-k_2 r} \tag{3.12}$$

where $\boldsymbol{r}$ is the normal coordinate from the surface. What's more, as will be encountered below, this damping of the photoelectron, after dissociation from the remaining photohole, leads to the measured lifetime of a 3D electron becoming the superposition of the photohole and photoelectron lifetimes whereas for $D < 3$ the measured lifetime only reflects the decay of the photohole in response to scattering events occurring in the remaining N-1 electron system in the crystal. The practical effect of this problem is to render the results of ARPES on one and two dimensional systems vastly simpler to interpret than on three dimensional systems. This has led to a natural predisposition in the ARPES community in recent years to study intrinsically one and two dimensional states, e.g. in layered systems like the cuprates and dichalcogenides or the surface states of metal crystals, as opposed to more general 3D systems. In light of the realization that the physics of one and two dimensional systems is often much richer than that of 3D systems this has not presented much of a handicap to the technique. In any event, we will from here on out confine ourselves to the study of ARPES in systems of reduced dimensionality except where necessary.

3.2 Quantum Theory of Photoemission

What have been described so far are the purely kinematic relations used to link the various quantities measured in ARPES, e.g. electron kinetic energy and emission angle, to the quantities of interest in a solid sample, binding energy and crystal momentum. This analysis does not tell us anything about the quantum mechanical nature of the photoemission process. The quantum theory of ARPES tells us how to interpret what we are measuring whereas the kinematics only tell us how to measure it. The quantum theory of ARPES is typically broken down into two flavors called the three step and the one step models, respectively [22]. The three step model, described in more detail below, consists of picturing the photoemission process as i) photodissociation of an electron from the background sea of electrons in the crystal by an impinging photon ii) propagation of the photoelectron to the surface and iii) emission of that electron into the vacuum while the remaining N-1 electron system relaxes by annihilating the residual photohole. The one-step model on the other hand makes use of the sudden approximation in which a photoelectron is excited directly from an occupied state in the crystal to a final state in the vacuum followed by relaxation of the N-1 electron system. The approximation made in this case is that the photoelectron does not interact with the residual system before exiting the crystal or, put another way, that the decay of the photohole occurs on a time scale much longer than the escape time of the photoelectron. In what follows we assume a system with dimensionality less than three.

The photocurrent $I(k,\omega)$ (where $\hbar\omega$ is the electron binding energy; recall we use $h\nu$ as the photon energy) is directly proportional to the transition matrix element from some N electron initial state (in the sample) to some $N-1$ final state in the sample along with the state of the photoelectron removed to create it. This transition is calculated according to Fermi's Golden Rule such that

$$I(k,\omega) \propto \frac{2\pi}{\hbar} |\langle f| \hat{H}_{int} |i\rangle|^2 f(\omega)\delta(E_f - E_i - h\nu)\delta(\boldsymbol{k}_{\|,i} - \boldsymbol{k}_{\|,f} - \boldsymbol{G}) \tag{3.13}$$

where $\langle f|$ and $|i\rangle$ are the initial and final state wavefunctions, respectively, $\hat{H}_{int}$ is the Hamiltonian operator connecting the two states and $\boldsymbol{G}$ is a reciprocal lattice vector. The Fermi function $f(\omega)$ enforces the fact we can only probe occupied electronic states. The delta functions as usual enforce energy and momentum conservation. Because we are here focusing on low dimensional systems for which $k_\perp = 0$ we have dropped that delta function. For photon energies with insufficient momentum to excite q transitions beyond the first Brillouin zone, as is the case for the low energy laser, $\boldsymbol{G}$ becomes irrelevant and so can be dropped so long as we are working in the reduced zone scheme.

In photoemission the interaction Hamiltonian $\hat{H}_{int}$ is the electromagnetic dipole operator introduced by the photon source. It is formed by modifying the electronic momentum to reflect the electron's coupling to an electromagnetic field,

$$\hat{\boldsymbol{p}} \to \hat{\boldsymbol{p}} + \frac{e}{mc}\boldsymbol{A} \tag{3.14}$$

so that the Hamiltonian becomes

$$\hat{H}_{int} = \frac{1}{2m}(\hat{\boldsymbol{p}} + \frac{e}{c}\boldsymbol{A})^2 - eV. \tag{3.15}$$

After expanding the square, keeping in mind the noncommutativity of $\hat{\boldsymbol{p}}$ and $\boldsymbol{A}$ (which can be written in terms of harmonic oscillator creation and annihilation operators) this becomes

$$\hat{H}_{int} = \frac{e}{2mc}(\boldsymbol{A} \cdot \hat{\boldsymbol{p}} + \hat{\boldsymbol{p}} \cdot \boldsymbol{A}) - eV + (\frac{e}{mc})^2 \boldsymbol{A} \cdot \boldsymbol{A} \tag{3.16}$$

where $\boldsymbol{A}$ is the magnetic vector potential, $\hat{\boldsymbol{p}}$ is the canonical momentum operator $-\imath\hbar\nabla$, V is (here) the scalar electric potential, e and m are again the electron charge and mass, and c is the speed of light. This is simplified by working in the Coulomb gauge so $V = 0$. The first term is simplified by commuting $\hat{\boldsymbol{p}}$ through $\boldsymbol{A}$ yielding

$$\frac{e}{2mc}(\boldsymbol{A} \cdot \hat{\boldsymbol{p}} + \hat{\boldsymbol{p}} \cdot \boldsymbol{A}) = \frac{e}{2mc}(2\boldsymbol{A} \cdot \hat{\boldsymbol{p}} + \imath\hbar\nabla \cdot \boldsymbol{A}). \tag{3.17}$$

Assuming that the divergence of the photon field is negligible over the region near the surface photoemission occurs (another point we shall return to later) $\nabla \cdot \boldsymbol{A} = 0$ and ignoring the second order term in Eq. 3.16$\boldsymbol{A} \cdot \boldsymbol{A}$ that describes the field energy, $\hat{H}_{int}$ is reduced to the rather slim form

$$\hat{H}_{int} = \frac{e}{mc}\boldsymbol{A} \cdot \hat{\boldsymbol{p}}. \tag{3.18}$$

In a three step model the initial state $|i\rangle$ is formally decomposed as

$$|i\rangle = \phi^{i}_{k,\omega}\, |N-1, i\rangle \tag{3.19}$$

where we have explicitly identified the wavefunction of the electron undergoing excitation from the sample as being separate from the (properly antisymmetrized) many body wavefunction describing all the remaining electrons in the system. The final state wavefunction can be similarly decomposed;

$$\langle f| = \phi^{f}_{E_{kin}}\, \langle N-1, f'| \tag{3.20}$$

In this case we distinguish between the final state f of the photoelectron and the final state f' of the remaining system. Noting that the dipole operator only acts upon the photoelectron wavefunction the photocurrent can then be written as

$$I(k,\omega) \propto \sum_{f,i,k} \sum_{f'} |\, \langle \phi^{f}_{E_{kin}}|\, \boldsymbol{A}\cdot\hat{\boldsymbol{p}}\, |\phi^{i}_{k,\omega}\rangle\, \langle N-1, f'|\, c_k\, |N, i\rangle\, |^2 f(\omega)\delta(E_f - E_i - h\nu)\delta(\boldsymbol{k}_{\parallel,i} - \boldsymbol{k}_{\parallel,f}) \tag{3.21}$$

where we have absorbed various constants into the proportionality, explicitly introduced the fermion annihilation operator c_k into the overlap integral between the initial and final states of the many bodied system "left behind" by photoemission and summed over all initial and final states of the photoelectron and (separately) summed over all final states of the $N-1$ body system as is required to evaluate the golden rule.

The first bracket in the above equation denotes the fact that the electrons most likely to be emitted are those with their momentum aligned along the polarization vector of the incident light. This effect, which is essentially geometrical, imparts a modulation to the spectral intensity that can under some circumstances be externally controlled (say by changing the relative angle of incidence and polarization of the exciting light) but usually does not contain any easily determined quantitative information. In the literature of ARPES this is what is usually referred to as the ARPES matrix element,

$$M_{f,i}^2 = |\, \langle \phi^f_{E_{kin}} |\, \boldsymbol{A} \cdot \hat{\boldsymbol{p}}\, | \phi^i_{k,\omega} \rangle \,|^2. \tag{3.22}$$

The second factor describes the response of the system itself to the introduction of a photohole by how that hole subsequently decays due to scattering of the remaining electrons into that vacancy, thus yielding the density of states component of Fermi's Golden Rule. Put all together the photocurrent is proportional to

$$I(\boldsymbol{k},\omega) \propto M_{f,i}^2 A(\boldsymbol{k},\omega) f(\omega) \delta(E_f - E_i - h\nu) \delta(\boldsymbol{k}_{\parallel,i} - \boldsymbol{k}_{\parallel,f}). \tag{3.23}$$

where $A(\boldsymbol{k},\omega)$ is called the spectral function, which we shall examine more closely in a moment.

While the three step model has the somewhat pedagogical advantage of explicitly separating the matrix element contribution to the photocurrent from the density of states defined by the spectral function it is not so easy to derive obvious or experimentally useful conclusions from this formulation. On the other hand, though the validity of the sudden approximation might be called into question, it has the distinct advantage of relating transitions between relatively easily identified single particle states in the crystal to what one can expect to measure. In the one step model one simply avails one's self of the Golden Rule as is,

$$I(\boldsymbol{k},\omega) \propto |\langle f|\, \boldsymbol{A}\cdot\hat{\boldsymbol{p}}\,|i\rangle|^2 f(\omega)\delta(E_f - E_i - h\nu)\delta(\boldsymbol{k}_{\|,i} - \boldsymbol{k}_{\|,f}). \tag{3.24}$$

where we can take the final state wavefunction to be either a free electron (in the vacuum) or a so-called inverse LEED state and we can take the initial state to be that of a Bloch electron in the crystal as described by $A(\boldsymbol{k},\omega)$. Whether one can trust such a simple picture at low energies in the cuprates has been a matter of debate. For photoemission at VUV and higher photon energies the photoelectron certainly escapes the sample long before the system relaxes, if only because of the increased sensitivity to the surface. It has been shown recently that the sudden approximation seems to hold at least for the nodal state of Bi2212 at low T with a 6 eV excitation source by comparing a laser excited spectrum to data taken at a more conventional synchrotron beamline. Jumping ahead a little bit, it is well known that the low energy states of optimally doped Bi2212 show lifetimes τ between scattering events on the order of picoseconds whereas an electron imparted with 6 eV of kinetic energy, even $L = 10$ nm below the surface of a sample, requires a mere handful of femtoseconds t, in a classical calculation, to find its way out of the sample. Thus it is natural to expect that for a given system the sudden approximation will hold so long as $t \ll \tau$. Note that this may not be true for systems in which scattering is very strong, implying a very short electron lifetime, or even for more conventional systems at high temperatures. In any event, the same reasoning shows that a state would have to have an intrinsic width on the order of 100 meV before the sudden approximation might begin to fail.

Returning to the quantum theory of ARPES, we find that the spectral function contains the information about the electronic density of states as a function of energy and momentum we seek to measure,

$$A(\boldsymbol{k},\omega) = |\langle N-1, f'|\, c_k \,|N,i\rangle\,|^2 \delta(\omega - \epsilon_{\boldsymbol{k}}) \tag{3.25}$$

where $A(\boldsymbol{k},\omega)$ is the spectral function of the system. The spectral function is the probability that an electron can be removed (added) from (to) the occupied (unoccupied) part of the electron distribution in a sample. The power of ARPES to probe the many body physics of crystalline solids stems from the simple relationship derived in quantum field theory between the spectral function and the one electron Green's function $G(\boldsymbol{k},\omega)$,

$$A(\boldsymbol{k},\omega) = \frac{1}{\pi}|\operatorname{Im} G(\boldsymbol{k},\omega)|. \tag{3.26}$$

In field theory applied to solid state physics the Green's function contains all the information about the bare dispersion of electrons as would be obtained from, for example, a tight binding model of band structure, as well as information about the interactions, in the form of self energies, that can alter that dispersion as well as lifetimes. The trivial example of this property is the bare (unrenormalized) Green's function that contains no information about electron interactions with themselves or with the medium in which they reside;

$$G_0^{-1}(\boldsymbol{k},\omega) = \omega - \epsilon_{\boldsymbol{k}} - \imath\eta \tag{3.27}$$

which, per Eq. 3.26, yields the spectral function

$$A^0(\boldsymbol{k},\omega) = \delta(\omega - \epsilon_{\boldsymbol{k}}). \tag{3.28}$$

The spectral function shows that in the absence of interactions one expects electrons to appear as Dirac delta functions located at the poles of the noninteracting Green's function, which are themselves defined by the bare dispersion relation $\epsilon_{\boldsymbol{k}}$. The minus sign in front of the infinitesimal $-\imath\eta$ term has been chosen to uphold the causal properties of positive and negative energy states. The power of the Green's function method is that it allows the relatively simple introduction of many body interactions into a formalism that modifies the bare dispersion through the additive inclusion of self energy terms. The self energy is included after renormalization of the bare dispersion by, for example, the electron-electron (Coulomb) interaction, the electron-phonon interaction or any other electron-boson interaction that might be relevant to the system. In field theory these interactions can be represented graphically in terms of Feynman diagrams. Mathematically, self energies can be simply added to the renormalized Green's function,

$$G_{Int}^{-1} = \omega - \epsilon_{\boldsymbol{k}} - \Sigma(\boldsymbol{k},\omega) \tag{3.29}$$

where the self energy term $\Sigma(\boldsymbol{k},\omega)$ is itself a complex quantity,

$$\Sigma(\boldsymbol{k},\omega) = \mathrm{Re}\,\Sigma + \imath\,\mathrm{Im}\,\Sigma. \tag{3.30}$$

The spectral function that derives from this Green's function is

$$A_{Int}(\boldsymbol{k},\omega) = \frac{1}{\pi}\frac{\mathrm{Im}\,\Sigma}{(\omega - \epsilon_{\boldsymbol{k}} - \mathrm{Re}\,\Sigma)^2 + (\mathrm{Im}\,\Sigma)^2} \tag{3.31}$$

which is a Lorentzian line shape. Recalling that the Dirac delta function is the limit of a Lorentzian taken to zero width, one views the non-interacting case, not surprisingly, as one in which the electrons have infinite lifetime. Physically the interacting spectral function reflects the fact that in the presence of new interaction channels electrons have some probability for scattering into new states by absorbing and emitting particles and so have a finite lifetime τ. This lifetime is related to the Lorentzian width in energy measured in the EDC of a photoemission experiment as

$$\frac{\hbar}{\tau} = 2\,\mathrm{Im}\,\Sigma \equiv \Gamma \tag{3.32}$$

where Γ is the full width at half maximum of the Lorentzian defined by Eq. 3.31.The real part of the self energy shows up as a shift in the dispersion of the band away from the bare dispersion. In the presence of an interaction well defined in energy and momentum this shift takes the form of a “kink” in the dispersion at $\omega = E_F \pm \hbar\Omega$ where Ω is the mode energy. The real and imaginary parts of the self energy are self consistently related to each other through causality so that by measuring one you can extract the other by

Kramers-Kronig (KK) transformation. This last fact is especially important for the philosophy of measuring these things in an ARPES experiment because while one can clearly measure linewidths in EDC's without much ambiguity, one certainly has a right in the absence of experimental evidence to ask whether one can legitimately measure the shift in dispersion implied by the real part of the self energy versus a bare dispersion that by definition doesn't really exist. In practice one can often assume a bare dispersion defined by theory and, after measuring the lifetimes, use the KK relations to check for self consistency.

Eq. 3.31 is often written in the somewhat more transparent form used for actually fitting data as,

$$A(\boldsymbol{k},\omega) = \frac{1}{\pi}\frac{Z_{\boldsymbol{k}}\Gamma_k(\omega)}{(\omega - \epsilon_{\boldsymbol{k}} - \xi(\omega))^2 + \Gamma_k^2(\omega)} + (1 - Z_{\boldsymbol{k}})A_{Inc} \tag{3.33}$$

where for brevity we've set $\mathrm{Re}\,\Sigma = \xi$ and introduced what's called the quasiparticle weight $Z_{\boldsymbol{k}}$ and the so-called incoherent part of the spectral function A_{Inc}. Z_k is essentially the normalization constant that arises because the total probability for electron existence, viz a viz Eq. 3.21, must be equal to unity when the incoherent part of the spectrum is taken into account. A_{inc} is not currently well described by theory.

3.3 The Modern Practice of ARPES

The above relations are essentially true for the case of photoemission from one and two dimensional states. Before moving on a few words about how modern ARPES is really performed are in order. This discussion leads in short to the technique of measuring Momentum Distribution Curves (MDC's). Whereas EDC's are a measure of the spectral weight as a function of binding energy for a fixed angle of emission, MDC's measure the momentum (or, more directly, the angular) distribution of the spectral weight for a fixed binding energy. For reasons that will become clear when we turn our attention to low energy ARPES, we must be very careful *not* to use the emission angles θ and ϕ interchangeably with the crystal momentum $\boldsymbol{k}$ per Eqs. 3.8 and 3.9, as is often the case in photoemission using VUV photoexcitation.

Most modern photoelectron spectrometers are based upon what is referred to generically as a hemispherical deflection analyzer (Fig. 3.4). The basic ingredients of a hemispherical analyzer are a collection lens based on electron optics, a channel formed by two concentric hemispheres at different voltages and a detector that registers the electrons. The sample sits at the focus of the electron lens, which collects electrons over some range of emission angles and focuses them so they enter one side of the hemisphere perpendicular to a slit. The width of the analyzer entrance slit defines the best possible energy resolution obtainable in the hemispheric deflector. The outer and inner hemispheres have negative and positive potentials such that electrons entering through the slit have their trajectories bent into an arc whose radius depends upon their kinetic energies. By changing these voltages, electrons with different kinetic energies can be sent on trajectories that land them on the center of the detector on the other side of the hemisphere. By sweeping the voltages one acquires on the detector a photoelectron intensity as a function of kinetic energy, an EDC. Before the advent of the angle multiplexing spectrometers discussed below one would collect EDC's as a function of emission angle as described above, one for every angle; by taking a scan of $I(\omega)$ and then changing the position of the sample relative to the analyzer (or vice versa if the analyzer was small enough to fit in the vacuum chamber) and then taking another scan. The detector of choice for this type of measurement is called a channeltron. It effectively registers a number proportional to the total electron flux impinging upon it. The presentation of this type of data looked much like Fig. 3.5b.

The mathematics governing the orbits of electrons through this type of analyzer are ancient [22]. For an electron of kinetic energy E entering the hemisphere through an entrance slit at a radius R_0 half way between the inner and outer radii R_{in} and R_{out}, respectively, traversing this radius around the hemisphere and landing in the center of the detector at the same radius on the back end, the inner and outer potentials are

$$V_{out} = E \times (3 - 2\frac{R_0}{R_{out}}) \tag{3.34}$$

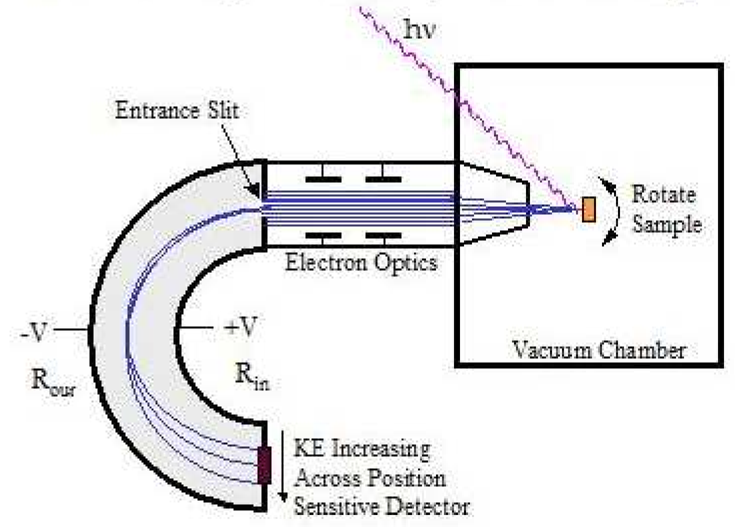
and likewise

$$V_{in} = E \times (3 - 2\frac{R_0}{R_{in}}). \tag{3.35}$$

The energy resolution of such a spectrometer depends ideally upon the width of the entrance slit, S and the kinetic energy these same electrons have as they enter the analyzer. This energy, called the pass energy, E_{pass}, is usually much lower than E_{kin} and is achieved in the lens section of the analyzer by electrostatic deceleration of the incoming photoelectrons. The energy resolution can be approximated as

$$\Delta E = \frac{E_{pass} S}{2R_0}. \tag{3.36}$$

a) Sorting Electrons by Kinetic Energy



b) Sorting Electrons by Emission Angle

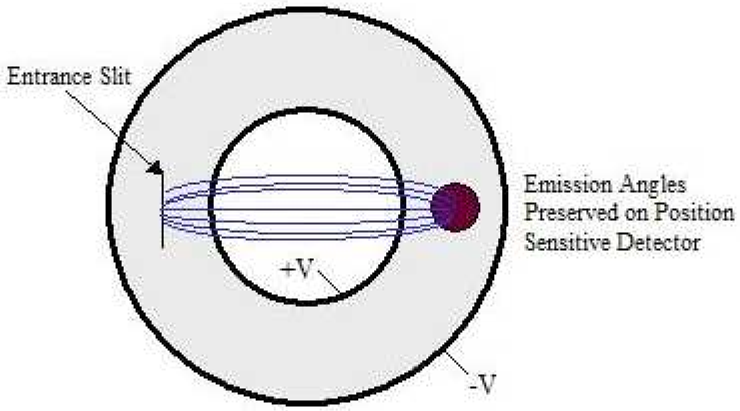


Figure 3.4: Cutaways of generic hemispheric electron spectrometer used in ARPES showing a) kinetic energy sorting geometry and b) angle sorting geometry in hemisphere. Note that the entrance slit can only define energy resolution, not angular resolution.

Our own analyzer has an R_0 of 200 mm and comes equipped with several slits of different widths. As an example, for laser ARPES using 6 eV light, the 1 eV pass energy and a slit of .5 mm we have a theoretical best energy resolution for the spectrometer of 1.25 meV. Experiments performed with synchrotron radiation from a storage ring more often utilize pass energies of 5 or 10 eV and somewhat larger slits. Regardless, the high theoretical performance of this type of spectrometer is more often limited by the bandwidth of the light source and "extrinsic" factors such as analyzer fabrication quality, power supply stability, magnetic fields in the spectrometer and so on.

During the 1990's two developments radically altered the way these spectrometers were used. The first was the introduction of microchannel plates (MCP's) as the detector on the back of the hemisphere. These position sensitive detectors allowed the rapid collection of EDC's by registering the photoelectron intensity of more than one energy at a time across the detector. Referring to the discussion above it is clear that for a given setting of voltages electrons with higher energy will travel longer orbits and so strike a detector at a position of higher radius and vice versa for electrons of lower kinetic energy. Thus in the same way that removing the entrance slight of a photon spectromer and replacing it with a CCD allows the simultaneous measurement of multiple wavelengths so to the removal of the analyzer "exit slit" in favor of a position sensitive detector like an MCP allows detection of electrons with multiple kinetic energies simultaneously. An earlier version of this technique using a lower "pixel count" segmented detector was pioneered by S. D. Kevan [25].

A more important advance followed when it was realized that by tuning the focusing properties of the electron lens at the front of the analyzer one could essentially form an image on the MCP whereby many angles and energies could be measured simultaneously. This development essentially takes advantage of the fact that in a large hemisphere electrons entering along different points along the length of the entrance slit will maintain their relative spatial positions to high accuracy as they are sorted in energy as described above. The curvature of the analyzer does introduce a geometrical abberation that mixes angles and energies to some degree. In later model analyzers this is correctable with a set of curved slits of varying widths. The net result is that rather than taking an EDC for every angle of sample position one could take many angles at once, much faster and with much higher angular resolution. With angle multiplexing spectrometers can effectively make images of electronic bands and so see the fine structure of dispersion and line shapes that had previously been inaccessible.

Angle multiplexing allowed the introduction of new types of analysis on photoemission data. Most importantly, the fine spacing of angles, and therefore momentum points, with resolution on the order of $.1^o$ allowed the analysis of ARPES spectra not only in terms of EDC's but in terms of cuts taken as a function of momentum at constant binding energy, called momentum distribution curves or MDC's. The three common modes of representing an ARPES spectrum, in terms of EDC's, MDC's and false color image plots combining the two is shown in Fig. 3.5.

It turns out that for highly dispersing bands the MDC lineshape typically has a very nearly pure Lorentzian form and a very simple (usually flat) background. This property, contrary to the historical difficulties in fitting EDC's,

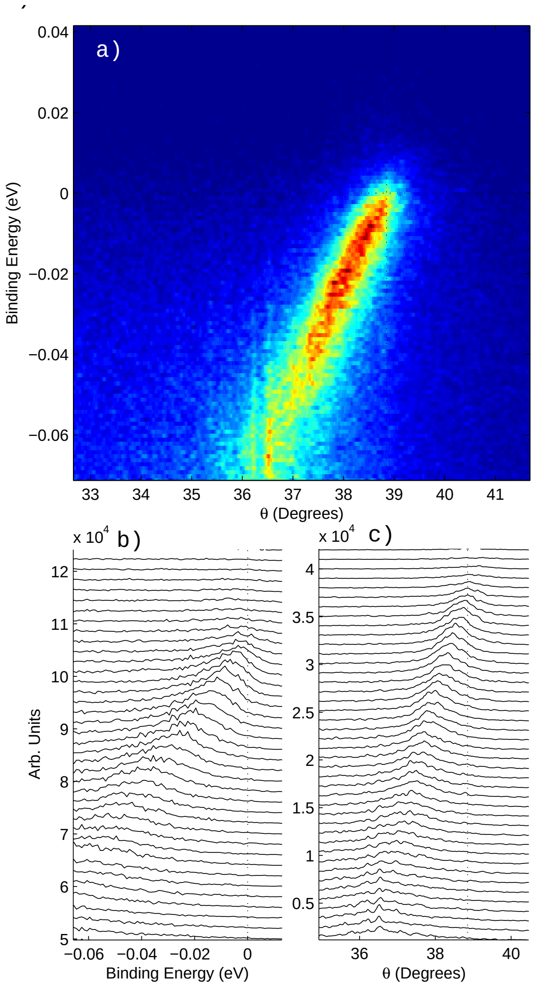


Figure 3.5: An early sample of 6 eV laser ARPES data taken at 20 K on optimally doped Bi2212 in the nodal region. Panel a) shows the 2D spectrum in energy and angle space. The horizontal and vertical dashed lines mark E_F and k_F, respectively. Intensities are mapped to the false colorscale per the colorbar at the top of the panel. Panels b) and c) show EDC and MDC stack plots, respectively, as are often found in the literature.

allows for a relatively simple analysis of many highly dispersing bands seen in ARPES. The quantities measured by an MDC, namely peak position and peak full width at half maximum (FWHM), have a very simple relationship to the quantities measured by EDC's. The FWHM of an MDC, Δk is a measure of the mean free path of the electron in that state (as opposed to the lifetime Γ measured by fitting an EDC),

$$\ell = \frac{1}{\Delta k}. \tag{3.37}$$

By fitting many MDC's to locate the peak positions in momentum space at many binding energies one can easily determine the band velocity and, near E_F, the Fermi velocity v_F,

$$v_F = \frac{1}{\hbar}\frac{\partial \epsilon_k}{\partial k}_{k=k_F} \tag{3.38}$$

The imaginary part of the self energy is related to the bare Fermi velocity v_F^0[26] by

$$\hbar v_F^0 \Delta k_F = 2\,\mathrm{Im}\,\Sigma_F \tag{3.39}$$

where v_F^0 is in m/s. (The natural measure of band velocity in ARPES is actually $eV\mathring{A}$, k and Δk having units of $\mathring{A}^{-1}$.) Not that in the event the band structure is renormalized by the presence of an electron-boson coupling v_F^0 can differ significantly from the renormalized Fermi velocity v_F. In general,

$$\mathrm{Im}\,\Sigma(\omega) = \frac{\hbar}{2\ell}\frac{\partial \varepsilon_{\boldsymbol{k}_\parallel}}{\partial k}. \tag{3.40}$$

We note in passing that the nearly perfect Lorentzian line shape of MDC's indicates the weak momentum dependence of interactions in many systems, relative to their energy dependence.

With the theoretical basics of ARPES in hand it is time to move on to the possible advantages, and disadvantages, of using a very low energy laser for photoemission versus a more traditional source of ultraviolet photons. To begin with, in recent years photoemission has in general been viewed primarily as a probe of the electronic structure of *surfaces*. This mistaken belief derives from the fact that the inelastic mean free path of photoelectrons has a minimum in the VUV energy range most often used in synchrotron and helium discharge lamp based experiments. To see why this is the case, the so-called "universal curve" of photoemission, exemplified by the data in Fig. 3.6, was often pointed to [2]. Because we are only interested in electrons that escape a sample without first scattering such a curve clearly implies that high resolution photoemission as it is most often used is a surface science technique. It is true that a great deal of ARPES has been done on 2D surface states of metals as well as inherently 2D systems such as the dichalcogenides and the cuprates. However it has also been demonstrated that one can equally well obtain band structures even for 3D materials that agree well with theoretical calculations.

This problem of bulk versus surface sensitivity has been compounded by facts of life that have their origin in the photon sources traditionally used in photoemission. While sources of x-rays for XPS (X-ray photoemission spectroscopy) have been around for quite a long time, and while these higher energies certainly yield a bulky signal as far as the universal curve is concerned, even in the event that one could obtain a very narrow bandwidth x-ray beam

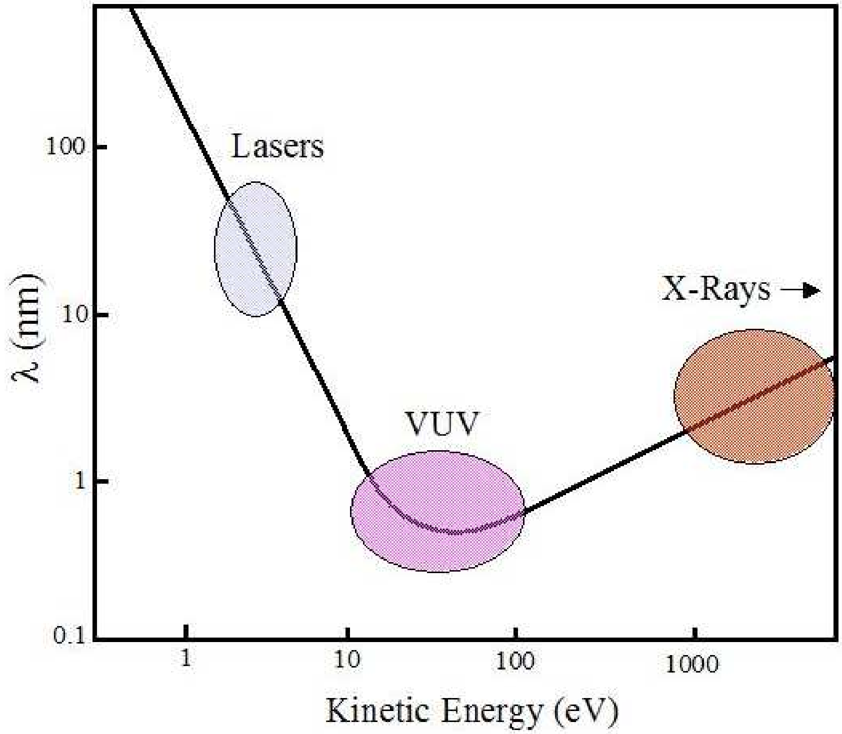


Figure 3.6: Inelastic mean free path of electrons in a crystal, in nanometers, as a function of electron kinetic energy. Adapted from reference [2]

the resolution in k space, viz a viz Eq. 3.10, is certainly not optimal for probing the details of sharp, highly dispersive states. On the other hand, light sources for UPS – ultraviolet photoemission spectroscopy – while somewhat more plentiful now, clearly sit around the minimum of the universal curve, in the vicinity of a 4-5 $\AA$ probing depth. Before synchrotron radiation sources became "widely" available the workhorse of photoemission light sources was the venerable Helium discharge lamp utilizing the He Iα line at 21.22 eV. Synchrotron beamlines can in principle scan a very large range of photon energies from the visible up to many tens of eV depending on their configuration. Our own undulator beamline, U13UB at the NSLS, is equipped with a normal incidence monochrometer yielding a nominal tuning range between 15 and 30 eV, and so is still essentially a surface probe.

In recent years some pointed reminders of the limitations of referencing the universal curve have come to light. For one thing, it must be recalled that the data that went into these curves was taken using the so-called overlayer method in which one takes a spectrum on a clean and well understood reference surface and then monitors the photoelectron intensity of the reference as one deposits thicker layers of another type of atom on the surface. While trustworthy in its context, it is hard to say from first principles exactly how well this type of situation applies to the study of intrinsically layered materials, such as the cuprates. Recent examples in which the simple picture of probing depth has met some difficulty in modern high resolution ARPES include the cases of Sr_2RuO_4, YBa_2CuO_{7+x} and to some extent $Bi_2Sr_2CaCu_2O_{8+x}$.

It was found in strontium ruthenate, for example, that cleaving a crystal in a very good vacuum at low temperatures results in a reconstruction of the surface layer of atoms into a structure different enough from the bulk material that it produced a set of states and Fermi surfaces totally unlike that of the bulk. It was later found that one can tune the photon energy used (and hence the matrix element $M_{f,i}^2$) to enhance or suppress the photoemission intensity of the bulk versus the surface. It was also found that by aging the surface for some amount of time in the vacuum chamber that the surface states were eventually destroyed to the extent one could study the bulk states from underlying layers with relative clarity. In the case of YBCO, at least at optimal doping, it was found that the rather violent process of cleaving this relatively isotropic crystal results in a fairly disordered surface made up partly of exposed BaO planes and partly of broken CuO chains. The result by some as yet unknown mechanism is that the topmost layer of $\mathrm{CuO_2}$ planes becomes maximally overdoped to the point that it does not exhibit any evidence of superconductivity or even any sort of interaction that would result in the strong renormalization of bands typically found in the cuprates. It has been shown that by finding a “good” spot on the surface or, again, by tuning to some special photon energy one can obtain spectra with varying contributions from the bulk and surface components of the spectrum.

Bi2212, mercifully, lies on the extreme other end of surface born complications. In that material the layers are held together very weakly by Van der Waals forces and so have a natural cleavage plane at the BiO layer. The exposed BiO surface is an inert band gap insulator and so the photoemission results entirely from buried CuO_2 layers. The samples live an incredibly long time (days, versus hours) as a result and one only loses photoemission signal as the surface is eventually covered by atoms of residual gas in the vacuum chamber too thick for the photoelectrons to penetrate without scattering strongly. These properties of insensitivity to surface impurity atoms and ease of cleavage have made the Bi2212 system the workhorse of ARPES on the cuprates.

All these complications aside, if one is interested in obtaining spectra that have the greatest relative sensitivity to probing bulk electronic states while maintaining or even improving upon the momentum resolution described by Eq. 3.10 one must go to lower photon energies than are typically utilized in discharge lamps and synchrotron beamlines. This was one major though as yet not decisively tested reason for attempting to use very low energy light from a laser source to perform ARPES. Our own efforts, described in greater detail in later chapters, utilized the harmonic conversion of light from a Ti:Sapphire laser to reach 6 eV. Additional benefits of using a laser are its incredibly high brightness relative to a synchrotron or lamp, its intrinsically narrow bandwidth and, at these very low energies, the ability to use transmitting optics to focus and alter the polarization state of the beam. Additionally, UV lasers represent the first application of truly coherent light for the purpose of high resolution single photon ARPES and so require some modification to how we usually think about the first principles of the photoemission process outlined above. As before I will lay out the peculiar kinematics of very low energy ARPES and then discuss how one might modify the quantum theory of photoemission to take into account the coherent nature of the excitation source.

3.4 Very Low Energy ARPES

Low energy ARPES, particularly utilizing a laser excitation source [27][28] has become an important tool in the arsenal of high resolution photoemission in recent years. In principle one of the most positive effects that arises from using 6 eV light as an excitation source, regardless of its origin or other properties, arises from the momentum resolution one can achieve. Assuming a work function of 4.5 eV (for most materials this number falls between 4 and 5 eV as a rule of thumb) one finds trivially from Eq. 3.10 that the momentum resolution at 6 eV vs. 21 eV increases by a factor of 3.3 and by a factor of 5.8 vs. 55 eV regardless of angular resolution. There is, however, a complication. In general, in three dimensional systems, the lifetime of a state measured by photoemission contains contributions from both the photohole and photoelectron lifetimes because the electron waves normal to the surface are damped. This *measured* lifetime is written as

$$\Gamma_{exp} = \frac{\Gamma_h/v_{h\perp} + \Gamma_e/v_{e\perp}}{|(1/v_{h\perp})[1 - \frac{mv_{h\|}}{\hbar k_\|}\sin^2\theta] - (1/v_{e\perp})[1 - \frac{mv_{e\|}}{\hbar k_\|}\sin^2\theta]|} \tag{3.41}$$

where $v_{e,h;\perp,\|}$ are the electron and hole parallel and perpendicular band velocities, respectively, and $\Gamma_{e,h}$ are the intrinsic electron and hole lifetimes. For the moment we are restricting ourselves to "one angle" equations like Eq. 3.9 for simplicity's sake. Here m is still the electron rest mass. In the event the states of interest are one or two dimensional, $v_{h\perp} = 0$ and the above equation is simplified to

$$\Gamma_{exp} = \frac{\Gamma_h}{|1 - \frac{mv_{h\|}}{\hbar k_\|}\sin^2\theta|}. \tag{3.42}$$

The important point is that in this case the contribution to the measured lifetime from the photoelectron drops out so that one can measure the photohole lifetime independently.

It was shown experimentally by Hansen *et. al* [29] in the days before angle multiplexing that if one attempts to measure the EDC of a very highly dispersing band ($v_{h\|}$ large) at low photon energy, so that θ is large while $k_{\|}$ is relatively small, one can observe the lifetime of that state as being *narrower* than it truly is. This effect was called kinematic compression and results from the fact that at high angles and low photon energies the conversion between measured emission angle and the corresponding $k_{\|}$ changes very rapidly. If the dispersion of this state is positive one will observe a lifetime that is broader rather than narrower. We have dubbed this effect kinematic decompression. A less obtuse way to see this is to examine Eq. 3.9 which shows explicitly that, for a given angle, $k_{\|}$ changes as the square root of the kinetic energy of the photoelectron. It turns out that under most normal combinations of UPS photon energies, interesting emission angles and for low binding energies the square root hardly changes fast enough for the conversion from θ to $k_{\|}$ to appear as anything but a constant for all sampled binding energies. In that case one often sets the k scale at E_F and the conversion for all binding energies is the same.

It turns out however that for all states of interest in the cuprates, ARPES with a 6 eV laser situates you firmly in the regime of high angles and low kinetic energies to which the above analysis applies. To some extent one can handle this complication with relative ease. Whereas for the older style of data acquisition one could only infer the correction to the lifetime of a state from Eq. 3.42, it is a simple matter to scale from θ to $k_{\parallel}$ for every recorded energy, i.e. to scale every MDC appropriately with Eq. 3.9. The resulting algorithm effectively "tilts" and deforms the spectrum in the process. An example of the result of such a procedure is given in Fig. 3.7 in which the distorting effects of having an instrument that measures angles instead of momenta directly is clearly visible. Unfortunately, our ability to apply such a correction to laser ARPES data is limited by the fact that we can only take linear "cuts" in momentum space, such cuts being determined by the entrance slit of the analyzer. The problem with this approach becomes clear if we plot lines of constant $\theta(\omega)$ on a Euclidean grid of momenta and energies as they're measured by the instrument and examines what points in θ space are sampled by constant $k_{\parallel}$ cuts. Evidently one can only perfectly correct spectra in this manner for situations in which the cut taken (or the slit orientation) trace a line back to $\theta = 0$, e.g. the Γ point of any Brillouin zone. The further off such a geometry one attempts to scan, the more the lines of θ in $(\omega, k_{\parallel})$ space change perpendicular to the cut defined by the analyzer slit. In the worst case, as is illustrated in Fig. 3.8 where the slit is oriented perpendicular to a high symmetry line that intersects the Γ point, no correction of this type is possible because as one scans to lower kinetic energies (higher binding energies) there is no simple way to scale from angle to the rapidly changing momentum variable. The problem is only compounded when we consider both polar and azimuthal angles of emission simultaneously as with Eq. 3.8. While in principle one could stack a very high density of this worst sort of cut in order to extract lines of constant $k_{\parallel}$ from a large number of spectra, in practice this would be an exceedingly challenging procedure and would not likely achieve

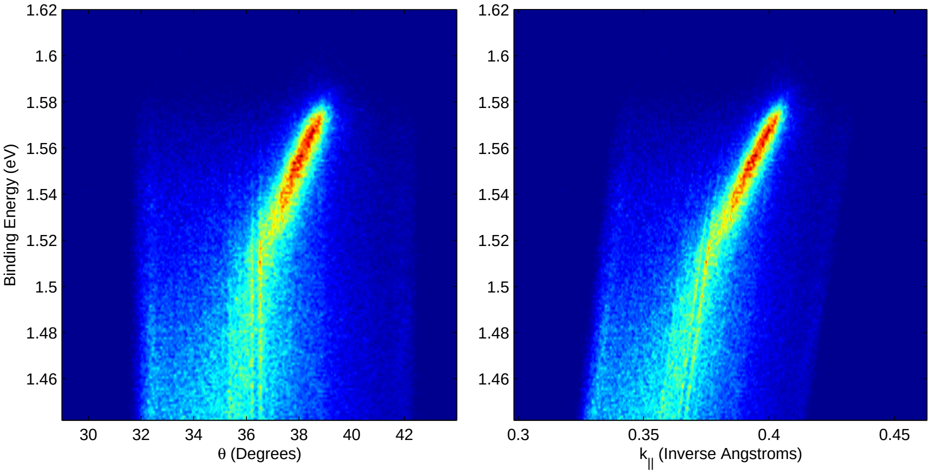


Figure 3.7: A typical 6 eV laser ARPES spectrum near the node of Bi2212. Panel a) shows the spectrum as it's acquired by the electron spectrometer. Panel b) shows the result of kinematic compression upon converting from θ to $k_{\|}$ space. Regions of the MCP detector beyond those where spectra are typically acquired are shown for clarity. The main effect of the conversion from angle to momentum space is the "tilting" and compression of the spectrum in the θ direction and the subsequent sharpening in energy of the observed dispersing band.

the same *angular* resolution as is easily achieved by a single scan at higher photon energies.

3.5 Coherence in the ARPES Light Source

To conclude this chapter on a more positive note we turn to a brief discussion of the ramifications of using a coherent light source, a laser, versus the more traditional incoherent lights sources such as synchrotron radiation. What difference there may or may not be between coherent and incoherent light sources is a very natural question. To begin with, we recognize that

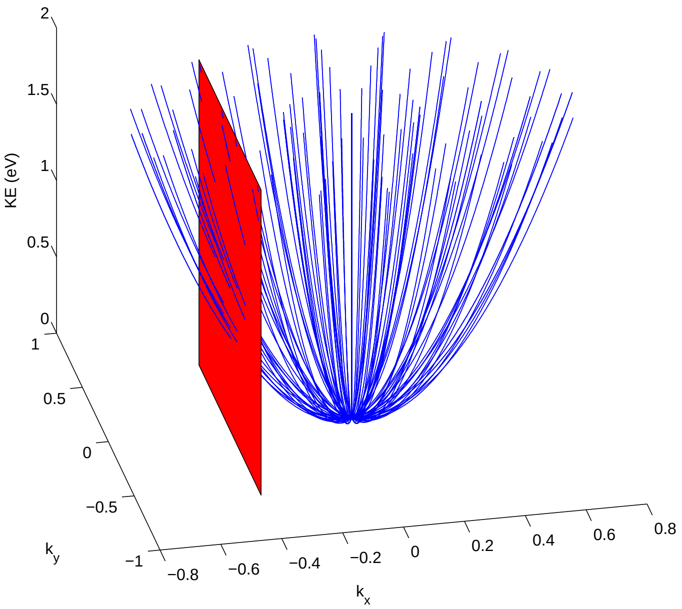


Figure 3.8: Lines of constant angle in k space. The red cut through (k_x, k_y, E), easily obtainable at high photon energies, must be constructed from many cuts though angle space in low energy ARPES.

whatever the properties of the light we choose to use in a photoemission experiment, the relevant quantity of interest is the vector potential that appears in the interaction Hamiltonian, Eq. 3.16. The electromagnetic field is quantized by representing the vector potential in explicit operator form and writing it in terms of a sum over harmonic oscillators, one for each mode $\boldsymbol{k}$ [30],

$$\hat{\boldsymbol{A}} = \sum_{\boldsymbol{k}} \sqrt{\frac{\hbar}{2\epsilon_0 V \omega_{\boldsymbol{k}}}} \boldsymbol{\varepsilon}_{\boldsymbol{k}} \{\hat{a}_{\boldsymbol{k}} e^{-\imath\omega_{\boldsymbol{k}} t} e^{\imath \boldsymbol{k}\cdot\boldsymbol{r}} - \hat{a}^{\dagger}_{\boldsymbol{k}} e^{\imath\omega_{\boldsymbol{k}} t} e^{-\imath \boldsymbol{k}\cdot\boldsymbol{r}}\} \tag{3.43}$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{k}}$ is the polarization vector, V is the volume under consideration, $\omega_{\boldsymbol{k}}$ is the photon mode energy and $\hat{a}_{\boldsymbol{k}}$ and $\hat{a}^{\dagger}_{\boldsymbol{k}}$ are the harmonic oscillator annihilation and creation operators, respectively, that here play the role of photon creation and annihilation operators. Note that in this section we are using $\boldsymbol{k}$ as the photon wave vector and not as the crystal momentum. A general beam of incoherent light can be represented by number states $|n\rangle$ with the usual properties of harmonic oscillator states. There is no quantum uncertainty in photon number in this instance ($\Delta n = 0$) because there is no phase relationship amongst the photons.

In general, an incoherent beam of light possesses no microscopic temporal structure and thus consists of a well defined number of photons or, more formally, can be represented by a complete basis of well defined number states. Coherent light on the other hand is represented as a rather special superposition of number states

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_n \frac{\alpha^n}{(n!)^{\frac{1}{2}}} |n\rangle \tag{3.44}$$

where α satisfies the eigenvalue equation for the annihilation operator

$$\hat{a}_{\boldsymbol{k}} \left|\alpha\right\rangle = \alpha \left|\alpha\right\rangle . \tag{3.45}$$

α, which is complex, can be usefully represented with a phase so that

$$\alpha = |\alpha| e^{\imath\theta} \tag{3.46}$$

where we now use θ as the phase angle. For the purposes of evaluating the photoemission matrix element $M^2_{f,i}$ in the presence of a coherent excitation source we first take the expectation value of the vector potential for a coherent state.

$$\left\langle\alpha\right| \hat{\boldsymbol{A}} \left|\alpha\right\rangle = \sum_{\boldsymbol{k}} \sqrt{\frac{\hbar}{2\epsilon_0 V \omega_{\boldsymbol{k}}}} \boldsymbol{\varepsilon}_{\boldsymbol{k}} \{ |\alpha_{\boldsymbol{k}}| e^{\imath\theta_{\boldsymbol{k}}} e^{-\imath\omega_{\boldsymbol{k}} t} e^{\imath\boldsymbol{k}\cdot\boldsymbol{r}} - |\alpha_{\boldsymbol{k}}| e^{-\imath\theta_{\boldsymbol{k}}} e^{\imath\omega_{\boldsymbol{k}} t} e^{-\imath\boldsymbol{k}\cdot\boldsymbol{r}} \} \tag{3.47}$$

which reduces to the rather nice form

$$\left\langle\alpha\right| \hat{\boldsymbol{A}} \left|\alpha\right\rangle = \sum_{\boldsymbol{k}} \sqrt{\frac{2\hbar}{\epsilon_0 V \omega_{\boldsymbol{k}}}} \boldsymbol{\varepsilon}_{\boldsymbol{k}} |\alpha_{\boldsymbol{k}}| \cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega_k t + \theta) \tag{3.48}$$

This form of the vector potential (and, similarly, the expectation values of the quantized electric and magnetic fields) overtly displays the complementarity of coherent light with the classical form of an electromagnetic wave in the limit of large α for a given mode. As a result of this, an explicit time (for a traveling wave) or space (for a standing wave) dependence in the form of a well defined phase enters the computation of the transition matrix element $M^2_{f,i}$. As before we retain only the $\boldsymbol{A} \cdot \hat{\boldsymbol{p}}$ term in $\hat{H}_{int}$ and so find explicitly that

$$M_{f,i}^2 = (\frac{e}{m})^2 \sum_{\boldsymbol{k}} (\frac{2\hbar}{\epsilon_0 V \omega_{\boldsymbol{k}}}) |\alpha|^2 | \langle f; \alpha | \cos(\boldsymbol{k} \cdot \boldsymbol{r} - \omega_k t + \theta) \boldsymbol{\varepsilon}_{\boldsymbol{k}} \cdot \hat{\boldsymbol{p}} | i; \alpha \rangle |^2. \quad (3.49)$$

There are several things of note that appear in this equation. First we find an explicit dependence of the magnitude of the matrix element on a factor $\omega_{\boldsymbol{k}}^{-1}$ arising solely from the quantized form of the vector potential regardless of the nature of the photon states it acts upon. Evidently the coupling of initial to final states has a very general dependence on the photon energy and so we find that states should appear intrinsically "brighter" as the photon energy is decreased in a photoemission experiment, other factors not withstanding.

The obvious effects of using coherent light appear to be manifold yet should only exhibit themselves in a real experiment under certain conditions. The factor $|\alpha|^2$ appearing outside the bracket is the *mean* number of photons in the beam, $\langle \alpha | \hat{n} | \alpha \rangle = |\alpha|^2$ where $\hat{n}$ is the harmonic oscillator number operator. In the limit of large α (bright beams) this should asymptotically approach the eigenvalue obtained from the number states $\langle n | \hat{n} | n \rangle = n$. Conversely in the limit of low α (weak beams) the number-phase uncertainty relation ($\Delta n \Delta \theta \geq \hbar/2$) will, for a beam with well defined phase, guarantee an inherent instability in the number of photons found by our sample in the beam and thus in the photoemission intensity.

Another effect of using coherent light is the explicit introduction of spatiotemporal oscillations in the field impinging upon the sample. In general even for rather high photon energies the wavelength of light is such as to render the field essentially constant over the volume of sample being probed and so this is not important. Time variations of the field are also generally too fast to be observed in a photoemission experiment with pulsed lasers and so one should expect the time average of the variation in the field to be the relevant quantity in the matrix element. Nevertheless it is not hard to imagine that for the case of a continuous wave (CW) laser one could through a combination of polarization phase relationships in the beam (to modulate couplings to different states) and positioning of the wave nodes of the CW beam relative to the surface achieve a sort of depth modulated photoemission. Even in the simplest case one could imagine positioning the beam node at the sample surface and so achieve a photoemission signal manifestly due to excitations from the bulk. In such a case, or even in the event of a very high f number focus of the beam, one might have to include the $\hat{\boldsymbol{p}} \cdot \hat{\boldsymbol{A}}$ term in consideration of the experiment.

This is certainly not all one can do with the coherent nature of light in a photoemission experiment. The use of orbital angular momentum states of light might be used for selectively coupling light to electronic states of certain symmetries in a manner analogous to what is already done with circularly polarized light in magnetic circular dichroism experiments. The literature on laser physics is certainly vast and many books on the topic contain excellent treatments of the theory of photoemission from the point of view of optical and atomic physics rather than solid state physics. What's more, as photoelectron analyzers continue to push the bounds of angular and energy resolution, as well as temporal resolution in various kind of time of flight detector arrangements, it is certainly worth considering how one might apply the methods of ARPES to the study of light itself. Further consideration of such topics is well beyond the scope of this work and the intention has been only to point out how one might reconsider what has become well trod ground in the ARPES community in light of the arrival of a powerful new set of tools.

Chapter 4

Harmonic Generation

The generation of laser harmonics by nonlinear processes in optical crystals was first demonstrated in 1961 shortly after the discovery of the original ruby laser [31]. The work was stimulated by the then new ability to produce (by the standard of the time) very intense, monochromatic and well collimated laser light beams which, when focused to a point, produced transient local voltages well in excess of what had been previously possible. Since that time harmonic generation has found wide application in the operation and manipulation of laser beams in many experimental and engineering circumstances. In recent years it has been found that by using multiple harmonic generation steps one can produce laser beams of sufficiently high energy and intensity so as to be extremely useful for the practice of angle resolved photoemission spectroscopy[27]. While these processes result in deep ultraviolet (DUV) laser beams that are by many standards extremely weak, even a very weak DUV laser is often more than an order of magnitude more intense than a beam of corresponding energy produced, for example, in a typical synchrotron beamline or gas discharge lamp. Further, the laser offers these beams in a form with high purity of polarization, collimation and for which the focusing properties of Gaussian optics come into play. Some of these advantages have been alluded to in earlier chapters. It is the purpose of this chapter to first give a brief summary of the electromagnetic theory of harmonic generation in dielectric media and to then demonstrate by numerical simulation the feasibility of producing a suitable DUV beam from picosecond laser pulses. In particular we shall highlight the relative challenges of using picosecond, versus femtosecond, laser radiation for such purposes as well as the ultimate benefits of doing so despite the inherent difficulties of such an approach.

As might be expected by the general heading of “nonlinear optics” in association with harmonic generation, any theoretical description of the generation of frequency overtones of a laser beam in a ponderable medium will necessitate the consideration of nonlinear terms in the equations of motion describing the interaction between oscillating electromagnetic fields, e.g. from the laser, and the atoms and molecules of the system with which they interact. While nonlinear equations are generally impossible to solve analytically some reasonable conclusions can be arrived at by the usual judicious employment of suitable approximation schemes. Further, while the full description of the interaction of photons with charged particles necessitates a quantum theory, as is often the case in optics one can arrive at some useful and more or less physically meaningful conclusions from examining classical and quasi classical approximations to the full problem. Thus, to keep the discussion of second harmonic generation physically transparent and on a relatively accessible level for non-specialists in this field of optics the following discussion is drawn almost exclusively from that contained in reference [32]. The object is merely to provide a context for understanding the results of the full blown numerical simulations of the harmonic generation scheme employed in the present photoemission experiments. Such simulations are vital for determining the best choice of, for example, optics, crystals and laser wavelengths with which to produce a DUV laser beam suitable for performing ARPES.

4.1 Nonlinear Mechanics of a Driven Harmonic Oscillator

The starting point for the theory of laser-matter interactions that will lead to the generation of harmonics of an incident laser beam is the classical harmonic oscillator model of an electron, bound to an ionic core experiencing the time dependent force generated by the electric fields of an incident laser beam. Assuming the electron at equilibrium resides in a spatially radially varying potential well $V(r)$ one begins expanding the potential in a Taylor series about the minimum. Thus,

$$V(r) = V(r_0) + r(\frac{dV}{dr})_{r=r_0} + \frac{1}{2!}r^2(\frac{d^2V}{dr^2})_{r=r_0} + \frac{1}{3!}r^3(\frac{d^3V}{dr^3})_{r=r_0} \dots \qquad (4.1)$$

where r_0 is the location of the potential well minimum. The equation of motion for the electron driven by the laser electric field depends not upon the potential but upon the force, which enters as derivatives of the potential. Therefore we can set the first term, a constant, to zero. The first derivative of the potential is also zero at the well minimum and so this term is dropped as well. The quadratic term simply gives the usual result for the harmonic oscillator so that the coefficient can be identified as

$$(\frac{d^2V}{dr^2})_{r=r_0} = m\omega_0^2. \tag{4.2}$$

By going to higher order, to the third term, we introduce a first order nonlinearity into the problem. Not knowing *a priori* what the constant associated with the derivative should be (this can evidently be calculated from first principles in quantum mechanics) we shall simply assign it a symbol,

$$A = \frac{1}{3!}(\frac{d^3V}{dr^3})_{r=r_0}. \tag{4.3}$$

Consideration of higher order terms is unnecessary for the derivation of the primary nonlinear effects we are interested in here.

Recalling that $F = -dV/dr$ we find that the terms in the classical equation of motion for the electron in a potential well deriving from that potential are

$$F = -\frac{dV}{dr} = -m\omega_0^2 x - 3Ax^2 - ... \tag{4.4}$$

so that the full equation of motion, when the driving force $eE(t)$ is included, is

$$\frac{e}{m}E(t) = \ddot{r} + \omega_0^2 r + ar^2 + ... \tag{4.5}$$

where we have set $\frac{3A}{m} = a$ for the sake of a notational simplicity whose advantage will become clear momentarily. Finally, for a well defined (linearly polarized) laser beam we can, as in chapter 3, set $E(t) = E_0 \cos(\omega t)$ so that our compact expression, now dropping the ellipses, is

$$\frac{e}{m}E_0\cos(\omega t) = \ddot{r} + \omega_0^2 r + ar^2. \tag{4.6}$$

As alluded to above, solving nonlinear differential equations such as Eq. 4.6 analytically is in general not possible without making some major concessions to approximation. One such scheme that is standard in mechanics problems if not frequently taught in the schools is to recognize that a first order, approximate solution can be arrived at by recognizing the smallness of the nonlinear term (a good physical approximation in this case) and then using that solution to iteratively derive higher order corrections that explicitly include the progressively smaller nonlinearities. Thus the first approximation consists of solving that most familiar of mechanical equations (minus a linear restoring force)

$$\ddot{r}^{(1)} + \omega_0^2 r^{(1)} = \frac{e}{m}E_0\cos(\omega t). \tag{4.7}$$

The parenthetical superscript here denotes the order of approximation. The complementary solution to this equation, the same as would be got in the absence of a driving force, is simply $r_c^{(1)} = A_1 exp(\imath\omega_0 t) + A_2 exp(-\imath\omega_0 t)$, A_1 and A_2 arbitrary constants. In the event the perturbation is weak, as is our stated case, this simply represents a perpetual sympathetic oscillation of no interest. In the event the perturbation were strong, we may introduce a linear restoring force of the form $k\dot{r}$ which would in any event cause this solution to exponentially decay in time and so cause it to remain irrelevant. The particular solution to the driven harmonic oscillator problem we are interested in is

$$r^{(1)} = \frac{e/m}{\omega_0^2 - \omega^2} E_0 \cos(\omega t). \tag{4.8}$$

The nonlinear equation approximate solution scheme follows by using this solution for $r^{(1)}$ as the coordinate for the nonlinear term ar^2 while solving for the next iteration of r, $r^{(2)}$, as though it were a harmonic oscillator with a now modified driving force. Thus,

$$\ddot{r}^{(2)} + \omega_0^2 r = \frac{e}{m} E_0 \cos(\omega t) - a(r^{(1)})^2 \tag{4.9}$$

where we plug in the solution to Eq. 4.8 on the right hand side for $r^{(1)}$. The particular solution to this equation for $r^{(2)}$ is [31]

$$r^{(2)} = \frac{e/m}{\omega_0^2 - \omega^2} E_0 \cos(\omega t) - \frac{a}{2\omega_0^2} (\frac{e/m}{\omega_0^2 - \omega^2})^2 E_0^2 - \frac{a}{2} \frac{1}{\omega_0^2 - 4\omega^2} (\frac{e/m}{\omega_0^2 - \omega^2})^2 E_0^2 \cos(2\omega t). \tag{4.10}$$

And so we find the central result that the second order approximate solution to the nonlinear harmonic oscillator contains a component of the motion for which the electron oscillates at *twice* the frequency of the incident laser beam. The physical ramifications of this effect for the absorption and emission of radiation will be addressed in the next section.

Expressions such as these can be greatly simplified by switching to the more practical exponential notation for oscillatory phenomena whereby the fundamental electromagnetic wave is written as

$$E_0 \cos(\omega t - kz) = \frac{1}{2} [E_\omega(z) e^{-\imath(\omega t - k_\omega z)} + c.c.] \tag{4.11}$$

where z is the linear coordinate along the propagation direction inside the crystal medium and the wave oscillating at twice the fundamental frequency is written as

$$E_0\cos(2\omega t - k_{2\omega}z) = \frac{1}{2}[E_{2\omega}(z)e^{-\imath(2\omega t - k_{2\omega}z)} + c.c.] \tag{4.12}$$

where the second harmonic takes a different wave vector from the fundamental reflecting the difference in the frequency dependent index of refraction for the medium,

$$k(\omega) = n(\omega)\frac{\omega}{c}. \tag{4.13}$$

In this form the second harmonic electron displacement, now written explicitly as a function of time, becomes (assuming $E(z)$ is real)

$$r^{(2)}_{2\omega}(z,t) = -\frac{a}{4}\frac{1}{\omega_0^2 - 4\omega^2}(\frac{e/m}{\omega_0^2 - \omega^2})^2 E_\omega^2[e^{-2\imath(\omega t - k_\omega z)} + c.c.] \tag{4.14}$$

in which the factors of 2 in the exponentials arise for the same reason as the factor of 2 in the cosine of Eq. 4.10.

4.2 The Wave Equation in Ponderable Media

Thus far we have only made the process of harmonic generation plausible by examining the effect of a nonlinear term in the equation of motion for an oscillating electron-ion system. With this result in hand we now turn our attention to the details of directly connecting the behavior of highly stimulated matter to the wave equations governing the propagation of a laser beam and its harmonics in a dielectric medium. The aim of this section is to provide a more direct contact to the physically measurable quantities associated with materials and electromagnetic waves that will ultimately provide the tools necessary to determine whether and under what circumstances harmonic generation can be practically achieved in the laboratory. While the discussion to follow on the production and distribution of primary and second harmonic fields in the medium appears a bit abstract at first it will ultimately produce the several simple and vital results that delineate when and how harmonic generation can occur.

The connection between the physical properties of a material, a material's behavior when subject to the presence of electromagnetic fields and how the fields themselves are modified by the interaction are summarized by the equations for electrical polarization and electric susceptibility, respectively, which are

$$P = Ner = \epsilon_0 \chi E \tag{4.15}$$

where P is the polarization vector, N is the number of electron-ion systems per unit volume and χ is the electric susceptibility. (Though P and E are vector quantities there is no loss in generality when working in one dimension at this stage.) The essential point here is that the polarization directly connects the nonlinear displacement of the electron subject to the oscillating field as described in the previous section to the material susceptibility, including the frequency dependent index of refraction, and the electric field itself. This connection is made at an even more fundamental level by examining the Maxwell wave equation with a source

$$\nabla^2 E - \epsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2} = \mu_0 \frac{\partial^2 P}{\partial t^2}. \tag{4.16}$$

The implication is that the nonlinear polarization induced by the fundamental, $P_{2\omega}^{NL} = Ner_{2\omega}^{(2)}$ acts as a source generating an oscillating field $E_{2\omega}$, e.g. the second harmonic. This is

$$P_{2\omega}^{NL} = -\frac{Nea}{4} \frac{1}{\omega_0^2 - 4\omega^2} (\frac{e/m}{\omega_0^2 - \omega^2})^2 E_\omega^2 [e^{-2\imath(\omega t - k_\omega z)} + c.c.]. \tag{4.17}$$

Additionally, one must consider the linear polarization induced in the medium by the second harmonic itself, $P_{2\omega}^{L}$, which is identical to the first term in Eq. 4.10 with the replacements $\omega \rightarrow 2\omega$, $k_\omega \rightarrow 2k_{2\omega}$ and, $E_\omega \cos(\omega t) \rightarrow \frac{1}{2}[E_{2\omega} exp(-\imath(2\omega t - k_{2\omega} z)) + c.c.]$. Thus the full polarization to be considered when evaluating Eq. 4.16 is $P = P_{2\omega}^{NL} + P_{2\omega}^{L}$.

Following some rather unilluminating algebraic manipulations carried out in reference [32] Eq. 4.16 can be reduced to a relatively simple form relating the nonlinear polarization responsible for generating the second harmonic wave to the electric field of that wave

$$\frac{dE_{2\omega}(z)}{dz} = \imath\omega \sqrt{\frac{\mu_0}{\epsilon_{2\omega}}} \bar{d} E_\omega^2(z) e^{\imath \Delta k z} \tag{4.18}$$

where $\bar{d}$ is the relevant component of the nonlinear susceptibility tensor relating fundamental field intensities to nonlinear polarizations as (for example)

$$P_{2\omega}^{NL} = \bar{d} E_\omega^2(z) \tag{4.19}$$

and

$$\Delta k = 2k_\omega - k_{2\omega} = 2\omega\sqrt{\epsilon_0\mu_0}[n(\omega) - n(2\omega)]. \tag{4.20}$$

These last three equations, when considered together, encode the most important practical aspects of harmonic generation. That is, the strength of the second harmonic field $E_{2\omega}$ generated in a ponderable medium depends directly upon the *intensity* E_ω^2 of the fundamental beam, the strength of the nonlinear susceptibility $\bar{d}$ connecting the fundamental field to the generation of a nonlinear polarization in that medium and the magnitude of difference in the indices of refraction in that medium between the fundamental and second harmonic beams. The success of harmonic generation depends upon the control of these three parameters. In the following section we shall explore how one might go about optimizing these parameters. This discussion will lead directly to the specific examples of second harmonic generation upon which our photoemission experiments rely.

4.3 Phase Matching and Crystal Properties

Besides the fundamental beam intensity, some practically relevant properties of which we shall investigate in the next section, the most important considerations for the successful production of useful second harmonic beams reside in the Δk term of Eq. 4.18 and in the nonlinear susceptibility denoted above by $\bar{d}$. Both quantities are governed by physical properties of the nonlinear medium, most often a birefringent crystal.

Eq. 4.18 can be easily integrated from $z = 0$ to L and squared to yield an equation for the intensity of the second harmonic beam in terms of the fundamental intensity, relative wavevectors of the fundamental and second harmonic, Δk, and the crystal length L

$$E_{2\omega}^2(L) = \frac{\mu_0 \omega^2 \bar{d}^2}{\epsilon_{2\omega}} |E_\omega(0)|^4 L^2 (\frac{\sin(\Delta k L/2)}{\Delta k L/2})^2 \tag{4.21}$$

where L is the crystal length. In this form, the direct dependence of the efficiency of second harmonic generation upon Δk is manifest. For a given L, the second harmonic intensity varies according to the Airy function $\sin^2(x)/x^2$ with maximum intensity at $\Delta k = 0$. The process by which Δk is minimized is called phase matching, with $\Delta k = 0$ critically phase matched. A second harmonic generation process for which $\Delta k \neq 0$ is referred to as quasi phase matched. The phase matching condition is most physically transparent in Eq.4.20, from which it becomes clear that optimal phase matching, and thus second harmonic generation, occurs when the index of refraction for the fundamental beam equals the index of refraction for the second harmonic beam. Physically, an index mismatch causes the fundamental and second harmonic beams to travel through a crystal with different phase velocities $c/n(\omega)$. Over time the waves fall out of phase with each other and no more energy can be transferred from the fundamental to the second harmonic. Thus the name phase matching. During quasi phase matched harmonic generation processes the harmonic beam, due to group velocity mismatches, falls behind the fundamental pulse and is said to walk away. Further, a mismatch in indices of refraction will, over a long enough crystal, cause the second harmonic beam to travel along a slightly different spatial trajectory than the fundamental beam thus further weakening it. In this case, the second harmonic is said to walk off. In general a coherence length $L_c = |\pi/\Delta k|$ is often defined, indicating the point beyond which walk off and walk away effects render further lengthening of the crystal ineffective.

Because it is far too much to ask of any random crystal that its indices of refraction be identical for wavelengths separated by a factor of exactly two, useful nonlinear crystals generally exhibit a very strong birefringence. Birefringent crystals have different indices of refraction along different crystallographic directions. Classic examples of such materials include calcite and quartz, high quality samples of which can be found in nature. Unpolarized light entering a birefringent crystal from the appropriate direction will appear to be split as the different polarization components of the light are separately refracted. This effect is historically called double refraction. Birefringence does not occur in crystals with centrosymmetric structural symmetry such as those with cubic type structures. A crystal whose structural symmetry is broken in only one direction is said to have a uniaxial birefringence. The optic axis of such a crystal is defined as the propagation direction for which all polarizations of light experience the same index of refraction, termed the ordinary index, n_o. Light waves incident upon the crystal from a direction normal to the optic axis are termed either ordinary waves (o waves) if they are polarized normal to the optic axis or extraordinary wave (e waves) if they are polarized parallel to the optic axis. The situation, often stated but for some reason seldom drawn, is diagrammed in Fig. 4.1.

The extraordinary index of refraction n_e is defined for an angle of incidence of $\pi/2$ relative to the optic axis. The index of refraction actually experienced

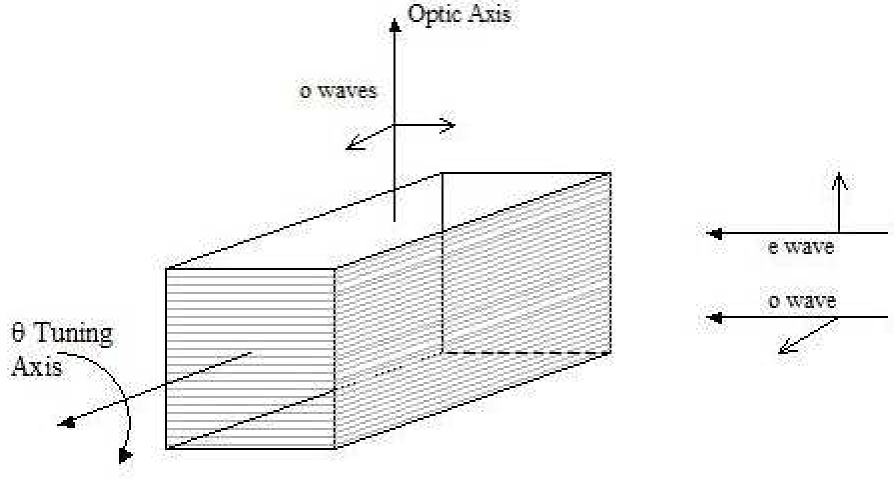


Figure 4.1: Schematic of polarization properties of a uniaxial birefringent crystal. Gray lines indicate the manner in which the crystal structural symmetry is broken. The optic axis is defined by the wave propagation direction for which all polarizations experience the same index of refraction. Ordinary (o) polarized waves experience the same index of refraction n_o as if they traveled along the optic axis. Extraordinary (e) polarized waves experience a fraction of the extraordinary index of refraction n_e depending upon the tuning angle θ of their angle of incidence.

by an e wave can be continuously tuned between it's maximum value of n_e and n_o by changing its angle of incidence relative to the optic axis. This tuning is expressed quantitatively by the relation

$$\frac{1}{n_e^2(\theta,\omega)} = \frac{\cos^2\theta}{n_o^2(\omega)} + \frac{\sin^2\theta}{n_e^2(\omega)}. \tag{4.22}$$

A crystal for which $n_e(\omega) > n_o(\omega)$ is called a positive uniaxial crystal. Conversely, a negative uniaxial crystal has the property $n_e(\omega) < n_o(\omega)$.

This ability to tune the effective n_e while leaving n_o constant is the key to the so-called angle tuning method of phase matching. The object of angle tuning is to find a crystal that, for some angle of incidence, matches $n_e(\omega)$ to $n_o(2\omega)$ (for positive uniaxial crystals) or $n_o(\omega)$ to $n_e(2\omega)$ (negative uniaxial). For positive uniaxial crystals the phase matching angle θ_M is determined by setting the ω and 2ω indices equal and solving for θ_M. Thus

$$\sin^2\theta_M = \frac{\frac{1}{n_o^2(\omega)} - \frac{1}{n_o^2(2\omega)}}{\frac{1}{n_o^2(\omega)} - \frac{1}{n_e^2(\omega)}}. \tag{4.23}$$

For negative uniaxial crystals the role of e and o in the denominator is reversed.

In the laboratory there are two further considerations when angle tuning a nonlinear crystal. The phase matching angle θ_M is the angle made between the propagating waves *inside* the crystal and that crystal's optic axis. For a real crystal the actual angle of incidence of the fundamental beam impinging upon the surface will in most cases differ from θ_M because the beam will be refracted according to Snell's law. For this reason crystals destined for application with particular wavelengths are often cut and polished so that the surface normal lies close to θ_M rather than some strange angle relative to the optic axis. This decreases the Snell's law effect as well as mitigating problems associated with reflection losses to the s polarized o wave beam at the crystal surface. Further tricks can be played to mitigate such effects such as using crystals cut to a Brewster angle so one wave can pass in or out without intensity losses to reflection. Losses to absorption, scattering and optical damage induced by high laser fluences sadly depend upon the innate absorption spectrum of the crystal as well as it's manufacturing quality and purity. All these consideration aside, the attraction of angle tuning to achieve good phase matching is its inherent simplicity. Because phase matching only depends upon one's ability to tune one or both of the indices of refraction of a birefringent crystal several other methods exist. These include, for example, temperature tuning, pressure tuning and combinations of methods which exploit, for example, $n(\omega, T, P, \theta)$. These methods, not used in the experimental work described in this thesis, will not be treated here.

Returning to the physical properties of nonlinear crystals it is worth saying a few words about the nonlinear coefficient $\bar{d}$. The factor $\bar{d}$ introduced above as the nonlinear coefficient is actually the component of a nonlinear susceptibility tensor describing the ability of a material to generate harmonics of one polarization at one frequency from another polarization and frequency. For the purpose of examining second harmonic generation it suffices to point out that a) the bigger the $\bar{d}$ the more efficiently a second harmonic will be generated and b) by its relation to the electric susceptibility it is also directly dependent upon the index of refraction of a material at the fundamental and second harmonic frequencies. This last relationship often means that the larger the index of refraction of a material the more likely it is to have a large nonlinear coefficient. Ultimately, $\bar{d}$ is a material property and to the extent that one has a limited number of choices of nonlinear crystals appropriate to a given harmonic generation scheme, you frequency double with the $\bar{d}$ you have, not the one you want. Some tricks can be played e.g. with crystal temperature, etc., but by and large the crystal with the largest $\bar{d}$ that satisfies the phase matching condition is the one you use.

4.4 Fundamental Considerations

Clearly, setting aside all other contributions for the moment, the easiest way to increase the intensity of the second harmonic is to increase the intensity of the primary input beam. There are several ways to accomplish this, most of which are taken advantage of in the current experiment. First, if possible, one can simply apply brute force to the problem by using a more powerful laser. In our own experiments, the physical description of which is provided in more detail in a later chapter, this was accomplished by increasing the power of the fundamental beam from .7 Watts to 1.5 Watts. Because the price of modern lasers tends to scale with their output power this is a simple albeit expensive method of improving harmonic generation efficiency. This approach is limited ultimately by space available – which is always at a premium in optics laboratories and at synchrotron facilities – and by the amount of power it is practically possible to get out of a given type of laser. In the case of the conventional mode locked Ti:Sapphire laser 2 Watts is generally the limit.

Given an average power limited beam there are several ways to further maximize harmonic generation efficiency prior to considering the dielectric medium and its radiation response. A simple way to increase the intensity of a beam inside the medium, where it is needed, is to focus the beam with lenses or curved mirrors so that the power density is at a maximum in the interaction region. Thus the power figure of merit for the primary beam inside the crystal is not the overall power of the beam but the power per area, usually expressed as Watts/cm^2, or fluence. Because the local propagation direction of the beam with respect to the crystal axes is vitally important for maximizing second harmonic generation inside a nonlinear crystal, it is usually a good idea for the entire confocal length of the focused beam to be contained inside the crystal. This ensures that the wavefronts of the exciting beam are essentially plane wave like in nature. A suitable choice of initial beam size, beam waist and confocal length (or Rayleigh range, which is half the confocal length), for a given crystal length, which remains to be determined, can be ascertained in the vast majority of cases by examining the equation for focusing of a Gaussian beam. The relationship between the Rayleigh range and the beam waist is

$$z_0 = \frac{\pi w_0^2}{\lambda} \tag{4.24}$$

where w_0, the beam waist, is the beam radius at the point of focus, λ is the photon wavelength and z_0, the Rayleigh range, is the distance from the point of focus to the point where the beam radius w equals $w_0\sqrt(2)$.

The third method employed in the present experiment for increasing the power of the incident beam is to use a mode locked laser. A mode locked laser outputs the same time averaged power as a continuous wave (CW) laser but with the beam now arriving in short, discrete pulses rather than continuously. In this fashion the total power is maintained but the peak power is greatly increased and continues to increase as the duration of individual pulses and the number of pulses emitted from the laser per second is decreased. Mode locked Ti:Sapphire lasers typically emit pulses at a repetition rate of about 80 MHz with pulse durations of 2 to 3 picoseconds for long pulse lasers and tens to a few hundred femtoseconds for so-called ultra fast lasers. When using a pulsed laser for spectroscopy *and* for harmonic generation some care must be taken. Recall the energy-time Heisenberg uncertainty relation

$$\Delta E \Delta t \geq \frac{\hbar}{2} \tag{4.25}$$

for which ΔE and Δt are one standard deviation of a Gaussian spread in energy and time, respectively. Laser pulses of the type considered here are generally produced in a state very close to that of being transform limited, meaning in this instance that the relationship between the pulse duration and its energy spectrum are approximately governed by the inequality in Eq. 4.25 becoming an equality. CW beams, for example, have an essentially infinitely narrow spectral bandwidth and an infinitely long "pulse duration". Considering an example relevant to the current experiment, a pulse with a Gaussian intensity envelope in time of 300 fs, full width at half maximum (FWHM), has a FWHM spectral bandwidth of 6.08 meV (recalling that one standard deviation $\sigma = 2\sqrt(2\ln 2)$) whereas a pulse of 2 ps in duration will have a bandwidth of a mere .91 meV FWHM. On the other hand, let's say we have a 2 Watt laser with a rep rate of 80 MHz so that each pulse contains $2.5 \cdot 10^{-8}$ Joules of total energy. The 300 fs pulse will deliver over the FWHM of it's duration the equivalent of 83333 Watts whereas the 2 ps pulse will deliver a relatively soft 12500 Watts, more than a factor of 6 worse in terms of effective intensity and thus relative harmonic generation efficiency. Thus we find that as far as the generation of harmonics is concerned we generally gain energy resolution in our beam, which is highly desirable for high resolution ARPES, at the expense of the intensity of harmonics that can be generated. Fortunately ARPES does not require intense beams.

Lest this analysis appear to paint too dire a picture for the generation of harmonics with high spectral purity it should be recalled that many other factors come into play when considering the design of a harmonic generation scheme. Eq. 4.21 indicates that one can increase the amount of second harmonic produced simply by increasing the length of crystal through which the fundamental propagates, subject to the limits of phase matching for quasi phase matched processes. This fact is of much greater consequence for picosecond pulses than femtosecond pulses because very short pulses, containing very many more Fourier modes than their longer cousins, are much more prone to chromatic dispersion in dielectric media. Because femto pulses will rapidly broaden in a crystal, they can only be effectively used for harmonic generation in very short crystals, typically with lengths on the order of hundreds of microns, whereas pico pulses can typically be used with great effect in crystals of up to several millimeters in length. Further, nonlinear crystals typically have an energy bandwidth acceptance, meaning that for pulses that are too short some of the beam won't excite the second harmonic because it contains a significant number of frequencies outside of the phase matching range. On the other hand, longer crystals may not do better in cases for which absorption of either or both of the fundamental and harmonic beams is a serious issue as might be the case for the harmonic generation of a DUV laser beam. Also, for high efficiency conversion processes one must also consider the depletion of energy from the fundamental beam as it produces the second harmonic, for the quantum picture of second harmonic generation is that for every second harmonic photon emitted, two fundamental photons are necessarily absorbed. For the most part such considerations are most easily handled in the context of the sort of direct numerical simulations described below for particular crystals and wavelengths.

So far the simplest case of harmonic generation, in which a single laser beam or pulse is passed once through a crystal, has been considered. Not surprisingly there exist many other more sophisticated techniques for increasing the effective intensity of the fundamental laser beam seen by the nonlinear medium. These include various methods of intracavity harmonic generation in which a nonlinear crystal is placed directly into the fundamental laser cavity, as well as several other external cavity methods such as optical parametric amplification (OPA). Further, there are so-called three and four wave mixing processes such as sum and difference frequency generation in which two beams of different wavelengths, combined in a suitable nonlinear medium, produce a third beam at either the sum or difference of their frequencies. Such topics are beyond the scope of the present work.

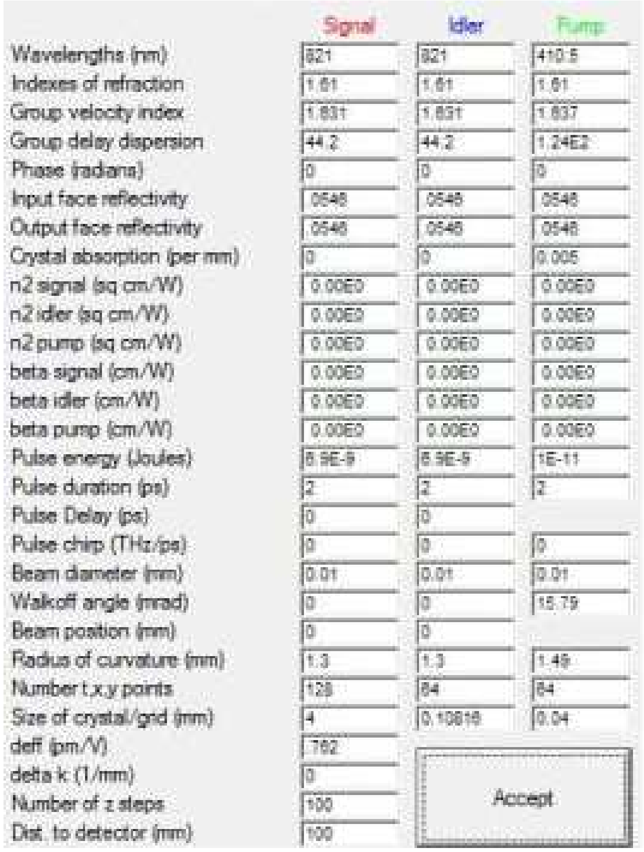
4.5 Simulation

The scheme proposed for laser ARPES at the NSLS was to generate the fourth harmonic of a Ti:Sapphire laser by first doubling the frequency of its near IR beam at 820 nm and then doubling the frequency of that second harmonic at 410 nm in a second crystal to reach the fourth harmonic at 205 nm. For all the reasons outlined above regarding crystal and laser beam properties it was not obvious from the outset that this would work, our laser being of the picosecond variety and the fourth harmonic, being in the DUV, subject to the severe strictures of strong absorption in most materials. It did not help that our custom Ti:Sapphire laser had an unusually high repetition rate of 105 MHz versus the usual 80 MHz repetition rate, resulting in a 24 percent decrease in the energy per pulse of our laser relative to the garden variety oscillator. To settle these issues as well as to aid the overall design of the double harmonic generation scheme both harmonic generation steps were examined using numerical finite difference calculations.

Calculations were performed using SNLO, a software package freely available from Sandia National Laboratory [33]. The package uses a split-step finite difference numerical method to simulate the mixing of beams in nonlinear crystals. The package includes codes for variously detailed levels of simulation including long and short pulse mixing in the plane wave approximation as well as more detailed, fully two dimensional codes for the accurate simulation of realistically focused beams. Additionally there are codes to aid the design of OPO and OPA systems. Included in the package are several useful tools for aiding the design of optical systems and the selection of nonlinear crystals, including a “phase matching calculator” that determines the beam polarization, launching angles θ and ϕ (for biaxial crystals), and nonlinear coefficients for a given nonlinear crystal, temperature and combination of beams, be it for second harmonic generation or sum or difference frequency generation. Given the rapid gains in computing power that have occurred even since the outset of this project it has become a trivial matter to run the full blown 2D simulations with a very tight spatial and temporal grid. Typical run times on a 2.4 GHz quad core PC were on the order of five minutes.

Though the first second harmonic generation step, converting from 820 nm to 410 nm, was performed using a commercial “harmonic generator”, simulation of the process nevertheless provided a good check and valuable insight into applying the above considerations in the laboratory. Given the relative ease of measuring the output characteristics of the second harmonic, being well within the blue end of the visible part of the electromagnetic spectrum, this check proved to be particularly straightforward. To avoid as much of the DUV absorption in the second step as possible, the first simulation was actually carried out at 821 nm. The crystal of choice for this step was Lithium Triborate (LBO). LBO is actually a biaxial crystal, meaning it has no optic axis as described above, and so has three different indices of refraction depending upon polarization direction and crystal axis. For LBO, optimal phase matching is achieved for propagation of the beam along the XY plane ($n(x) < n(y) < n(z)$) at $\theta = 90^o$ and $\phi = 29.6^o$.

All nonlinear crystal parameter calculations were performed assuming a crystal temperature of 300 K using the QMIX tool in the SNLO package. SHG occurs for the input of the 821 nm beam as an o ray with the second harmonic produced as an e ray. While the phase velocities of the e and o rays are matched for this configuration so that $\Delta k = 0$, there is a small mismatch in the group velocity indices of 1.631 and 1.665 for the near infrared (NIR) and blue rays, respectively, resulting in a small amount of walkoff. The walkoff angle is 15.79 mrad. Because we are using 2 ps pulses, the spectral bandwidth acceptance does not come into play here. Also included, but of minimal consequence for picosecond pulses, is the group dispersion that tends to broaden the pulse, in this case a factor of 44.2 fs^2/mm for the NIR beam and 123.7 fs^2/mm for the blue beam. Finally the $\bar{d}$ for this process is .762. The parameters used in the "2D-MIX-SP" simulation of this second harmonic generation process, using a 4 mm long crystal and 2 ps long pulses, is presented in Fig. 4.2, which is a screen shot captured from the actual run. Reflectances were calculated with Fresnel's equation assuming normal incidence to the crystal face. The beams are labeled according to standard usage as the "Signal", "Idler" and "Pump", in descending order of intensity. For Type I SHG in which a single input beam functions as both the signal and idler, and has a single polarization (o in this case), the initial pulse energy is split evenly between the two columns, everything else being identical as well. The simulation assumed a fundamental power of 1.5 W which, at 105 MHz, yields a pulse energy of $1.38 \cdot 10^{-8}$ Joules. The pulse energy of the second harmonic, the pump, is seeded with an energy at least an order of magnitude smaller than what is expected. As this number is made smaller the result for pulse energy should ultimately converge to a single number as the whole mixing calculation is ultimately self consistent. The radii of curvature of the input and output beams in air at the crystal surface are calculated using the FOCUS tool native to the SNLO package.



	Signal	Idler	Pump
Wavelengths (nm)	821	821	410.5
Indexes of refraction	1.61	1.61	1.61
Group velocity index	1.831	1.831	1.837
Group delay dispersion	44.2	44.2	1.24E2
Phase (radians)	0	0	0
Input face reflectivity	.0546	.0546	.0546
Output face reflectivity	.0546	.0546	.0546
Crystal absorption (per mm)	0	0	0.005
n2 signal (sq cm/W)	0.00E0	0.00E0	0.00E0
n2 idler (sq cm/W)	0.00E0	0.00E0	0.00E0
n2 pump (sq cm/W)	0.00E0	0.00E0	0.00E0
beta signal (cm/W)	0.00E0	0.00E0	0.00E0
beta idler (cm/W)	0.00E0	0.00E0	0.00E0
beta pump (cm/W)	0.00E0	0.00E0	0.00E0
Pulse energy (Joules)	6.9E-9	6.9E-9	1E-11
Pulse duration (ps)	2	2	2
Pulse Delay (ps)	0	0	
Pulse chirp (THz/ps)	0	0	0
Beam diameter (mm)	0.01	0.01	0.01
Walkoff angle (mrad)	0	0	15.79
Beam position (mm)	0	0	
Radius of curvature (mm)	1.3	1.3	1.49
Number t,x,y points	128	64	64
Size of crystal/grid (mm)	4	0.10816	0.04
deff (pm/V)	.762		
delta k (1/mm)	0		
Number of z steps	100		
Dist. to detector (mm)	100		

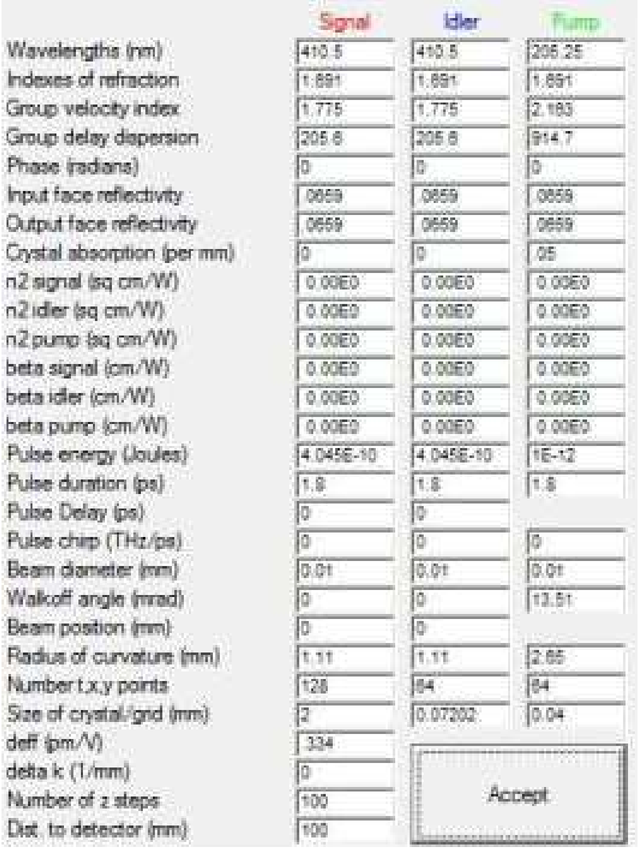
Figure 4.2: Initial parameters for the SNLO 2D short pulse mixing algorithm for second harmonic generation in LBO of a 410.5 nm beam from the 821 nm fundamental.

The results of the simulation are summarized in Fig. 4.4, panels a)-c). The second harmonic pulse energy yielded by this configuration was $8.09 \cdot 10^{-10}$ Joules giving rise to an average power of 85 mW at 105 MHz. This number is consistent with the typical SHG power observed at U13A using the setup described in Chapter 5. The observed power was between 75 and 100 mW, depending on the details of the coupling and the amount of power delivered to the crystal at the NSLS. Interestingly this mixing process gives rise to a small compression of the second harmonic pulse width relative to that of the fundamental as well as a corresponding broadening of the second harmonic spectral width. This new pulse width, as well as pulse energy of the second harmonic were used as inputs into the simulation of fourth harmonic generation by doubling the frequency of the second harmonic in a second nonlinear crystal.

The second SHG step was carried out using a Beta-Barium Borate (BBO) crystal. BBO was selected for it's relatively high transparency in the DUV region as well as the possibility for phase matching 410.5 nm and 205.25 nm light. It is a negative uniaxial crystal of the type described earlier in this chapter. The second harmonic is sent into the crystal as the o wave and the fourth harmonic exits as the e wave. The phase velocities are again matched with o and e indices of refraction of 1.691. The angular bandwidth did not come into play as an optical configuration was chosen so that the crystal length, 2 mm, was less than the Rayleigh range of the beam. The phase matching angle θ_M for this configuration is 85.9^o. Unlike the previous case, the group velocity indices were significantly different at 1.775 for the second harmonic and 2.163 for the fourth harmonic. Similarly the group delay dispersion is 205.6 fs^2/mm for the second harmonic and 914.7 fs^2/mm for the fourth harmonic. This results in a significant temporal delay, shown in Fig. 4.4d) and e), of the fourth versus second harmonic pulses. $\bar{d}$ for this process is .334. The input parameters for the simulation are summarized in Fig. 4.3 using a pulse length of 1.8 ps obtained from the output of the previous simulation. The output of the simulation is summarized in Fig. 4.4 panels d)-f). The output pulse power was $5.26 \cdot 10^{-12}$ Joules or .552 mW average power at the 105 MHz repetition rate. Taking into account losses from optics downstream of the crystal, this number is consistent with a photoelectron yield between 20 and 50 times greater than what is observed on gold and Bi2212 samples examined identically with the laser and with the U13UB undulator synchrotron beamline, whose output power at 21 eV is equal to or less than .01 mW.

4.6 Outlook

Given the strictures imposed upon this experiment by the type of laser used and the distances overcome in shuttling the beam to the harmonic generation apparatus the method presented in this chapter of generating an ARPES-qualified DUV beam at high energy resolution can be judged a success. Further, the process is robust enough that even under dire circumstances it does not appear difficult to generate DUV beams of significantly greater power than what is available, for comparable energy resolution, from synchrotron beamlines. The advantage of the current scheme is its simplicity in utilizing two successive SHG processes, both of which are easy to implement and highly efficient in terms of conversion, even while using picosecond laser pulses in a single pass configuration. While one might gain power or energy bandwidth



	Signal	Idler	Pump
Wavelengths (nm)	410.5	410.5	205.25
Indexes of refraction	1.891	1.891	1.891
Group velocity index	1.775	1.775	2.183
Group delay dispersion	205.6	205.6	914.7
Phase (radians)	0	0	0
Input face reflectivity	.0659	.0659	.0659
Output face reflectivity	.0659	.0659	.0659
Crystal absorption (per mm)	0	0	.05
n2 signal (sq cm/W)	0.00E0	0.00E0	0.00E0
n2 idler (sq cm/W)	0.00E0	0.00E0	0.00E0
n2 pump (sq cm/W)	0.00E0	0.00E0	0.00E0
beta signal (cm/W)	0.00E0	0.00E0	0.00E0
beta idler (cm/W)	0.00E0	0.00E0	0.00E0
beta pump (cm/W)	0.00E0	0.00E0	0.00E0
Pulse energy (Joules)	4.045E-10	4.045E-10	1E-12
Pulse duration (ps)	1.8	1.8	1.8
Pulse Delay (ps)	0	0	
Pulse chirp (THz/ps)	0	0	0
Beam diameter (mm)	0.01	0.01	0.01
Walkoff angle (mrad)	0	0	13.51
Beam position (mm)	0	0	
Radius of curvature (mm)	1.11	1.11	2.65
Number t,x,y points	128	64	64
Size of crystal/grid (mm)	2	0.07202	0.04
deff (pm/V)	.334		
delta k (1/mm)	0		
Number of z steps	100		
Dist. to detector (mm)	100		

Figure 4.3: Initial parameters for the SNLO 2D short pulse mixing algorithm for second harmonic generation in BBO of a 205.25 nm beam from the 410.5 nm second harmonic of a Ti:Sapphire laser.

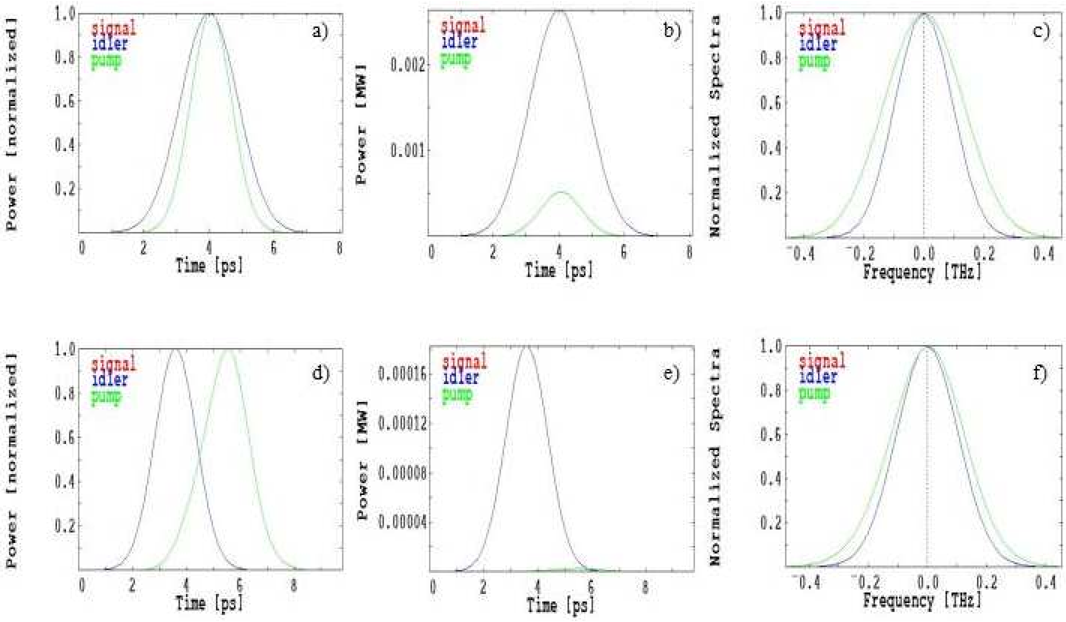


Figure 4.4: Plots of power and spectral output of SNLO simulations of second and fourth harmonic generation. a),b) are the normalized and absolute power, respectively, as a function of time and c) is the spectral distribution of the fundamental (Idler) and second harmonic (Pump) pulses after the harmonic conversion in LBO initialized with the parameters in Fig. 4.2. d)-f) same as a)-c) for the result of simulating the doubling of the second harmonic with parameters in Fig. 4.3, which were themselves obtained from the results of the previous simulation.

by using, for example, cavities of various types, including for CW beam conversion, the current method produces at least as many photons as can be used in an actual ARPES experiment and often more.

On the other hand, a photon energy of 6 eV is just high enough to perform ARPES and the technique would clearly benefit from higher photon energies. The technique outlined here, relying as it does upon phase matching to the fourth harmonic in a BBO crystal, is ultimately limited by the availability of nonlinear crystals. BBO can only phase match to 205 nm. Quasi phase matching in a three wave mixing process, mixing the fundamental with the third harmonic of a Ti:Sapphire laser, one can reach 193 nm (6.5 eV) with BBO, though this scheme is both much more difficult and much less efficient than the one outlined above. More encouragingly, efficient SHG phase matching down to 177 nm (7 eV) has recently been demonstrated in the new nonlinear crystal $KBe_2BO_3F_2$, (KBBF), and already applied to laser PES [34]. Unfortunately, this crystal is currently very difficult to synthesize and is not yet commercially available. A few research grade samples exist in the ARPES and optics communities but availability to the wider scientific community is currently limited. Barring the development of any more UV transparent nonlinear crystals KBBF is likely to remain the gold standard for nonlinear crystals that can be used for ARPES for quite some time.

Chapter 5

Experimental Apparatus

Thus far we have taken a mostly theoretical approach to superconductivity, ARPES, harmonic generation and the new methodologies of deconvolution and Fermi normalization. Where appropriate and necessary we have used the results of real experiments to make these discussions more concrete. We have also highlighted the ways in which some particular aspects of the experimental apparatus used in modern photoemission experiments directly impact the interpretation of data in light of these various theoretical topics. Before moving on to the presentation and interpretation of the new physics captured in the results of our laser ARPES experiments it is first necessary to explain in detail how these results were actually acquired. As we shall see below, low energy laser ARPES experiments add a layer of difficulty on top of what is already recognized as an extremely complex experimental undertaking. In the case of our own laser ARPES facility this difficulty was compounded by our ambition to implement the technique not in the warm confines of a laboratory as has been recently accomplished by other groups but at a long established synchrotron beamline endstation. While the unique challenges in this endeavor were immense, the potential payoff – the ability to use both synchrotron and laser radiation simultaneously or in tandem for the study of low dimensional and complex correlated electron systems – more than justifies the effort. The goal was to obtain an experimental apparatus that combines the photon energy tunability and large momentum space sampling provided by a high resolution synchrotron beamline with the high brightness, small spot size and momentum and energy space "zoom lens" properties of low energy lasers, all in a single facility. In short, we wanted one endstation with two beamlines.

5.1 Laser System

The minimum photon energy required for any ARPES experiment is set by the work function at a sample surface, discussed in Chapter 3, which has a value that seldom exceeds 5 eV and lies between 4 and 5 eV for the vast majority of solids. The exception to this rule are the elements comprising the first column of the periodic table. While it has long been known that the introduction onto a crystal surface of a small amount of Cesium or Potassium, for example, suppresses the work function of a surface down to 1 to 3 eV, their presence in a complex system such as BSCCO might overly complicate matters e.g. by changing the effective doping of the surface. Thus, at least while studying the cuprates, it is preferable to avoid this method when possible. The simplest method of producing a laser beam with photon energy in excess of 5 eV is the generation of even harmonics of a pulsed laser in nonlinear optical crystals. A brief explanation of this physics as well as simulations relevant to showing the feasibility of using relatively long pulses for this experiment were presented in Chapter 4. It suffices here to recall that with the crystals available to us we could obtain a beam with a photon energy of 6 eV, wavelength of 205 nm, by first frequency doubling a beam at 820 nm to 410 nm and then doubling that beam.

At the NSLS we were able to avail ourselves of the laser facilities previously constructed for use in time resolved, two photon infrared (IR) spectroscopy at the U6 laser hutch. That system initially consisted of a Mira 900-P Titanium doped Sapphire (Ti:Sap) passively mode locked oscillator from Coherent pumped by a Verdi V-5 5 Watt, frequency doubled Nd:YVO laser, also from Coherent. The oscillator was equipped with the so-called X Wave broadband optics set giving it a nominal wavelength tuning curve running from 700 nm to 950 nm, with optimum intensity falling fortuitously around 820 nm. Most Ti:Sap oscillators, when mode locked, generate transform limited pulses of tens to hundreds of femtoseconds length by dispersion compensation in a matched prism pair inside the cavity. Our cavity was of the somewhat unusual picosecond pulse length variety using a thin film grating optic to achieve the required dispersion compensation. Additionally, our oscillator had a custom cavity length; rather than running at the typical pulse repetition rate of around 80 MHz the cavity length was shortened to produce pulses at a rate of 105.2 MHz, exactly twice the repetition rate of the VUV ring running in seven bunch mode. This was required for timing operations at the IR beamlines of the VUV ring. While odd, this configuration in principle had several advantages for ARPES. First, having a high repetition rate and thus a relatively low intensity per pulse is desirable in an ARPES experiment as it allows the use of a higher flux beam overall before space charging near the sample surface becomes a problem. Second, though harmonic generation becomes much less efficient as the pulse length is increased, the transform limited nature of the pulses ensures a much narrower intrinsic bandwidth for the laser. For a 2 ps pulse length we should have a photon bandwidth of about 1 meV FWHM.

Because harmonic generation efficiency, especially in the single pass design outlined below, decreases linearly as the power per pulse is decreased and the length of a pulse is increased, the pump laser was upgraded to the 10 Watt Verdi V-10 model at the outset, raising the maximum power output by the oscillator to roughly 2 Watts at 820 nm from less than a Watt when environmental conditions allowed optimum tuning of the cavity optics. The power increase also vastly improved the stability of the mode locking, which we found empirically to suffer somewhat from the custom design of the cavity. Another unfortunate result of the custom cavity design was the introduction of a relatively severe ellipticity of the beam profile as it exited the cavity. This was corrected for by the introduction of a pair of cylindrical lenses immediately downstream of the oscillator.

Another good reason to increase the output power of the Ti:Sap laser is illustrated in Fig. 5.1. Because the lasers used in this experiment were all Class IV and therefore extremely dangerous in an uncontrolled environment like the experimental floor of the VUV ring, a great deal of effort had to be expended to carry out operations in a manner consistent with the safety guidelines laid out by BNL in general and the NSLS in particular. Because the IR experiments to which the laser was originally dedicated only required intense, pulsed light at a particular rep rate it was sufficient for those investigators to enclose the laser system, optics bench and other working parts in a light tight hutch with appropriate interlocks and other safety features, and distribute the light to various beamlines by fiber optic cable assemblies. Harmonic generation on the other hand requires intense, coherent beams which, owing to the distance between the U6 hutch and our own endstation and surroundings, presented something of a dilemma. The distance, about 200 feet as the crow flies, was too long to use fiber to carry the fourth harmonic directly to the sample chamber and would have in any event resulted in a beam quality much degraded from what was desired. Similarly there was no way to transport the fundamental beam or its second harmonic via fiber for later stage harmonic generation near the experimental chamber. Ultimately it was decided that the beam requirements, coupled to the limited table space available in U6 and the short coherence length of DUV light at 205 nm in air, rendered every option but the transport of the fundamental beam through free space, to the vicinity of U13, untenable.

As a result of these considerations a steel pipe was erected between U6 and U13A. It was composed of 15 sections of stainless steel tube, each section approximately ten feet long and 1.5 inches in diameter. Sections of pipe were joined end to end by quick flanges (QF) so that the pipe could ultimately be purged with nitrogen or evacuated as the need arose. It was also a simple way to keep the pipe securely light tight. Steel was used because the more common material for this application, PVC, was deemed unsafe and a health hazard in the event of fire. As we chose to transport the beam in a straight shot across the VUV ring, passing through twenty years' worth of cabling, pipes and equipment necessary to keep the storage ring functioning, the pipe diameter

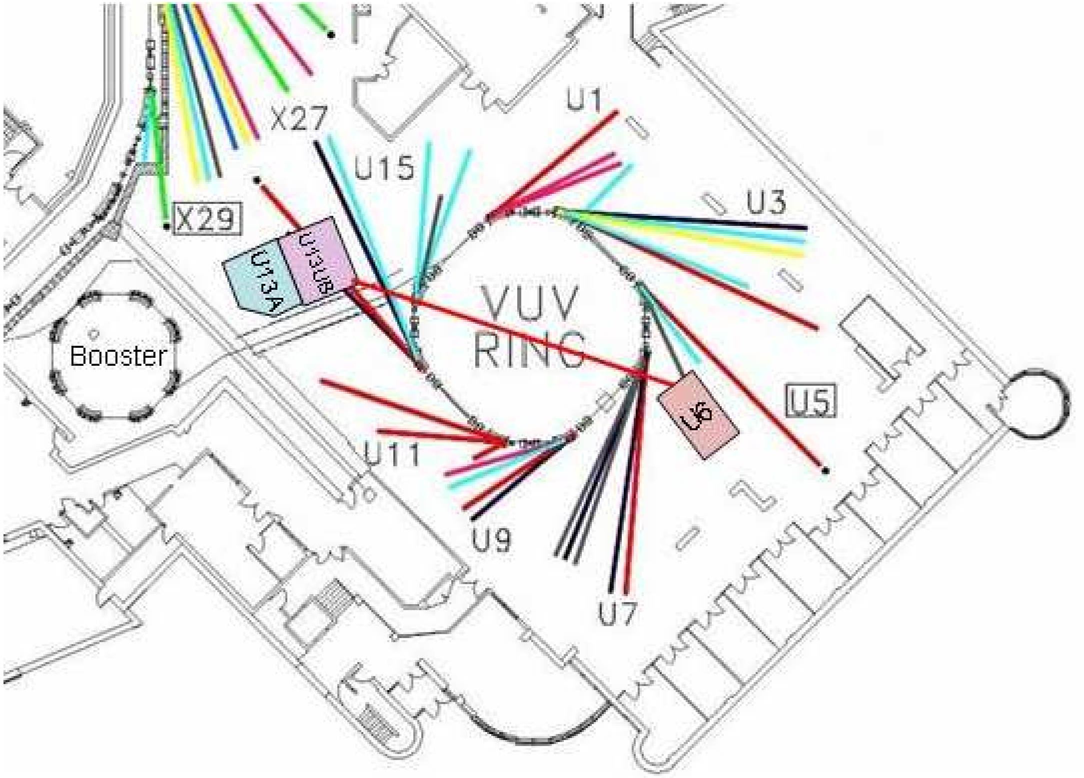


Figure 5.1: Schematic of the VUV ring experimental hall of the NSLS circa 2008. The laser hutches at U6 and U13A as well as the hutch enclosing U13UB are shown. The steel laser pipe connecting U6 to U13A is shown as the thick red line connecting the two hutches.

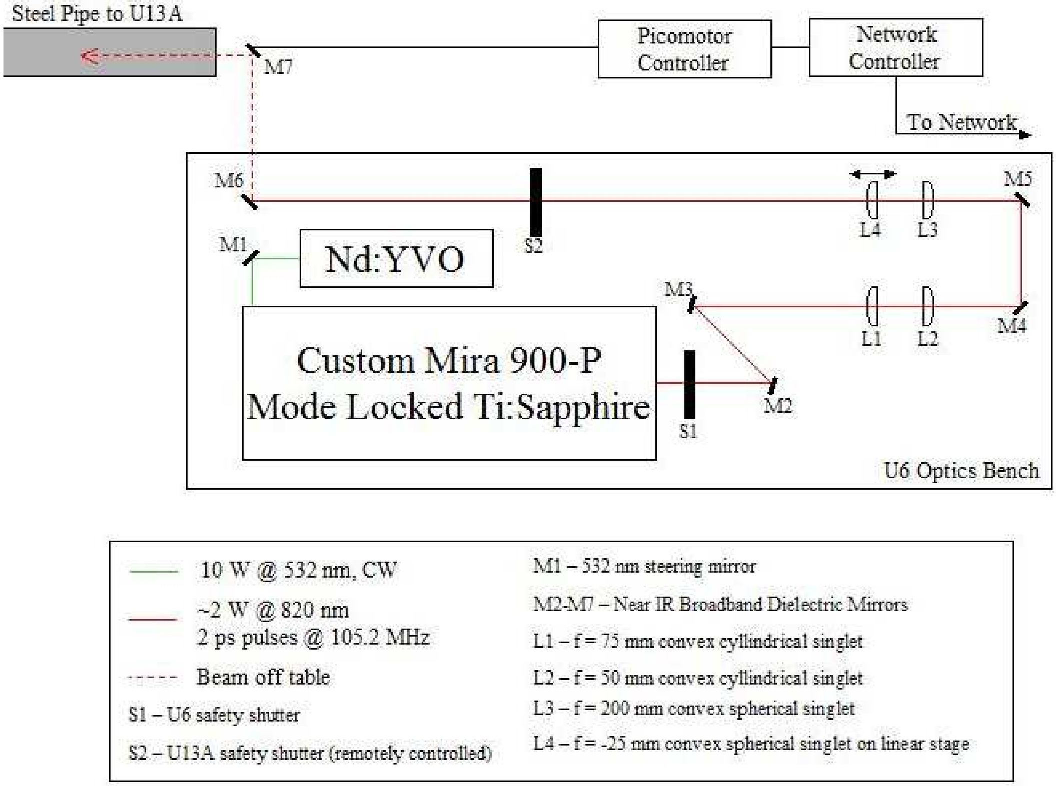


Figure 5.2: Arrangement of optics and lasers in the U6 hutch. The Ti:Sapphire oscillator output was first corrected for ellipticity and then expanded before being steered through the beam pipe to U13. The last mirror, residing at the pipe entrance, was moroized and controlled remotely as needed.

had to be extremely small and the pipe had to be suspended by various means roughly eight feet above the ground. Keeping a 1.5 inch wide pipe aligned well enough over 200 feet to pass a laser beam through, while passing over a particle accelerator and attempting to connect two optics benches hidden behind walls at each end is no mean feat. Nevertheless with a great deal of effort and assistance from our post-doc, Hongbo Yang, and one of the master electricians at the NSLS, Jim Lacey, this was accomplished while only inducing the VUV ring to crash once while we were inside it.

To get the fundamental beam through the laser pipe it was necessary to first expand it with a pair of spherical singlet lenses by a factor of 8 so that the Gaussian beam waist would fall roughly in the middle of the traverse. The beam was then directed vertically to a mirror mounted just in front of the U6 entrance to the pipe. The layout of optics in U6 is diagrammed in Fig. 5.2. Because steering the beam through the pipe over such a distance required great precision, and because the mirror was located near the ceiling of the U6 hutch well above the optical bench, tipping and tilting of the kinematic mirror mount was achieved using New Focus Picomotor actuators. These actuators were connected to a control box that supplied appropriate voltages to them. The control box was connected in turn to a networking unit so commands controlling the velocity, acceleration, activation and deactivation of the two mirror actuators could be sent remotely from the command prompt of a Telnet terminal on any computer connected to the network via the TCP/IP protocol. A Thorlabs thermopile power meter affixed to the end of the pipe at U13 provided a voltage signal proportional to the power. This signal was fed to a multimeter in U6 to facilitate the gross alignment of the beam. Once aligned through the main pipe section a broadband dielectric mirror was inserted before the power meter to steer the beam through the last two meters of pipe into the laser enclosure constructed at U13A (details given below) where harmonic conversion and other optical manipulations occurred. Once aligned the motorized mirror only required minor adjustments every week or so to keep the beam well aligned. This was accomplished by monitoring the beam power and profile on the optical bench in U13A while sending commands to the motor controller via the NSLS local network.

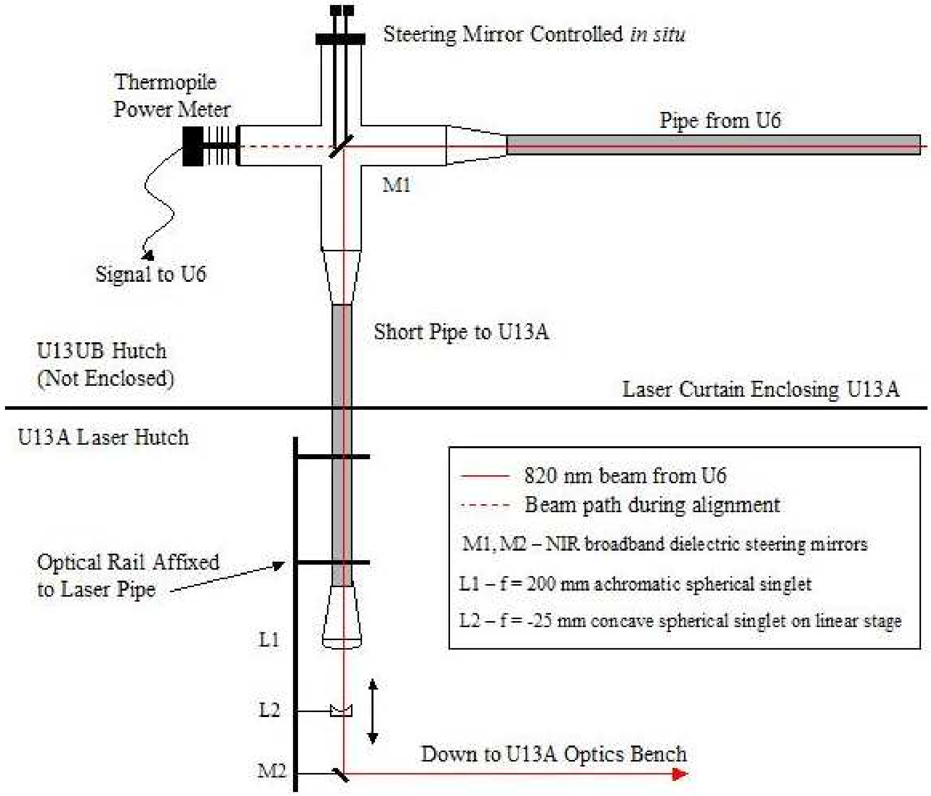


Figure 5.3: The turning mirror at the end of the pipe is removed during initial alignment of the 820 nm beam from U6 to U13. Alignment is established by monitoring the signal of a thermopile power meter. The mirror turning mirror is then replaced and the beam steered through the last two meters of pipe into the U13A laser enclosure.

The U13A end of the pipe was fitted with a 50 mm diameter achromatic doublet lens epoxied vacuum tight to a QF flange adapter. An optical rail affixed to this part of the pipe was fitted with a concave spherical singlet on a linear stage and a stearing mirror to bring the beam down to the table. This arrangement, diagrammed in Fig. 5.3, effectively prevented any air currents from moving through the pipe as it was only open at one end and also allowed for the recollimation of the beam at U13A so it could be easily manipulated with standard one inch optics. Once back on the table at U13A the beam typically had a diameter of 2 mm FWHM and an average power of 1.5 Watts at 820 nm. Though it had to travel a great distance the beam position and intensity was stable for periods on the order of a week before requiring correction. There was no detectable motion in the beam due to vibration induced by, for example, the low frequency sonic noise that permeates the facility due to the presence of vacuum pumps and experimenters at other beamlines.

The beam path and optics for harmonic generation in U13A are diagrammed in Fig. 5.4. After steering the 820 nm beam onto the table it passes through half waveplate to clean up the polarization and ensure it is horizontal as it enters the second harmonic generation stage. The Inrad harmonic generator (HG) designed to accompany the Mira Ti:Sapphire oscillator is used in the frequency doubling of the fundamental beam. Inside the HG the fundamental beam is focused through an LBO crystal by a spherical broadband dielectric mirror with a 50 mm focus. Because harmonic generation using picosecond length pulses is much less efficient than with femtosecond pulses relatively long crystals are used; the LBO crystal was 4 mm long (in the direction of the beam) and cut to 38.6 degrees. This choice of angle for the LBO crystal was dictated in the design of the harmonic generator by the desire to be able to achieve reasonably efficient phase matching over a wide range of wavelengths. Phase matching in the LBO crystal between the fundamental and second harmonic beams was achieved by angle tuning the crystal and then optimizing its position in the fundamental beam waist. The fundamental and second harmonic beams, exiting the crystal collinearly, were collimated by a second broadband dielectric spherical mirror. The fundamental and second harmonic beams were then separated by a thin film dichroic mirror. The fundamental was sent to a beam dump and the second harmonic, vertically polarized, was allowed to exit the harmonic generator. For a typical input power of approximately 1.5 W in the fundamental beam we routinely achieved second harmonic powers between 75 and 100 mW.

After steering the 410 nm beam to the fourth harmonic generating section of the setup the polarization was rotated in a second waveplate to a horizontal orientation. This was done so the polarization of the fourth harmonic at 205 nm would be vertical (s polarization) and the 410 nm beam horizontally polarized (p polarization) before entering the last set of optics for reasons that will become clear shortly. The 410 nm beam, after rotating its polarization, was focused through a BBO crystal by a 100 mm focal length lens. This focal length was chosen so the beam waist would lie entirely within the BBO crystal while maintaining a reasonably small beam diameter at the focus. The beam waist at the crystal position was measured to be 10 μm by a razor blade on a micrometer driven translation stage. The BBO crystal was 2 mm long and cut to 56 degrees from the optic axis so the impinging and exiting beams would lie close to the crystal surface normal. The exiting second harmonic beam and

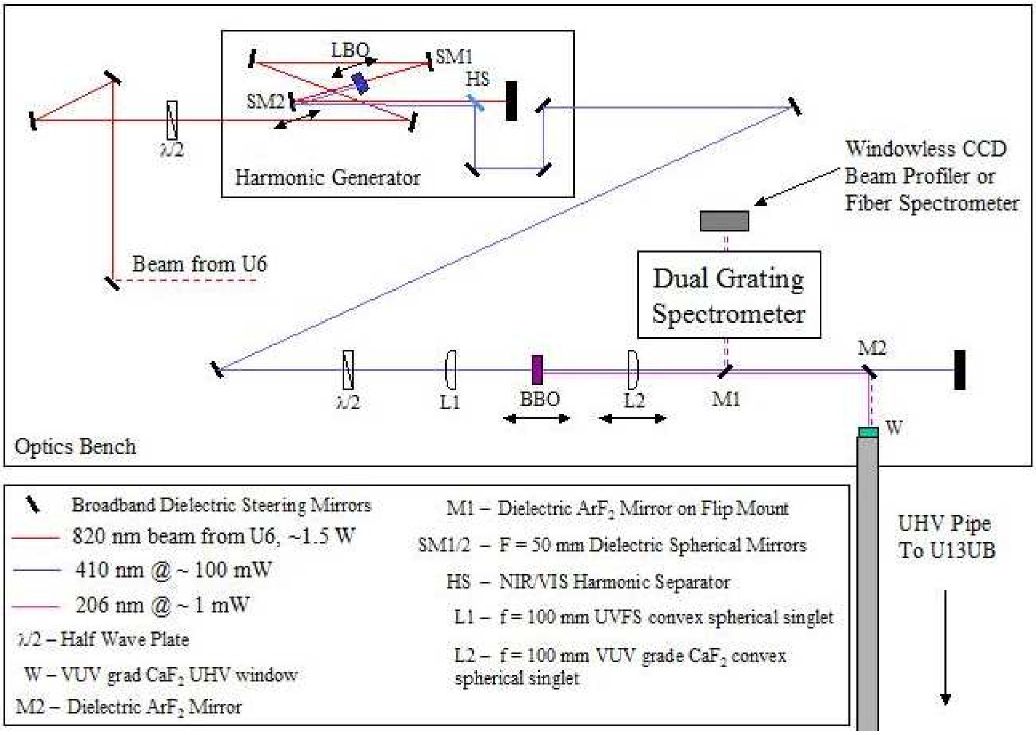


Figure 5.4: Layout of optics for generation and characterization of second and third harmonics of the Ti:Sap fundamental beam. An enclosure (not shown) was installed over the optics bench to limit dust and air currents that induced noise in the second and fourth harmonics. Waveplates ensured that the fourth harmonic was s-polarized, and the second harmonic p-polarized, with respect to the ArF_2 mirrors. This arrangement ensured their function as effective dichroic mirrors.

fourth harmonic beam at 205 nm were recollimated by a VUV grade CaF_2 spherical singlet lens mounted on a linear translation stage. The BBO crystal itself was mounted on a combination linear translation and precision rotation stage (not shown in the figure) to facilitate optimum angle phase matching and positioning of the crystal in the beam waist.

Working with 205 nm laser radiation, deep ultraviolet (DUV) radiation, versus vacuum ultraviolet (VUV) radiation presents some unique challenges. Unlike more common laser wavelengths in the IR, NIR and visible parts of the spectrum, detection of the beam can be problematic. Wavelengths longer than the visible are easily viewed with “night vision goggle” IR viewers. Unlike second harmonic generation, which is easily confirmed by the appearance of visible, in this case blue, light, DUV light is both invisible to most UV viewscreens and is easily swamped by the visible signal when attempting to generate and optimize the fourth harmonic. Furthermore, one’s choice of materials with which to construct even the most basic optical elements is extremely constrained and optics with antireflection (AR) and high reflection (HR) coatings for these wavelengths very difficult to come by. It was found that the simplest, most reliable and cost effective method for separating the visible from the second harmonic was to utilize the reflection, transmission and polarization dependent properties of mirrors originally designed for 193 nm eximer lasers. Though designed for 193 nm, the optic still works exceptionally well as a reflector for s polarized (vertical in our case) radiation at 205 nm and reflects p polarized radiation at 410 nm exceptionally poorly. Thus we are able to use the optic as a “poor man’s” dichroic mirror for separating the second from the fourth harmonic. In practice there remains a small amount, less than 1 mW, of visible radiation. This residual blue light, collinear with the DUV radiation, proves to be of enormous practical use when lining up and focusing the beam through the remainder of the apparatus, into the experimental chamber and onto a sample. It is easily viewed safely through a CCD camera whereas the DUV beam is not.

While the residual blue light is very useful it still poses difficulties with regard to detection and optimization of the fourth harmonic. That is because the quantum efficiency of most commercially available photon detectors and power meters is much higher in the visible than the DUV, and because at best we do not expect more than about .5 mW of DUV radiation anyway, another means of detecting the fourth harmonic is necessary. We thus use another ArF mirror mounted on a flip mount just after the recollimation lens (L2 in the figure) to direct the beams into a home made dual grating imaging spectrometer that can completely separate the visible from the DUV beams. The design and characterization of the spectrometer is discussed below. Detection and optimization of the fourth harmonic was achieved by first aligning the residual second harmonic beam through the spectrometer with the entrance slit wide open and onto the beam profiler using the long wavelength grating. Switching from long to short grating halved the wavelength to which the spectrometer was set, in our case from 410 nm to 205 nm. With the alignment thus assured, it was a simple matter to adjust the angle and position in the beam waist of the BBO crystal while observing the fourth harmonic output on the profiler. The harmonic conversion down to 205 nm was thus optimized by examining the shape and intensity of the image produced at the CCD without being swamped by signal from the second harmonic. A typical example of the

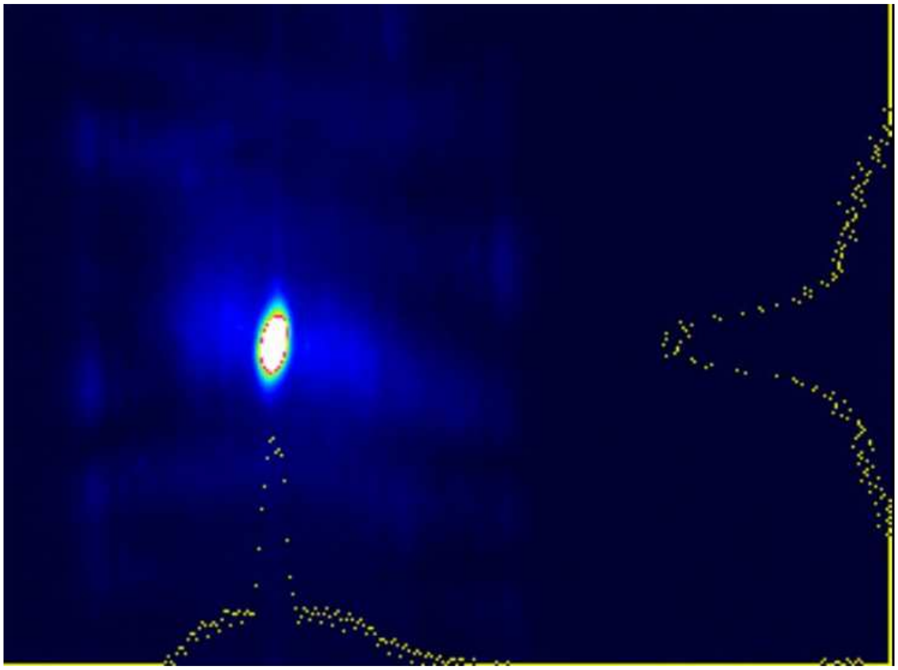


Figure 5.5: Image and real time x and y profiles of the fourth harmonic beam taken with laser beam profiler. The fourth harmonic beam has been separated from the second harmonic beam by the grating monochrometer.

resulting fourth harmonic beam profile is presented in Fig. 5.5.

After fourth harmonic generation has been optimized the first ArF mirror is flipped down and the beams are separated in the second, fixed ArF mirror. Because class 4 laser radiation must always be contained on the experimental floor and because the fourth harmonic is particularly sensitive to absorption by air and dust it is launched immediately into the laser beamline through an AR coated VUV grade CaF_2 window. The window was made by securing the coated optic in a CF flanged quick connect. The seal between the quick connect and the optic was formed by a Viton o-ring. The laser beamline and its relationship to the U13UB beamline and endstation, as well as the U13A laser hutch, is diagrammed in Fig. 5.6. The beamline is constructed of two main sections that can be isolated from eachother and from the experimental chamber by means of two gate valves, V1 and V2 in the figure. The beamline is constructed primarily of 1.5 inch diameter stainless steel tubes connected by 2 3/4” Conflat (CF) flanges. The first section comprises of the CaF_2 window, a standard ion gauge, pump out port for a removable turbo cart and a 50 l/s ion pump which is sufficient to maintain a base pressure of 10^{-8} Torr in the whole beamline. This section passes through the U13UB beamline and intersects a five way cross mounted on a breadboard. The cross hosts the next beam pipe section, two screened, standard viewports used for alignment and a 1” diameter VUV enhanced aluminum mirror. The mirror, which steers the beam through the last section of beamline into the experimental chamber, is attached to a micrometer driven linear translator used to move the mirror in and out of the beam path during alignment, a rotary platform and a micrometer driven tip-tilt mechanism for fine adjustments. The beam is aligned through the first section by centering it in window W1 with the mirror withdrawn. The mirror is then inserted and the beam is steered down the last beamline section. The last section of beamline is equipped with another pumpout port so the sections can be independently pumped out and baked during commisioning, a cold cathode ion gauge that can interface with the synchrotron beamline vacuum interlocks, and a paddle that can be inserted into the beampath for observation through a standard viewport. The beamline is isolated from the endstation by a second gate valve, V2, and a hydroform bellows that allows adjustments in the positioning of the endstation relative to the synchrotron beamline, which is also equipped with a bellows for this purpose, while maintaining connection to the laser beamline.

The last optic in the system, a 100 mm focus VUV grade CaF_2 spherical singlet lens, is mounted in a lens tube inside the chamber. The lens tube is anchored to a port alignment mechanism that allows small movements of the lens about the axis of the beamline to correct for small misalignments of the beamline relative to the sample position without moving the aluminum mirror. The lens alignment mechanism is in turn connected to a linear drive used to adjust the distance of the lens from the sample to optimize the focus. The main advantage of having the final lens so close to the sample inside the chamber is the ease with which a small spot size for the fourth harmonic – a mere ten microns – can be achieved. This is desirable because the angular resolution of the electron spectrometer depends on having a pointlike source. We can thus improve our angular resolution by using a smaller spot size than is typically achieved at the synchrotron, whose beam presents an elliptical profile on the sample with a vertical width of about 300 microns and a horizontal width defined by the exit slit of the beamline monochrometer ranging from 100 to 300 microns. Another good reason for using a small spot size is the presence in some samples, typically ones difficult to grow as well ordered single crystals, of spatially small domains.

The chromatic dispersion of the lens ensures that the residual blue beam is supremely unfocused when the DUV beam is focused properly on the sample. In order to avoid problems related to charging of the lens by the profusion of photoelectrons emitted from a sample it is recessed 25 mm in the lens tube behind a copper plate with a 10 mm diameter hole cut in the middle through which the beam passes. Besides absorbing most of the electrons emitted in the direction of the lens it can be shown that when a charge is placed behind a

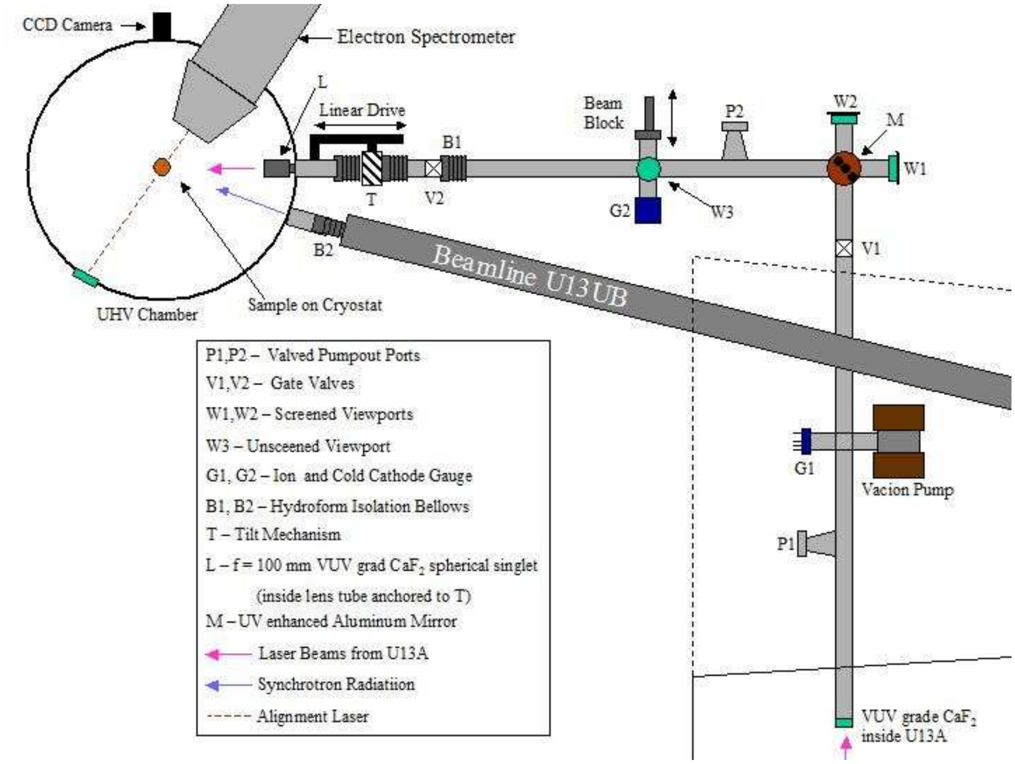


Figure 5.6: Laser beamline.

conductor with a circular hole cut in the middle the field on the other side of the hole from the charge is exponentially damped with distance from the plate[21]. The whole lens tube assembly and field plate are coated with graphite to match the work function of the graphite coated electron spectrometer. Finally, the lens tube is perforated with a half inch diameter hole on its side so it will vent properly under vacuum. While this necessarily connects the UHV environment in the endstation at $5 * 10^{-11}$ Torr to the 10^{-8} Torr vacuum in the laser beamline, the smallness of the link between the two ensures that the endstation vacuum is not adversely affected.

Alignment of the sample to the laser and the electron analyzer is a particularly tricky business when operating in the open environment of the NSLS experimental floor as all porthole windows on the chamber had to be made light tight. During normal operations with synchrotron radiation, assuming that the chamber and electron spectrometer are already well aligned with respect to the synchrotron beam by using the visible light of zero order specular reflection from the beamline's monochrometer. Such direct observation of the laser beam through the endstation chamber viewports is forbidden because of the safety concerns presented by the use of class 4 laser radiation. To get around this problem safely we mounted a CCD camera with 4x magnifying lens (shown in the figure) to an adapter that allowed us to mount it directly on a 2 3/4" viewport. The camera image was fed to a monitor on which samples and laser radiation could be clearly seen. Alignment was augmented by a a class 2 diode laser shot through a viewport on the back of the analyzer, through the entrance slits and onto the center of a viewport opposite the analyzer entrance. Because of the strong chromatic dispersion of the final lens for 410 nm light versus 205 nm light, the camera image could only facilitate alignment and not focus. Focusing and fine position tuning of the DUV beam and sample relative to the analyzer was accomplished by monitoring the transmission mode electron spectrum in real time. Transmission mode gives the position of electron emission versus kinetic energy rather than angle verses kinetic energy. By minimizing the width of the transmission spot and centering it on the detector we are able to align beam, sample and analyzer as well as optimize the focus of the beam and therefore the emission spot size. The ease of focus and alignment is especially important given the vast distance our laser must travel from U6; from experiment to experiment it is often necessary to tweak or otherwise work on the alignment. Also, because the focal length of the final focusing lens is relatively short, in order to get the small spot size we desire, the depth of field of the focus is extremely small rendering easy, precision adjustment of the lens an absolute necessity.

5.2 Photon Spectroscopy

Characterization of the the spectral and spatial characteristics of the 205 nm beam, while useful for ensuring the quality of the beam used in photoemission experiments, does not warrant the expense of a dedicated new commercial system with the sensitivity and resolution to perform this single task. On the other hand, the US national laboratories are well known for their reserves of aging yet entirely serviceable equipment that can be repurposed or otherwise modified with minimal effort or expense for new experiments. In this spirit a reasonably high resolution imaging spectrometer was constructed on the optical table in the U13A laser enclosure.

The overall layout of the system is diagrammed in Fig. 5.7. Spectral dispersion is accomplished with the diffraction gratings contained in an old Jarrell Ash model 82/410 dual grating monochrometer/spectrometer donated to the cause by Larry Carr at the NSLS. The device has slots at the beam input and output ports that accept slits of various widths. Sadly most of the slits for this device were lost over the years and the only remaining pair, both with slit widths on the order of millimeters, were insufficient to the task. Further evidence of the spectrometer's provenance from wide bandwidth, low resolution applications is observed in its mechanical construction; the grating is posi-

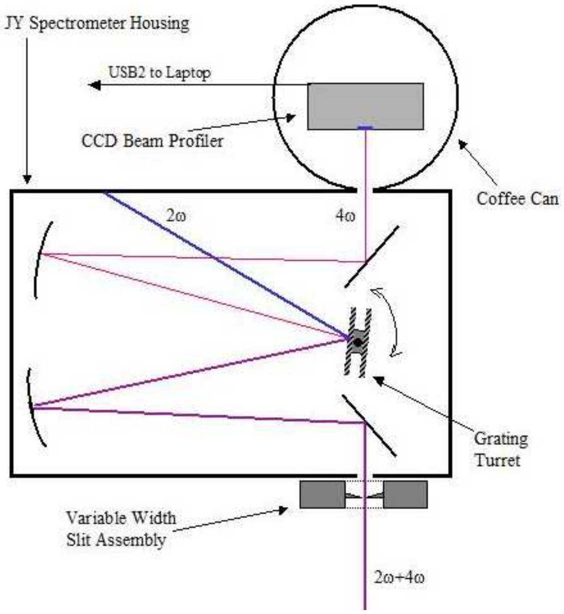


Figure 5.7: Layout of home built imaging spectrometer.

tioned by turning a large knob on the side of the housing that simultaneously drives a dial marked out in steps of .2 nm. Another knob on the housing flips the grating turret between the high and low blaze gratings. Nevertheless the absolute wavelength calibration of the spectrometer is more or less intact as is evidenced by its ability to reliably sort the second and fourth harmonics described in the previous section.

In olden times the lack of an exit slit, or rather the appearance of an exit aperture on the order of 2 cm, would render the device useless as the traditional method of measuring a spectrum involved a measurement of total beam intensity as the grating was swept through many wavelengths. To render the absence of an exit slit moot we located a windowless CCD based laser beam profiler (Newport model LBP-1 USB2 beam profiler) at the spectrometer exit port. The beam profiler is controlled by a computer that allows real time adjustment of electronic shutter speed, frame averaging, gain and beam profile characteristics as well as a convenient method of storing image data. The LBP-1, when used without a window, is sensitive down to 193 nm and so is well suited to our task. To keep the stray light observed by the beam profiler to a minimum a large coffee can with a slit cut out to let in light was put over the beam profiler (upside-down of course) and positioned so the entrance aperture of the can abutted the exit slit of the spectrometer. The inside of the coffee can was coated in graphite to minimize internal light scatter. Ultimately, as with the Scienta analyzer for electrons, one does not require an exit slit if one has a position sensitive detector and, with such a detector, one can acquire much more information much more rapidly anyway.

Because the center wavelength of the Ti:Sapphire fundamental beam and its harmonics are easily and precisely determined with the Ocean Optics fiber spectrometer it was not necessary to fine tune the home built system for absolute value measurements of wavelengths. All that is desired is a precise determination of relative wavelength across the detector so that one can convert the width of a feature measured on the CCD to a width in energy. This calibration was accomplished by observing the position on the detector of a small feature of light scattered off the diffraction grating as a function of wavelength setting on the spectrometer. Because the CCD chip is extremely small - about 1 cm across - relative to the path length between the grating in the spectrometer and the detector - about 50 cm - and because the change in wavelength on the detector as a function of spectrometer setting we wish to observe is relatively small, we expect the small angle approximation for motion of the

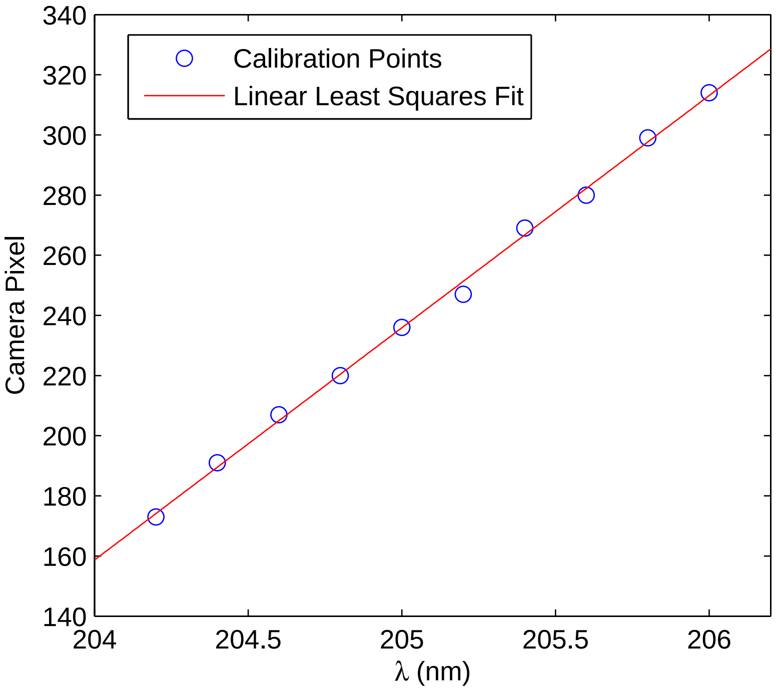


Figure 5.8: Calibration plot for home built imaging spectrometer. The λ points are read from the Jobin Yvon spectrometer and the pixel value from stored camera images taken at each wavelength calibration point.

beam across the detector to hold and thereby produce a linear relationship of pixels per nanometer. As demonstrated in Fig. 5.8, this is true and we derive a calibration constant for our device of 77 pixels/nm or 0.013 nm/pixel in the vicinity of 205 nm. At 205 nm this works out to a theoretical maximum resolving power $\lambda/\Delta\lambda$ of about 15600 for emission observed to land on a single pixel or, put another way, an energy resolution of about 0.38 meV.

Of course achieving this high resolution in a real measurement depends upon having an entrance slit comparable to the pixel size – in this case about 8 μ – while maintaining enough signal for the apertured beam to be detectable. In the absence of an entrance slit, the whole beam profile is imaged almost untouched by the dispersion of the grating. Without an entrance slit, in what might be called the beam imaging mode of the instrument, a limit on the resolution is set by the beam diameter. To operate the instrument in a truly spectroscopic mode it is necessary to provide a very narrow entrance slit. This was accomplished by situating a post mountable adjustable slit mechanism in front of the spectrometer's original entrance aperture. With this set of slits, which unfortunately are not knife edges, we were able to measure an upper bound on the bandwidth of the DUV beam of .01 nm or 2.5 meV. Presumably with better slits this number can be brought down further.

It is thus possible to establish the existence and optimize the intensity of the fourth harmonic beam with the instrument in imaging mode and to then close up the slit as much as possible in order to perform a spectroscopic measurement. The variable slit mechanism also has the desirable effect of blocking much of the residual second harmonic beam launched into the spectrometer even the scattered light of which can obscure observation of the much weaker fourth harmonic.

5.3 Vacuum System

Because the laser beamline was constructed to interface with an existing endstation a few, though significant modifications had to be applied to the endstation itself. One major modification, already addressed, was the introduction into the experimental chamber of the laser lens assembly. Two other significant modifications were made in the course of this work that will be described shortly. It is first necessary to describe the rest of the system as it was found. For any electron spectrometer and experimental chamber designed to perform high resolution, low energy ARPES must meet two fundamental requirements. First, the vacuum must be impeccably good and second the sample and photoelectron environs must be kept devoid of any stray electric or magnetic fields. While either of these requirements might be met alone with not too much difficulty, construction of a system in which both are met simultaneously can be extremely difficult.

An exceptionally good UHV environment is necessary for ARPES because of the limited escape depth of photoelectrons originating from or near the sample surface. When a sample is cleaved or otherwise prepared to reveal an atomically clean, fresh surface it will only remain in such a pristine state until it is covered by residual atoms or molecules in the vacuum chamber. Assuming we are in the Knudsen flow regime, and assuming that all atoms and molecules present will stick to the surface of our sample (a reasonable approximation at deeply cryogenic temperatures) we can expect approximately one monolayer of surface coverage per second at 10^{-6} Torr, or 1 Langmuir (1 L) = 10^{-6} Torr-seconds. Thus to retain a sample surface free of contamination and fresh enough to emit the electrons we wish to study rather than those originating from contaminants, a pressure of at least 10^{-10} Torr is necessary for an experiment lasting at least several hours. This number varies from sample to sample due to the details of adsorbate bonding at the surface and the constituents of any residual gas in the chamber. Bi2212, for example, can live for days in a vacuum in the 10^{-11} Torr range and for many hours in a worse vacuum due to the “protective” inert Bismuth oxide layer exposed at the natural cleavage plane. Samples that are very surface sensitive on the other hand might give you only a few hours even under the best circumstances. Additionally, the presence of contaminants such as water may be particularly problematic if they interact strongly with a sample surface, whose chemistry might well be different from that of the bulk. Thus in order to perform experiments in a reasonable amount of time for the vast majority of samples it is necessary to maintain a vacuum at least as deep into the 10^{-11} Torr range as possible.

To achieve and maintain such a vacuum the chamber is equipped with an array of pumping systems operating most efficiently at different pressure ranges. A high vacuum environment is easily obtained by the combination of a 150 l/s turbomolecular pump and 300 l/s ion pump, both mounted directly to the chamber. This combination is sufficient to reach the 10^{-8} Torr range. To reach the 10^{-11} Torr range it is necessary to bake the entire chamber (including pumps) and analyzer at 150 C and 110 C, respectively, for upwards of a week. This exceptionally long bake out time is imposed upon us by the large amount of surface area introduced by the many successive layers of magnetic shielding (discussed below) both inside the main chamber and inside the analyzer. Below 10^{-10} Torr the turbo pump is not very effective and so is valved off from the chamber. To reach and maintain better pressure the ion pump in combination with the occasional use of a titanium sublimation pump is necessary. This combination effectively maintains a pressure of $5*10^{-11}$ Torr. During experiments that require wide ranging changes in sample temperature this system is augmented by a liquid nitrogen cold trap in the titanium pump. This is necessary because residual gasses in the chamber will accumulate on the cryostat at low temperatures. When the temperature of the cryostat is raised past the point that accumulated gasses desorb the result can be a sudden increase in chamber pressure that, if not compensated for, can quickly coat the surface of the sample and prematurely terminate the experiment. While the Scienta analyzer on our endstation is equipped with an additional pump out port to compensate for the poor vacuum transmission and large surface area inside the device, we found that it is sufficient to connect this port via a flexible UHV hose to the main chamber. Finally, great care is taken at all times either when the chamber is vented to atmosphere for maintenance or during the routine introduction of samples to avoid contamination by high vapor pressure substances, especially of the organic variety. Identification of contaminants as well as the remaining water vapor pressure and helium leak detection is facilitated by a residual gas analyzer mounted inside the chamber.

5.4 Magnetic Shielding

Low energy, high resolution angle resolved photoemission experiments are exquisitely sensitive to the magnetic field environment traversed by the photoelectrons. Magnetic fields do no work and therefore cannot alter the actual kinetic energy of an electron. However because the force exerted upon an electron depends upon its velocity through the magnetic part of the Lorentz force law, $\boldsymbol{F} = e\boldsymbol{v} \times \boldsymbol{B}$, where e is the electron charge, $\boldsymbol{v}$ is its velocity and $\boldsymbol{B}$ is the magnetic field strength in vacuum, and because the photoelectron spectrum generally contains electrons with a large range of velocities, the presence of stray magnetic fields can have devastating consequences for the experiment. Magnetic fields in the vicinity of the sample induce an energy dependent alteration in the photoelectrons' trajectories which translates into a very complicated and malefactorous variation in the best angular resolution achievable. Because hemispheric spectrometers gain their kinetic energy resolving power by sorting electrons into differing orbits according to their energies, a stray magnetic field inside the hemisphere will not only alter their measured emission angle but their measured kinetic energy as well. Further, because the entrance slit to the spectrometer determines not only the energy resolution of the spectrometer but the location and depth of the cut in $\boldsymbol{k}$ space sampled by the spectrometer, the presence of stray fields can also cause errors in determining the location in the Brillouin zone at which a spectrum is acquired. Finally, stray fields that alter electron trajectories along the length of the slit will result in an energy dependent tilting and distortion of the spectrum thereby rendering any measure of a state's momentum dependence useless.

For all these reasons the quality and integrity of the magnetic shielding provided for an ARPES experiment is crucial to its success. The shielding problem becomes much more severe as one decreases the overall kinetic energy of the photoelectrons. A simple finite difference calculation, the results of which are plotted in Fig. 5.9, reveals the magnitude of the problem. In this calculation we assume a beam of electrons is launched from a point source into the $\hat{y}$ direction with velocity defined by a range of realistic kinetic energies and zero initial velocity in the $\hat{x}$ direction. A constant, uniform magnetic field B_z is applied in the positive $\hat{z}$ direction. For kinetic energies ranging between .5 eV and 2 eV, the energy range relevant for laser ARPES, the overall displacement of the beam is, not surprisingly, strongly dependent upon the field strength. Worse still is the strong energy dependence of the displacement for a given field strength over the range, yielding a strongly energy dependent angular

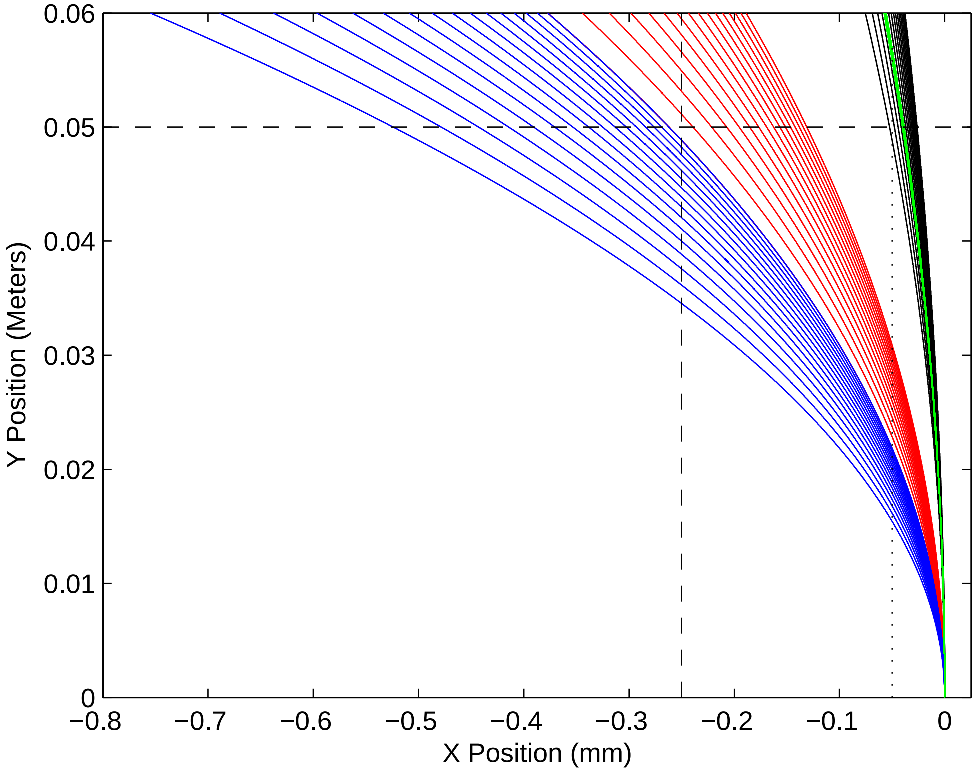


Figure 5.9: Finite difference calculation of electron trajectories normal to a magnetic field pointing out of the plane. Electrons originate from a point at (x,y)=(0,0) with initial velocities in the $\hat{y}$ direction. Trajectories of electrons with kinetic energies between .5 and 2 eV are shown for $B_z = 0.01$ Gauss (blue), 0.005 Gauss (red) and 0.001 Gauss (black). Electrons with kinetic energies between 21.5 and 23 eV in a 0.005 Gauss field – a condition similar to typical synchrotron experiments – are shown in green. The horizontal dashed line marks the distance from the sample to the entrance aperture of the SES-2002 electron spectrometer. The vertical dashed and dotted lines mark the half widths of a .5 mm and .1 mm analyzer entrance slit, respectively. Calculations were performed with time steps of 10^{-10} seconds.

resolution. This situation contrasts sharply with the situation at VUV photon energies. Also shown in the figure, for comparison, is the displacement of an electron beam with energies ranging between 21.5 eV and 23 eV in a 0.005 Gauss field. This situation is nominally representative of the fields and kinetic energies typically experienced at U13UB during synchrotron operations. Both the overall beam displacement and the range of displacements over the kinetic energy range are minimal in this arrangement. Clearly the angular (and thus momentum) resolution achievable in a laser ARPES experiment are highly dependent upon the ability to suppress magnetic fields over the flight path of the photoelectrons. While the orientation of fields in the chamber are essentially random, it is appropriate to consider the worst case for the best magnetic field we can achieve. Setting the goal of at least a 0.001 Gauss field for laser ARPES, we find that an entrance slit on the order of .5 mm width 10 cm from the source provides the narrowest meaningful cut through the angular distribution of photoelectrons.

The residual magnetic field generated by the Earth is on the order of 0.5 Gauss at its surface. On top of this, synchrotron facilities are filled with large and varied permanent and electrically driven magnets required for both experimental apparatus and the operation of the synchrotron itself. While we might reasonably expect minimal impact from these local fields because the magnetic dipole field falls off like $1/r^3$ in the far field, a sense of the difficulty of this problem was gained by observing the displacement of photoelectrons by up to several mm in real time in the spatial imaging mode of the Scienta analyzer at low kinetic energies while the booster ring and linac (Fig. 5.1) of the NSLS were run during injection into the VUV and X-Ray storage rings. In short, while synchrotron light sources are invaluable for many ARPES applications they are incredibly hostile environments and the hostility grows as the photon energy declines. Great effort must be expended to repel these deleterious fields.

The cornerstone of magnetic shielding for all parts of an ARPES experimental chamber and spectrometer is the judicious application of mu metal, a nickel-iron alloy that, when heat treated in a hydrogen atmosphere, obtains an incredibly high magnetic permeability μ. While the vacuum vessel of the Scienta 2002 electron spectrometer is fabricated from standard UHV grade, nonmagnetic stainless steel the electron lens and hemispheres are surrounded by two layers of 3 mm thick mu metal joined mechanically to form as complete a magnetic circuit as possible. Because the magnetic shielding properties of the mu metal degrade with the introduction of holes in the shielding through which fields can "bulge" and possibly reconnect, openings in this shielding are limited to those required for the entrance aperture to the lens, four electrical feedthroughs for the voltages operating the lens and hemispheres and the microchannel plate detector stack. The magnetic field inside the Scienta is assumed to be zero. This is difficult to check directly though it is observed to fall rapidly to zero upon inserting a magnetometer into the entrance of the lens assembly. The negligible magnitude of the field inside the analyzer is achieved largely through the quality and care of construction and through a design that minimizes field intrusion though apertures.

A somewhat trickier business is the shielding of the experimental chamber mated to the Scienta in which the sample and peripheral experimental equipment resides. The experimental chamber itself is based on a cylinder 18 inches in diameter and 3 feet in height with an assortment of different size necked flanges arranged around the circumference in three levels. The analyzer is mounted to a 10 inch flange on the middle level of the chamber. Samples, mounted on a cryostat, are introduced through a 1.5 inch hole in the top center of the chamber and lowered into the focus of the analyzer's electron lens. Magnetic shielding of the main chamber is accomplished with an overlapping pair of 2 mm thick mu metal shields lining the chamber walls with holes cut out in the locations of ports on the chamber which is itself also made of nonmagnetic stainless steel. The home made chamber shielding is mechanically mated by bolts flush to the shielding on the Scienta thus forming a magnetic circuit continuous around both the chamber and analyzer. Because of the many holes cut in the shielding for instrumentation and viewports allowing observation of the inside of the chamber, the strength of the magnetic field inside is not uniform. The field strength is however measured to range between 0.005 and 0.01 Gauss in the vicinity of the sample and the space between the sample and the entrance aperture of the analyzer. As demonstrated above, this level of shielding is adequate for high resolution ARPES experiments using synchrotron radiation as a photon source. Unfortunately, even at this level of shielding a small but noticeable energy dependent broadening of the electron beam was observed in transmission mode (i.e. imaging) spectra acquired with the laser. In angle resolved spectra this would translate to an unacceptable loss of angular and momentum resolution.

Our somewhat unorthodox solution to this problem was to supplement the passive magnetic shielding provided by the mu metal with a set of "Helmholtz" coils wound directly onto the chamber. Through some trial and error it was found that with two sets of coils, one wound about the top and bottom of

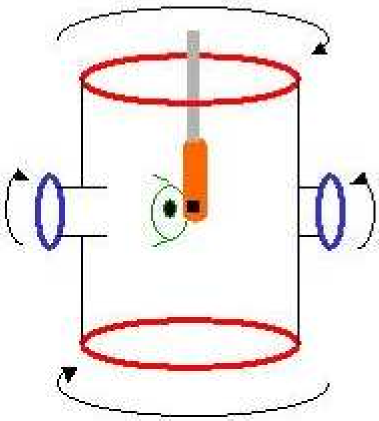


Figure 5.10: Arrangement of magnetic field shim coils. The vertical field canceling coils are marked red and the horizontal coils blue. The sense of the currents driving the coils are marked by arrows. The entrance to the Scienta analyzer is green. The position of the sample, mounted on a cryostat, relative to the coils and the analyzer is also shown.

the main chamber cylinder equidistant from the Scienta and another small set wound about two small flanges located ninety degrees from both the Scienta and the main chamber axis, it was possible to "push" the fields around in such a way as to null the field in the vicinity of the sample. The arrangement of the coils is diagrammed in Fig. 5.10. The large "vertical" coils were 18 inches in diameter and composed of 40 turns per coil. The small coils were five inches in diameter and also composed of 40 turns per coil. The coils were permanently affixed to the chamber by stainless steel shim bands spot welded directly to the vacuum vessel. The wire used to construct the coils was insulated by Teflon to they could be safely baked along with the rest of the chamber without needing to be removed. The coils were wired, like true Helmholtz coils, so current flowed in the same sense around each coil in a pair.

The leads from the coils were twisted and each set run from a separate DC power supply. The current driving each coil pair was calibrated by iterative manual adjustment while observing the reading on a flux gate magnetometer. It was found that the field in the sample region and between the sample and the spectrometer could be reduced to less than 0.001 Gauss in all directions by this procedure. The large coils require a current of 2.5 Amps and the small coils a current of 0.48 Amps. To check for alterations in the magnetic field environment as well as any drift in the system calibration the procedure was repeated after one year of operation and the same current values were found to yield the same vanishingly small fields. Measurements of the ambient magnetic fields around the chamber and analyzer with the coils in place and functioning yielded not obvious difference from the coil free environment. We are thus confident that our method of shimming the residual magnetic field inside the chamber with some well placed coils is highly robust and repeatable and serves as a flexible and cost effective way of improving upon the traditional magnetic shielding methods employed in ARPES.

5.5 Sample Handling

The mounting, cooling, manipulation and introduction of samples into the apparatus is fairly straightforward. Samples that require only *in situ* cleaving for preparation of a fresh surface (e.g. Bi2212) are affixed to a copper wedge cut to the azimuthal angle θ required to reach the specified part of the Brillouin zone for a given photon energy. The wedge face, typically with a surface area on the order of 3 cm^2, can accommodate several samples at once. The wedge is mounted as a tail piece with titanium screws to the bottom of a non-magnetic, UHV compatible liquid helium flow cryostat from Janis. A copper shield with an opening just large enough to show the face of the wedge is anchored to a thermal ballast several inches above the cold finger. The shield, which is cooled to at least liquid nitrogen temperatures by contact to the helium exhaust shroud of the cryostat, acts to dampen the amount of blackbody radiation from the world seen by the sample and cold finger, thus improving the system's performance at very low temperatures. The cryostat is equipped with a resistive heater and a calibrated Lake Shore thermal resistor sensor epoxied to the outside of the tail piece. The temperature can be set accurately to within a degree Kelvin with a Lake Shore temperature controller. The system can be run efficiently by cooling with helium gas from a dewar down to 10 K. By filling the bottom of the cryostat with liquid Helium the system can be maintained at about 4.5 K. By pumping on the exhaust, thereby lowering the vapor pressure of the liquid helium accumulated at the bottom of the cryostat, we have achieved temperatures as low as 3.2 K with a holding time on the order of 15 minutes.

The mounting of samples is often more of an art than a science. How this is done has a great deal to do with the properties of a given sample. Because the layers Bi2212 are held together by Van der Waals forces samples can be cleaved without a great deal of force. On the other hand, this same softness can, if care is not taken, cause the sample to flex or even crack upon cooling to cryogenic temperatures. Further, however one mounts a sample it is necessary to maintain good thermal and electrical contact between the sample and the cryostat. Traditionally samples are glued onto the wedge described above with an epoxy, either Torr seal or some kind of silver laced epoxy. However the best method for mounting thin, flat, easily cleaved samples is by far the "tape" method. In the version of this method developed for our system, a piece of double sided, UHV compatible, sticky silver tape is first mounted to the copper wedge with a very thin layer of silver epoxy in between to help it stick at low temperatures. The Bi2212 samples are stuck directly to this tape. The tape is strong enough that the samples can be "pre-cleaved" with Scotch tape to remove the weakest interlayer bonds. Then small aluminum posts are mounted on the samples with more silver epoxy. Finally the whole thing is spray coated with an aqueous graphite solution to cover the tape as well as provide a non-photoresponsive, even potential around the samples. Using this method we have found that the resistance between the sample surface and the cryostat is a mere few tens of Ohms. The rate of successful cleavage by this method is nearly 100 percent as it does not rely on one's skill at mounting very small crystals without the epoxy squirting out the sides grabbing the rest of the sample. The surfaces revealed are almost always very flat and mirror like. Another benefit of the tape besides ease of mounting and good electrical connection to the sample is that the tape appears to take up the mechanical strain that would otherwise be imparted to the crystal by a rigid epoxy upon the cooling to cryogenic temperatures and subsequent contraction of the copper tail piece.

The cryostat, with samples mounted, is introduced into the chamber after a 24 hour bake out in an isolated bellows section above the chamber. This section is pumped by a separate turbo pump. Because the cryostat must be able to be lowered from outside the main part of the chamber down to the center level at which the Scienta resides it is relatively long; one meter. The bellows is mounted on an XYZ stage with X and Y positions controlled manually with micrometers. The cryostat is affixed to a rotary platform at the top of the bellows allowing rotation of the cryostat and sample in the ϕ direction relative to the entrance slit of the analyzer. Because it is often useful to be able to introduce an electric potential between the sample and the analyzer for low energy ARPES the system was modified by inserting a high voltage ceramic insulating section between the cryostat and the rotary platform large enough for the cryostat to pass through. Smaller ceramic sections isolated it from its pumping unit. This configuration allows the whole cryostat to be electrically isolated without any special preparations for the sample. By applying a positive bias between the sample and analyzer it is possible to measure the energy difference between the Fermi level and the low energy cutoff of the photoelectron spectrum and thus derive a sample's work function. By floating the sample and introducing a picoammeter it is also possible to directly measure the photocurrent produced at the sample. Because the quantum efficiency of photoemission for different photon energies and samples can vary greatly this is a more useful characterization of laser photoemission than simply measuring the laser beam power. During normal measurements the sample is grounded.

5.6 Auxiliary Apparatus

In addition to the primary elements of the laser ARPES experiment described above the U13UB endstation is or can be equipped with a range of other surface science related pieces of equipment. The upper level of the chamber (above the Scienta) contains a reverse view Low Energy Electron Diffraction (LEED) apparatus and a wobble stick. The latter is used to hit pins epoxied to the top of samples thereby cleaving them. The LEED fires a collimated beam of electrons with energies usually in excess of 50 eV at a sample and images the diffracted electron pattern on a phosphor screen. This is useful for characterizing the surface structure of a sample (if, for example, it turns out to be different from the bulk) as well as for verifying the rotation angle of the sample about the ψ axis (as opposed to θ and ϕ). The LEED gun is mounted precisely ninety degrees in the ϕ direction from the electron spectrometer. By finding the ϕ angle relative to the LEED gun at which a change in electron kinetic energy produces no change in ϕ angle of the central diffraction spot the normal photoelectron emission angle is easily determined.

The analyzer level is equipped with an ion sputtering gun, usually used with an Argon source, used for cleaning the surfaces of non-cleavable samples such as metal crystals. For experiments requiring the deposition of thin metal films upon a sample substrate thermal and e-beam evaporators for various sources can be introduced through several auxiliary ports of the middle and upper levels. These are usually used in conjunction with a quartz crystal thickness monitor used to check the deposition rate. Also, as mentioned above, a flange mountable CCD camera adapted from an older Scienta analyzer no longer in use can be mounted on a 2 3/4 inch view port for remote visualization of the sample, a task especially important during laser operations when direct viewing of the inside of the chamber is prohibited for safety reasons. This camera, being sensitive to near IR wavelengths, is also useful for monitoring the onset of sample heating during experiments that require annealing of a sample for surface preparation.

Chapter 6

Lucy-Richardson Deconvolution and Fermi Normalization

6.1 Photoemission Above the Fermi Level

In Chapter 3 we explored the kinematic and quantum theories of ARPES and pointed out some of the considerations most relevant to interpreting the results of a photoemission experiments performed at very low photon energies. It was found that in general the detector intensity $I(\boldsymbol{k}_\parallel, \omega)$ was determined by Fermi's golden rule such that

$$I(\boldsymbol{k}_\parallel, \omega) \propto M_{f,i}^2 A(\boldsymbol{k}_{parallel}, \omega) f(\omega; T) \tag{6.1}$$

where the transition matrix element $M_{f,i}^2$ modulated the signal intensity, $A(\boldsymbol{k}_\parallel, \omega)$ gave the density of electronic states in the crystal and $f(\omega; T)$, the temperature dependent Fermi-Dirac distribution, enforcing the fact that ARPES can only probe the occupied part of the spectral function. The semicolon in the expression for the Fermi function reminds us that T is fixed for a given distribution in the energy variable ω. The proportionality absorbed the various constants in the problem as well as the solid angle distribution of photoelectrons $d\Omega/d\sigma d\omega$ sampled by our detector.

Two issues with this parametrization of photoemission spectroscopy were glossed over. The first issue, which is theoretical in nature, involves our definition of the spectral function. While the above statement that the spectral function gives the density of states for a system at both positive and negative binding energies is true, the way in which we enforce our probing of the occupied part of the spectral function is somewhat misleading. The full spectral function can be broken down into parts describing positive and negative frequencies corresponding to the advanced and retarded Green's functions from which they derive.

$$A(\boldsymbol{k}_\parallel, \omega) = A^-(\boldsymbol{k}_\parallel, \omega)_{\omega<0} f(\omega; T) + A^+(\boldsymbol{k}_\parallel, \omega)_{\omega>0}(1 - f(\omega; T)) \tag{6.2}$$

This form shows that what we were previously calling the spectral function A was actually A^-, the electron (occupied) part of the spectral function whereas A^+ is the hole (unoccupied) part of the spectral function. It also shows explicitly how A^- can, at finite T, display electron excitations above the chemical potential in proportion to the hole excitations below E_F described by A^+. It is this latter quantity that is probed, for example, in inverse photoemission spectroscopy and to some extant by two photon photoemission spectroscopy. In any event, we will from here out recognize that photoemission can only probe A^- and so revert to our former notation of simply calling the occupied part of the spectral function A. While $A(\boldsymbol{k}_\parallel, \omega)$ is what we seek to measure in photoemission, clearly what we actually measure then is $A(\boldsymbol{k}_\parallel, \omega) f(\omega; T)$. At $T = 0$ there is no mixing of electrons and holes and all single particle excitations reside below E_F. At finite T states above the chemical potential are filled in proportion to the fraction of electrons thermally excited to previously unoccupied states and so can be probed by photons. Thus, the density of states we actually measure in ARPES is

$$\tilde{A}(\boldsymbol{k}_\parallel, \omega; T) = A(\boldsymbol{k}_\parallel, \omega) f(\omega; T). \tag{6.3}$$

which, for finite T, includes states above E_F. The extent of electron-hole mixing can easily be determined for a give temperature by plotting $f(\omega; T)$, which we do for several temperatures in Fig. 6.1. Though states above E_F are filled in this way by thermally excited electrons, we still refer to them as unoccupied states, semantics notwithstanding. This is true so long as, in some sense, $A^+(T = 0)$ is not too different from $A^-(T > 0)$. So long as $T < T_c$ superconductors more or less fulfill this requirement. In any event, because the electron-hole mixing only extends to within $4k_BT$ of either side of the Fermi level, it is still true that $\tilde{A} = A^-$ so long as we only look at binding energies larger than this. Unfortunately, the excitations residing within this range of the Fermi level are usually those of greatest interest to solid state physics. What's more, as has become abundantly clear in the ARPES community over the last year or so, acquiring some understanding of the excitations above the Fermi level is crucial to resolving the physics of high T_c superconductors. Thus

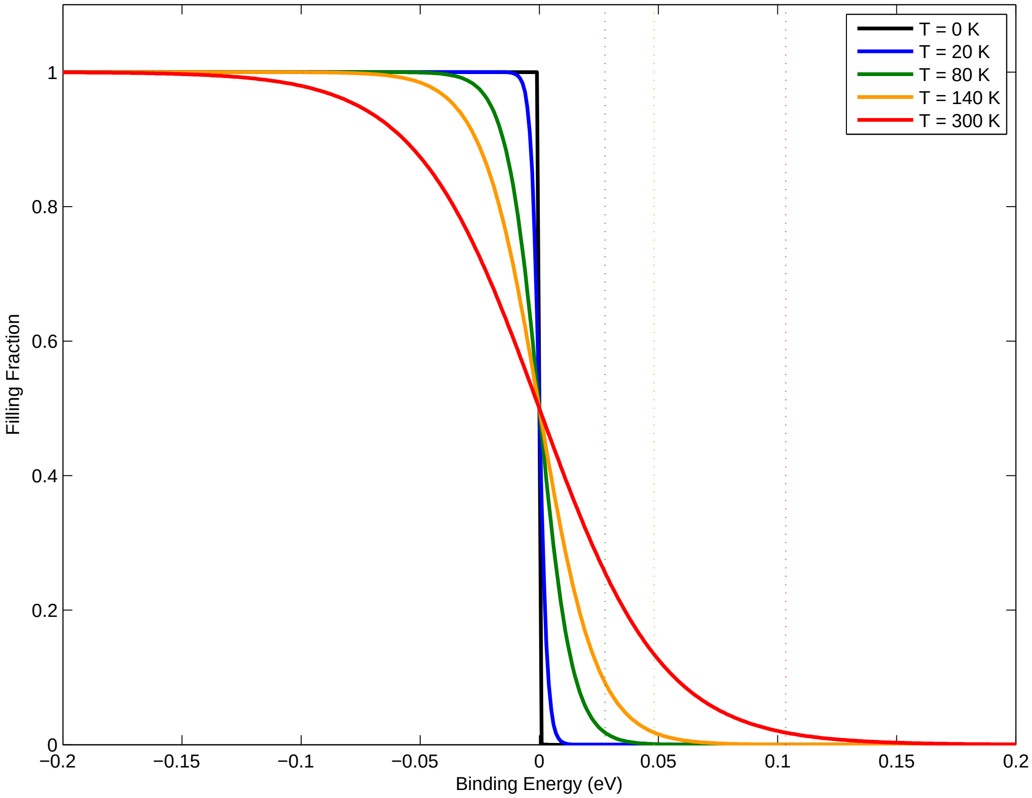


Figure 6.1: Fermi-Dirac distribution for several temperatures. For ARPES spectra with good counting statistics, dividing out the Fermi function reveals excitations up to $4k_BT$ above the Fermi level. The dashed lines indicate $4k_BT$ for the curves at 80 K, 140 K and 300 K; 27.6 meV, 48.3 meV and 103.4 meV.

it behooves us to try to use ARPES to probe these states.

6.2 The Resolution Problem

It has long been recognized in the photoemission community that one can in principle remove the effects of the Fermi function both above and below E_F from an ARPES spectrum, and thus recover $A(\boldsymbol{k}_\parallel, \omega)$, which is the theoretically relevant quantity, by simply normalizing the measured spectrum by a Fermi function[35]. Thus,

$$A(\boldsymbol{k}_\parallel, \omega; T) = \tilde{A}(\boldsymbol{k}_\parallel, \omega)/f(\omega; T). \tag{6.4}$$

The inversion of Eq. 6.4 to generate a Fermi normalized spectrum (FNS) is in principle a matter as trivial as the algebra done to get it. $\tilde{A}$ is simply what we measure and, if we know the sample temperature, we know $f(\omega; T)$ *a priori*. It turns out however that performing this normalization on real data in a rigorous way that can yield trustworthy, quantitatively accurate results is far from simple. The reason for this difficulty, which we will now explore at some length before returning to the problem of generating good FNS, is intimately related to the more general problem in ARPES which constitutes the second unresolved issue alluded to at the beginning of this chapter. An even simpler method of exploring unoccupied states, that of symmetrizing the spectrum in energy about E_F[36], has recently been shown to also suffer greatly from this resolution effect[37] on top of whatever physical deficiencies this method may have.

The issue at hand is that of all spectroscopies, namely, the instruments we use to perform measurements have a finite resolution in every variable. In the case of ARPES, photoelectron spectra are broadened by a finite energy resolution ΔE and a finite angular resolution $\Delta\theta$. The former is a function of the quality of the spectrometer being used, e.g. in terms of the stability of its power supplies or smoothness of its internal potentials and the width of the entrance slit described in Chapter 3, as well as the bandwidth of the exciting light. Clearly, for a dispersing band these effects will mix. The angular resolution is determined similarly by the spectrometer quality as well as by the spot size on the sample made by the impinging light. Assuming all these contributions to the various resolutions are Gaussian in nature, the resolution for a given experiment is usually parameterized in terms of overall Gaussian energy and angular resolutions. We will denote the function describing this overall broadening in energy and angle by the asymmetrical, normalized Gaussian distribution $R(\theta, \omega|\theta', \omega')$,

$$R(\theta, \omega|\theta', \omega') = \frac{e^{\frac{-\omega^2}{2\Delta\omega^2}} e^{\frac{-\theta^2}{2\Delta\theta^2}}}{2\pi\Delta\theta\Delta\omega} \tag{6.5}$$

where $\Delta\theta$ and $\Delta\omega$ are the standard deviations of the Gaussian resolutions in angle and energy, respectively. Of the angular resolution versus the momentum resolution, the angular resolution is the more natural of the two to use because that is the variable our spectrometers actually measure and also because it avoids possible complications viz a viz the odd nature of scaling from emission angle to crystal momentum for very low energy photoemission. The presence of the primed variables in R is a natural consequence of our defining it as a convolution kernel in what is about to follow.

We note here that there exists an additional, non-Gaussian contribution to the resolution function that can arise if we are not careful. If the count rate of an experiment is low the pulse counting (PC) method should be used to record data rather than the usual analogue to digital converter (ADC). Recall that the detector at the back end of the spectrometer (Chapter 3) is composed of a microchannel plate (MCP) stack that multiplies photoelectrons, the resulting amplified pulses of which strike a phosphor screen. Flashes on the phosphor screen are recorded by a CCD camera, the pixels of which have been calibrated in terms of energy and angle. In the traditional ADC mode of data acquisition the intensities registered by the camera from the phosphor screen are converted directly to grayscale intensities in the image. Thus, if a pulse is large enough, as it often is, a single hit is recorded across many pixels of energy and angle with some spread in intensity values. The result is a contribution to the 2D resolution resulting not from entrance slits and photon bandwidths (for example) but entirely upon the energy/pixel and angle/pixel calibration. The PC mode on the other hand uses software to register each pulse as a single count at the center pixel of the flash recorded by the CCD camera. Thus at the price of having a lower absolute value of intensity in the resulting image we remove this extra resolution effect. This situation is diagrammed in Fig. 6.2. We mention this here because the effect of using the ADC is to introduce possible non-gaussian resolution effects which we have not taken into account in our deconvolution kernel, Eq. 6.5. In the event the count rate is so high individual pulses cannot be discerned by the software the ADC mode must be used and the data must be oversampled enough relative to the resolution that this is not a problem. Such is the case for laser ARPES due to its intrinsically high count rate.

The general problem of spectroscopic resolution manifests itself in ARPES in the following way. In ARPES, as in most spectroscopies, one can achieve almost any resolution desired at the expense of signal intensity and signal to noise ratio. While in principle our electron spectrometers are well enough constructed to achieve at least .1 meV resolution in energy, this can be done only by narrowing the entrance slit of the analyzer, reducing the counts reaching

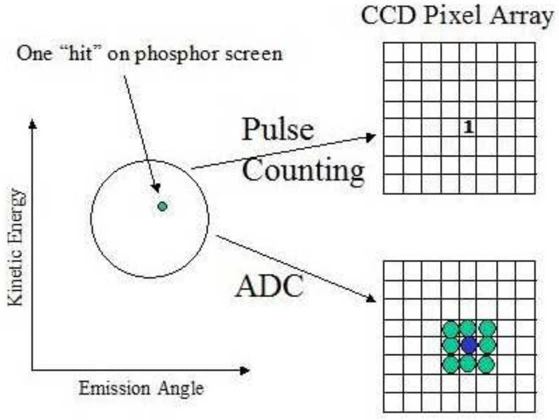


Figure 6.2: Schematic of ARPES data acquisition modes. In ADC mode hits on the detector phosphor screen are converted to grayscale intensities by the CCD camera spread across several pixels. In pulse counting mode each pulse is registered as a single count at the energy and angle position of the pulse center.

our detector in proportion. Similarly, we can narrow the bandwidth of incident light by passing it through a monochrometer in which the bandwidth, as well as the intensity, of light can be decreased by narrowing entrance and exit slits. Such is the case on all synchrotron beamlines as well as many lamps. (Of course in this respect lasers have the great advantage of being intrinsically bright while simultaneously having a very narrow width in energy.) All light sources face problems when the spot size on the sample, which defines the angular resolution, is considered. As one decreases the spot size one must simultaneously reduce incident photon flux to avoid space charging in the vacuum near the sample surface which would in turn reduce energy resolution. When all these contributions are taken together the experimentalist is forced to make a best compromise for a given experiment between a resolution sufficient to observe the details of a system while maintaining enough signal for the experiment to be performed at an acceptable signal to noise ratio and on a time scale comparable to the lifetime of a freshly cleaved sample surface. To put some numbers to the problem, a typical high resolution experiment at a synchrotron beamline usually produces results with energy resolution between 10 and 30 meV and an angular resolution between .1° and .2°.

If we properly account for the finite resolution of our experiment, as we did not do in Chapter 3, we find the spectral intensity measured in an ARPES experiment to be

$$I(\theta',\omega') \propto \int\int \tilde{A}(\theta,\omega)R(\theta,\omega|\theta',\omega')d\omega d\theta \tag{6.6}$$

where we have made the further simplification of absorbing the matrix elements $M_{f,i}^2$ into the proportionality because it represents only an overall intensity modulation in the spectrum. Because intensity information is recorded by an MCP detector, it is useful to view $I(\theta',\omega')$ as an image formed by an array of pixels such that individual counts are recorded on pixels given discrete values $\omega \pm \delta\omega$ and $\theta \pm \delta\theta$ where $\delta\theta$ and $\delta\omega$ are half the height (width) of a given pixel in the space of the measured variables. The effect of the convolution of the spectral function with the resolution kernel is to "smear" the image of $\tilde{A}$ in such a way that intensities are distributed about their mean value to neighboring pixels in a Gaussian fashion. This process does not alter the overall intensity of the image but rather blurs it. This is the general resolution problem of ARPES. If R is larger than some feature in $\tilde{A}$, e.g. a Lorentzian peak in the dispersion, that feature will be broadened accordingly and we say the spectrum is resolution limited. Considering the problem of generating

FNS, if $R(\theta) \gtrsim 4k_BT$ we clearly have

$$A(\theta,\omega) \neq I(\theta',\omega')/f(\omega;T) \tag{6.7}$$

or, in other words, $I(\theta',\omega') \neq \tilde{A}(\theta,\omega)$. In general, the resolution problem limits our ability to determine the intrinsic line shapes of the spectral functions as we would like to measure them from EDC's and MDC's. This problem is compounded at low T where, in the absence of thermal broadening, low energy electronic states can be quite narrow in many systems. The problem is especially burdensome in the vicinity of the Fermi level where states are theoretically the narrowest and where the presence of the Fermi distribution itself can, under the best of circumstances, distort the line shape in such a way as to render measurements of $\Gamma(\omega)$, k_F, v_F, etc. unreliable on the quantitative level. This last point is especially problematic for generating reliable FNS because a Fermi function broadened by convolution with a finite energy resolution produces another Fermi function of apparently higher temperature, T', such that

$$f(\omega;T') = \int f(\omega';T)R(\omega|\omega')d\omega' \tag{6.8}$$

where T' and T can be approximately related by the equation

$$4k_BT' = \sqrt{(4k_BT)^2 + \sigma_\omega^2}. \tag{6.9}$$

This expression is useful for determining the energy resolution of a measurement from a Fermi function acquired at a known temperature T.

In light of this relationship between apparent and physical Fermi functions one is faced with a serious quandary when attempting to produce reliable FNS from real spectra. Normalization of a broadened spectrum by the physical Fermi function produces undesirable intensity artifacts both above and below E_F thereby rendering the process highly unreliable. On the other hand, normalization of a broadened spectrum by a resolution broadened Fermi function, while appearing to mitigate this problem, results in equally unreliable results in which, for example, the intensities and binding energies of states above and below E_F can be seriously over or underestimated and the lineshapes and dispersions themselves distorted. This is the result of attempting to normalize a manifestly unphysical quantity, I, by another unphysical quantity, $f(\omega; T')$. The unphysical components of these quantities do not cancel. Clearly if we want to obtain more reliable quantitative results from ARPES FNS another method is required. The remainder of this chapter is devoted to the methodology we've devised for remedying the situation.

6.3 Lucy Richardson Deconvolution

There are several methods available for dealing with these problems, none of which constitute a silver bullet. The obvious solution is simply to find ways to improve the experimental energy and angular resolutions that can be routinely achieved. While great strides in this direction are certainly under way, and much has been done already, the constraints mentioned above as well as unforseen problems specific to different kinds of crystal samples will probably always render a certain level of high resolution unattainable except for a few select circumstances. For example, certainly as one scales up in photon energy into the VUV regime the problem of finding a light source bright enough to reduce the photon bandwidth significantly while maintaining reasonable levels of signal becomes a problem. It's also not unfair to say that at some point $\Delta\omega \propto ¥$, implying a law of diminishing returns for future experimental facilities dedicated to ARPES.

If one is unable or unwilling to improve the experimental resolution outright remedy must be sought in analytical and computational methods for treating the data. One method suggested in the literature is to generate a simulation of the underlying spectral function ($\tilde{A}$) based on theoretical calculations, broaden it with Eq. 6.6, and then compare the results to the experiment[38]. We call this the traditional approach. While there is certainly nothing wrong with this method in principle, it offers very little for experiments on systems for which there is no accepted theoretical description or for which there are several nearly degenerate theoretical models. The cuprates, and strongly correlated electron systems in general, are good examples of this problem as it is well known that band structure calculations fail spectacularly in many of these materials. What's more, if one is interested in discovering new physics, i.e. physics not yet contained in a model, the traditional approach offers little but to point out one's lack of knowledge in the best case and in the worst case to simply leave open the question of whether one should put more stock in theory or experiment, both or either of which may ultimately be flawed.

Another method for overcoming the ill effects of finite resolution in ARPES is deconvolution. Deconvolution is essentially a method for inverting integrals of the form of Eq. 6.6. The question we would like to answer is, given the result of the convolution (in this case I) and the convolution kernel itself (R), both of which we know *a priori* by performing an experiment, what is $\tilde{A}$? This very reasonable question it turns out is exceptionally ill defined mathematically. The most successful methods for accomplishing this task essentially take the form of clever curve fitting methods. That is, for reasons having to do with the nature of the statistics of conditional probabilities, called Bayesian statistics, one can only ascertain the function *that is most likely* to give the result I when convolved with a given kernel R. One method that has been applied with mixed success to ARPES is the so-called Maximum Entropy Method (MEM)[39][40][41][42]. In the cases presented in the literature the MEM was used to invert the equation for the electron-phonon coupling constant λ given in terms of the Eliashberg $\alpha^2F(\omega)$

$$\lambda = 2\int_0^\infty \frac{d\omega}{\omega}\alpha^2F(\omega) \tag{6.10}$$

where λ is directly related to the real part of the self energy $\mathrm{Re}\,\Sigma$ discussed in Chapter 3. The object was to retrieve $\alpha^2F(\omega)$ and thus the phonon spectrum from the ARPES spectrum. The MEM is perhaps well suited to such a task, at least in principle, because it allows one to explicitly introduce into the fitting procedure the constraint that the result should look like the Eliashberg funtion and one can effectively control how strongly the constraints are applied to the data. The problem with this approach to the more general problem of extracting line shapes from ARPES is that it explicitly introduces some idea of what the data *should* look like, *a priori* and enforces that to some degree decided by the analyst. In the sense that one can only get out physics like what one expects to find, the MEM can only be applied in situations for which one already has some degree of confidence that the traditional method outlined above can be useful. Thus, bias about the physical nature of the expected result is explicitly introduced into the analysis. Without a theory in hand, one could not easily apply the MEM to good data representing the edges of our understanding of physical systems. To paraphrase a recent Secretary of Defence, the MEM can help you resolve known unknowns but can do nothing in the presence unknown unknowns.

An alternative method of deconvolution, the application of which to ARPES has been recently pioneered by our group[43], is called the Lucy-Richardson method (LRM)[44][45]. This method, which has been applied to the analysis of medical and astronomical imagery and recently to other forms of electron spectroscopy [46] applies equally well to ARPES data. As explained earlier, ARPES data taken on modern detectors essentially forms an image on an array of pixels of angle and energy space. As will be demonstrated below, the LRM has the great advantage for ARPES that it does not require any assumptions to be made about the nature of $\tilde{A}$ or the physics that it encodes. All it requires from the analyst is the data itself and a knowledge of the resolution function R, the parameters of which depend only upon the apparatus used and which is easily determined during an experiment by simple calibration procedures. To restate the problem, given *a priori* knowledge of I and R from the experiment itself, what is the most likely physical $\tilde{A}$ to have underlain the convolution Eq. 6.6? This manner of asking the question is entirely agnostic as regards the underlying physics of $\tilde{A}$.

In the LRM, the kernel R is viewed as the probability that an electron originating from a binding energy $\omega \pm \delta\omega$ and angle $\theta \pm \delta\theta$ in the true spectral function of the sample is measured to have originated from the binding energy $\omega' \pm \delta\omega'$ and angle $\theta' \pm \delta\theta'$ where the δ's in the unprimed system originate from quantum mechanics and those of the primed system originate from the Gaussian nature of the resolution function. The former contribution is intrinsic to the physics of photoemission. The latter is extrinsic to the photoemission process and represents the instrumental contribution to the observed photoemission spectrum. This is a restatement of the "smearing" effect. It was shown by Lucy that, using Bayes' theorem one can define a quantity Q such that

$$Q(\omega, \theta|\omega', \theta') = \frac{\tilde{A}(\omega, \theta)R(\omega', \theta'|\omega, \theta)}{\int\int \tilde{A}(\omega, \theta)R(\omega', \theta'|\omega, \theta)d\omega d\theta} = \frac{\tilde{A}(\omega, \theta)R(\omega', \theta'|\omega, \theta)}{I(\omega', \theta')}. \quad (6.11)$$

Rearranging terms and integrating both sides in the primed (measured) coordinate system it is found that

$$\tilde{A}(\omega, \theta) \equiv \int\int I(\omega', \theta')Q(\omega, \theta|\omega', \theta')d\omega' d\theta' \quad (6.12)$$

where we have used that fact that $\int\int R(\omega', \theta'|\omega, \theta)d\omega' d\theta' = 1$ for a properly normalized probability distribution. From Eq. 6.12, we see that Q appears to be the inverse kernel of R and in principle reduces the problem of inverting an integral to the much simpler problem of solving one. Because Q is a functional of R, this equation cannot be solved by directly positing a Q. A partial solution can be acquired if one is able to devise an analytical approximation to Q by positing a functional form for R as well as the underlying line shapes of $\tilde{A}$ and has met some success. That practice has some disadvantages of its own stemming from the assumption of an underlying line shape and the need to approximate Q as a power series. The LRM on the other hand aims to generate successively better estimates of Q and therefore of $\tilde{A}$ by an iterative procedure. The initial guess of Q, Q^0 is obtained numerically from Eq. 6.11 using $I = I_0$ where I_0 is the observed spectrum. This guess is used to generate an initial estimate of $\tilde{A}$ using Eq. 6.12 such that,

$$\tilde{A}^1(\omega, \theta) = \int\int I_0(\theta', \phi') Q^0(\omega, \theta|\omega', \theta') d\omega' d\theta'. \tag{6.13}$$

With an estimate of $\tilde{A}$ in hand the cycle can be repeated. In general, where r denotes the iteration number,

$$\tilde{A}^{r+1}(\omega, \theta) = \int\int I_0(\theta', \phi') Q^r(\omega, \theta|\omega', \theta') d\omega' d\theta' \tag{6.14}$$

where (after the initial cycle)

$$Q^r(\omega, \theta|\omega', \theta') = \frac{\tilde{A}^r(\omega, \theta) R(\omega', \theta'|\omega, \theta)}{I^r(\omega', \theta')} \tag{6.15}$$

where

$$I^r(\theta', \phi') = \int\int \tilde{A}^r(\omega, \theta) R(\omega', \theta'|\omega, \theta) d\omega d\theta. \tag{6.16}$$

At this point the procedure can be simplified by using Eq. 6.15 in Eq. 6.14 so that

$$\tilde{A}^{r+1}(\omega, \theta) = \tilde{A}^r(\omega, \theta) \int\int \frac{I_0(\theta', \phi')}{I^r(\theta', \phi')} R(\omega', \theta'|\omega, \theta) d\omega' d\theta'. \tag{6.17}$$

So far we have only couched the results obtained by Lucy in terms familiar to the ARPES formalism. Lucy went on to show that this procedure has several features that are highly desirable for the manipulation of spectroscopic data. First, the LRM conserves the overall intensity of the spectrum;

$$\frac{\int\int I_0(\theta',\phi')d\omega' d\theta'}{\int\int I^r(\theta',\phi')d\omega' d\theta'} = 1. \tag{6.18}$$

That is, the LRM only "rearranges" intensities across pixels in a process that is the reversal of the blurring effect the resolution had (which also conserved spectral weight). Thus, overall intensity information is rigorously preserved. The next highly desirable feature of the LRM is that no negative values in $\tilde{A}$ are produced by the procedure so long as the initial guess to $\tilde{A}$ has no negative values, as should be the case for a spectral function. Thus

$$\tilde{A}^{r+1}(\omega,\theta) \geq 0 \forall \{\omega,\theta\}. \tag{6.19}$$

While a seemingly obvious spectral requirement for any integral inversion procedure we might use, it turns out this last point is highly non-trivial as the vast majority of procedures involving, for example, a direct assault on Eq. 6.6 or approximations to Eq. 6.12 can result in the production of negative intensities. Finally, the LRM is inherently insensitive to fluctuations in the original spectrum on scales much shorter than the R because it only attempts to get the I^r generated from estimates of $\tilde{A}^r$ to match I_0 over an integral. This smoothing out results from the integrals over R. In other words it will not see the statistical noise inherent in any experiment for many iterations. (Similarly, in a real measurement the instrumental resolution broadening tends to destroy whatever quantum statistical noise might appear in $\tilde{A}$ itself so that the high frequency noise actually observed is entirely due to the apparatus.) In principle, so long as R is not too much larger than any characteristic feature in the data, the procedure should converge to a good "fit" to $\tilde{A}$. Conversely, features much smaller than the resolution will be treated poorly, i.e. as noise, and so the requirement of performing a good experiment in the first place is not negated. However we can clearly compensate for the not unreasonable amount of broadening that accompanies the vast majority of ARPES experiments.

Because attempting to iterate to full convergence will ultimately enhance otherwise benign noise in the data it is useful to check explicitly for each iteration that one has not iterated too far. Even in the absence of noise related artifacts the use of Fourier transforms in the procedure will, if some care is not taken, produce other undesirable artifacts that must be guarded against. We do this by calculating a χ^2 per degree of freedom for the "fit" of $\tilde{A}$ by using I^r after each iteration.

$$\chi^2(r) = \frac{1}{N}\sum_N \frac{(\int\int \tilde{A}^r(\omega,\theta)R(\omega',\theta'|\omega,\theta)d\omega d\theta - I_0(\theta',\phi'))^2}{I_0(\theta',\phi')} = \frac{1}{N}\sum_N \frac{(I^r - I_0)^2}{I_0} \tag{6.20}$$

In practice, $\chi^2(r)$ shows either an asymptotic convergence to some value close to one, in which case it is best to cut off the procedure by hand before it begins to adjust noise-like features to fit the data or it has a clearly defined minimum, in which case one stops iterating there. Either way, as we show below, only a few iterations are needed in almost all circumstances of interest.

It turns out that the LRM is particularly well suited to the analysis or ARPES data because acquiring a detailed knowledge of R, the most important information we have to provide the algorithm besides the data itself, has always been a standard practice in ARPES. What's more, in ARPES, R takes on the particularly simple form of Eq. 6.5 so that we only have to supply the two parameters from experiment $\Delta\omega$ and $\Delta\theta$. This is a stark contrast to the much more complicated divination of the broadening kernel that was required, for example, to analyze early imagery from the Hubble Space Telescope when it was found to be suffering a severe astigmatism. This simplicity is a result of the fact that our spectrometers image phase space rather than real space, the former being governed by the simple and robust kinematics outlined in Chapter 3.

It is clear we now have a program for reliably generating FNS from resolution limited data. That is, we want to use the LRM to arrive at the $\tilde{A}$ that was most likely responsible for producing the I_0 observed in experiment. With $\tilde{A}$ in hand, we can normalize by an unbroadened Fermi function at the actual experimental temperature, T_{exp}, and retrieve A without the kinds of anomalies illustrated above. Thus,

$$A(\omega,\theta) = \tilde{A}_{LRM}(\omega,\theta)/f(\omega;T_{exp}) \tag{6.21}$$

which is the physically correct statement.

6.4 Deconvolution of Simulated ARPES Spectra

To make the above discussion concrete and to demonstrate the efficacy of the computational techniques explored above we would like to progress in as orderly a manner as possible. Thus we shall first demonstrate the LRM on simulated ARPES data encoding the simplest possible situation we are likely to encounter in a real experiment: a single Fermi liquid-like band crossing the Fermi level. We will then compare those results to the deconvolution of real data acquired on a similarly simple state. Then, with an eye towards exploring more interesting physics, we shall apply the LRM to a linear band with a BCS type superconducting gap opened at the Fermi level and compare that to data obtained from the superconducting state of Bi2212. At each stage we shall compare the results of Fermi normalization by the various methods outlined so far as well as, where applicable, observe how basic observables such as lifetime, binding energy and dispersion are affected by the LRM.

For simplicity we begin by modeling a linear band crossing the Fermi level, which is a good approximation to the low energy nodal excitations of Bi2212. The dispersion relation is therefore $\epsilon_{k_{\|}} = v_F(k_{\|} - k_F)$ where we've set $k_F = .4\mathring{A}^{-1}$ and $v_F = 1.2eV\mathring{A}$. The spectral function itself is modeled using Eq. 3.31 where we set $\xi = 0$ in the absence of a well defined bosonic interaction, $Z_k = 1$ and use $\Gamma(\omega) = \alpha + \beta\omega^2 + (\gamma k_B T)^2$ to get an energy and temperature dependent broadening similar to that predicted for a Fermi liquid. We've also added a small, constant background to model the presence of A_{inc}. In the simulations we've set $\alpha = 2meV$ to model a small temperature and energy

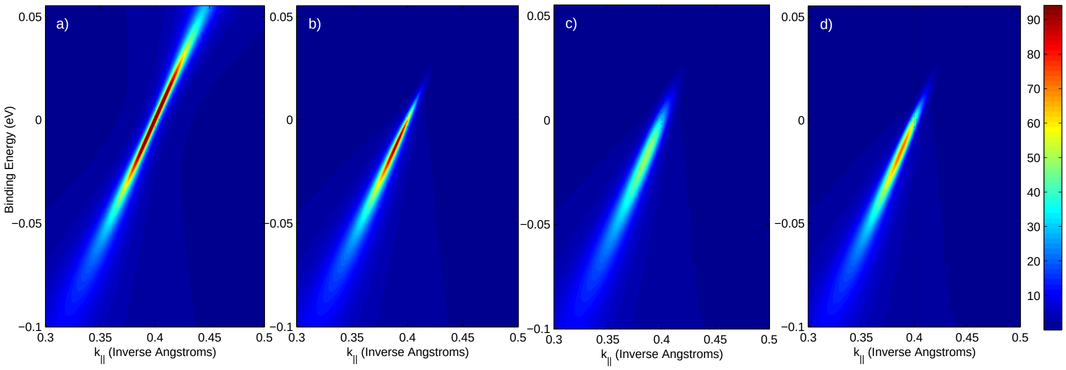


Figure 6.3: Simulations of a linear band crossing E_F. a) Theoretical spectral function $A^0(k_\parallel,\omega)$ b) $\tilde{A}^0(k_\parallel,\omega) = A^0(k_\parallel,\omega)/f(\omega;T_{exp})$ c) $I_0(k',\omega') = \tilde{A}^0(k_\parallel,\omega) \bigotimes R(k,\omega|k',\omega')$, d) $\tilde{A}^7(k,\omega)$

independent impurity scattering, $\beta = 3$, $\gamma = \pi$ and $T = 80K$. The energy and momentum broadening kernel R is parameterized by Eq. 6.5 with the energy and momentum resolutions set to $\Delta\omega = 10$ meV and $\Delta k = .003\mathring{A}^{-1}$ FWHM, respectively.

Image convolutions were performed using IMFILTER algorithm and the LRM was applied using the DECONVLUCY algorithm, both of which are contained in the Matlab Image Processing Toolkit. For simplicity we have not added detector noise to these simulations though its effect becomes obvious in real data. In Fig. 6.3, panel a) we plot the complete spectral function $A(k,\omega)$ and in panel b) the occupied part of the spectral function $\tilde{A}(k,\omega)$ where we have cut off the spectrum by multiplying $A(k,\omega)$ by $f(\omega;T_{exp})$. The result of convolving the image in panel b) by the resolution function to generate $I_0(k,\omega)$ is shown in panel c). Finally, in panel d) the result of deconvolving I_0 by the same resolution function is shown after seven iterations. This result is denoted by $\tilde{A}^7(k,\omega)$. All panels in the figure share the same color scale mapping of intensities denoted by the color bar. While the states we've chosen to simulate are rather sharply peaked compared to what one often encounters in real life the effect of even 10 meV energy broadening is profound. The rapid drop in peak intensity in panel c) is a typical effect as is the deleterious nature of intensity appearing at more positive binding energies than it ought to. Nevertheless, panel d) shows the LRM vastly improves upon the "as measured" spectrum.

A better idea of what's happening in the simulations can be obtained from examination of EDC's taken at k_F (Fig. 6.4a) and MDC's taken at E_F (Fig. 6.4b). While with this number of iterations we were unable to completely reconstruct the theoretical $\tilde{A}$, the results in both EDC's and MDC's show clear improvement over I_0 in terms of peak height and width. The shift and subsequent recovery of k_F, located at the MDC peak, is dramatic enough to be visible by inspection. The most promising feature to show up in the EDC's is the drastic reduction of spurious spectral weight above E_F relative to I_0. Also, as should be the case, the broadening and deconvolution procedures have no detectable effect on features broader than the resolution in either momentum or energy.

While various edge smoothing and image padding techniques have been employed to mitigate the edge effects associated with using Fourier transforms, there remains a small spurious feature in the EDC's at about -20 meV. It is also visible, though greatly reduced, in the MDC's. As we'll see below, this effect, common in the simulations, does not show up in the deconvolution of real data. This problem is related to our method of simulating I_0, that is, the act of convolving the original spectral function with a Gaussian produces medium

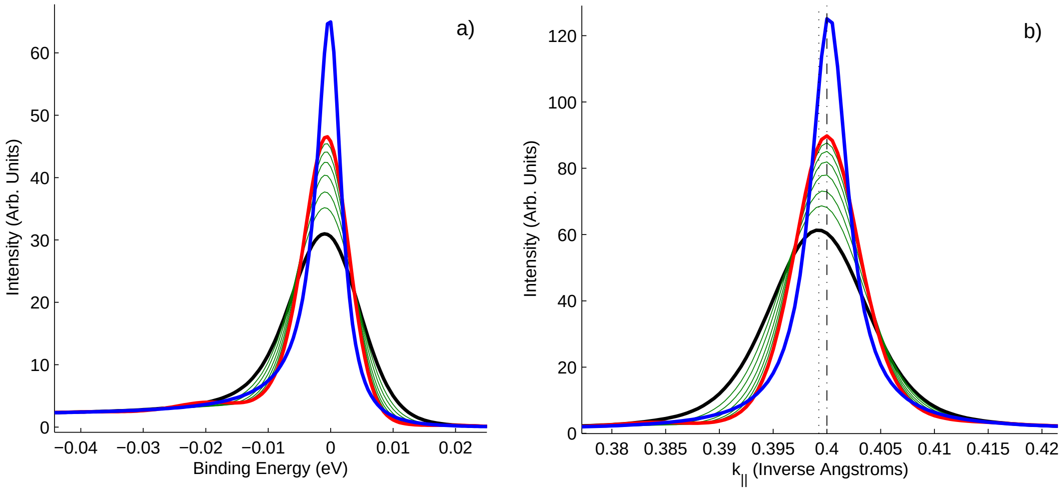


Figure 6.4: a) EDC's at k_F and b) MDC's taken at E_F. The thick blue lines are from the theoretical $\tilde{A}$, the thick black lines are from I_0, the thick red lines are from $\tilde{A}^7$ and the thin green lines are intermediate iterations.

wavelength ripples in the final image which, while hardly noticeable at first are amplified in deconvolution. The LRM assigns some non-zero statistical weight to the fluctuations introduced by the original broadening and thus appears to interpret those ripples as real features. In principle this problem could be remedied by, for example, using an analytical approximation to a Voigt function to simulate the broadening, thereby avoiding the necessity of the first round of Fourier transforms. For the time being, so long as we can properly interpret the results, this is beyond the scope of our present case.

Because ultimately we wish to determine the best method for generating FNS, the final step in our simulation is to compare what happens when we try to recover the full spectral function to the original theory. In Fig. 6.5a) we

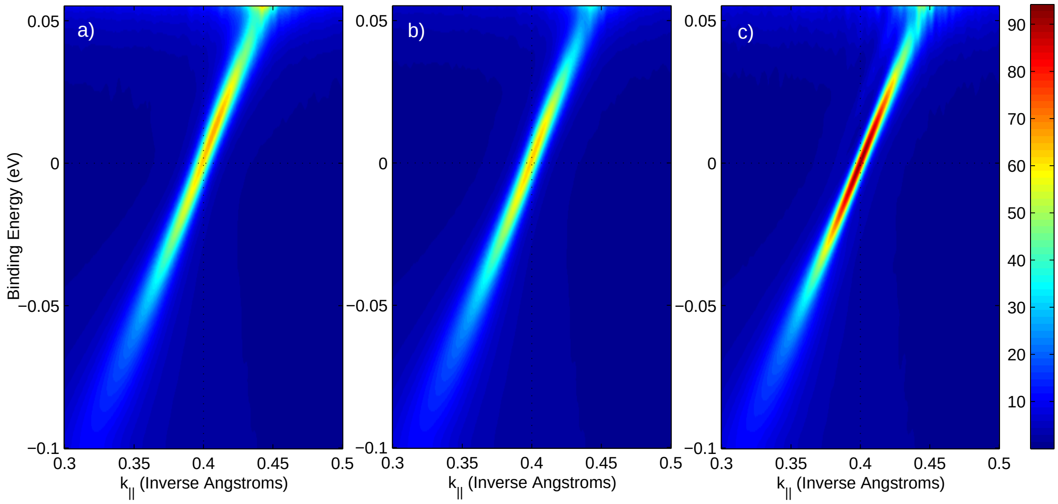


Figure 6.5: a) $I_0/f(\omega;T_{exp})$, b) $I_0/f(\omega;T')$, c) $\tilde{A}^7/f(\omega;T_{exp})$

have simply normalized I_0 by the Fermi function at the experimental temperature. The result is clearly asymmetric in binding energy, with the intensity peak falling well above E_F and the peak widths clearly unequal at positive and negative frequencies. In panel b) we simulate the method most often reported in the literature, that of normalizing I_0 by a resolution broadened Fermi function. While better than the result in panel a), the peak in the intensity still lies above E_F and the peak widths are still somewhat broader above than below E_F. Lastly, in panel c) we present the result of normalizing $\tilde{A}^7$ by the experimental temperature Fermi function. Panel c) compares rather well to the original theory, Fig. 6.3a) both in terms of the peak widths across the whole spectrum as well as well as the symmetry of the band above and below E_F. We also present Lorentzian fits to the MDC's of the spectra in Fig. 6.5 as well as to the theory, Fig. 6.3a). The results are summarized in Fig. 6.6 for the MDC peak widths Δk, spectral weights Z_ω and dispersion

$\epsilon_{k_\parallel}$. While the lineshapes we are fitting are clearly Lorentzian convolved with a Gaussian, known as a Voigt lineshape, that more honest type of fit adds a level of detail we are not currently aiming to grapple with. In general, fitting a Lorentzian to a Voigt profile will slightly underestimate the peak width and slightly overestimate the spectral weight. Because the Gaussian broadening is not very severe however the errors are only on the order of a few percent and in the present circumstance are small enough to be ignored.

In panel a) of the figure we find that the peak widths obtained by normalizing a broadened spectrum by a Fermi function at either T or T' give indistinguishable results. Normalization of the deconvolved spectrum by a Fermi function set at T on the other hand does an admirable job of recovering the widths above and below E_F with a relatively small residual error in the immediate vicinity of the Fermi level. The quasiparticle weights, not surprisingly, are the most affected by resolution broadening and the choice of Fermi function used in the normalization. Normalizing a broadened spectrum by $f(\omega, T')$ is terrible. Surprisingly, normalizing I_0 by the resolution broadened Fermi function actually yields *worse* results below E_F and overestimates the weight above E_F in manner that suggests the result will become unreliable somewhat before $4k_BT$. The deconvolved result does the best below E_F and is somewhat better, and at least linearly diverging from the theory, above E_F, up to $4k_BT$. Again, we expect the deconvolved results to perform better than this on real data because of the Fourier transforms we had to invoke to broaden the spectrum in the first place. Finally, the dispersions extracted for both broadened spectra are nearly identically misplaced relative to the dispersions extracted from the theory and the fits to the deconvolved spectrum, which are

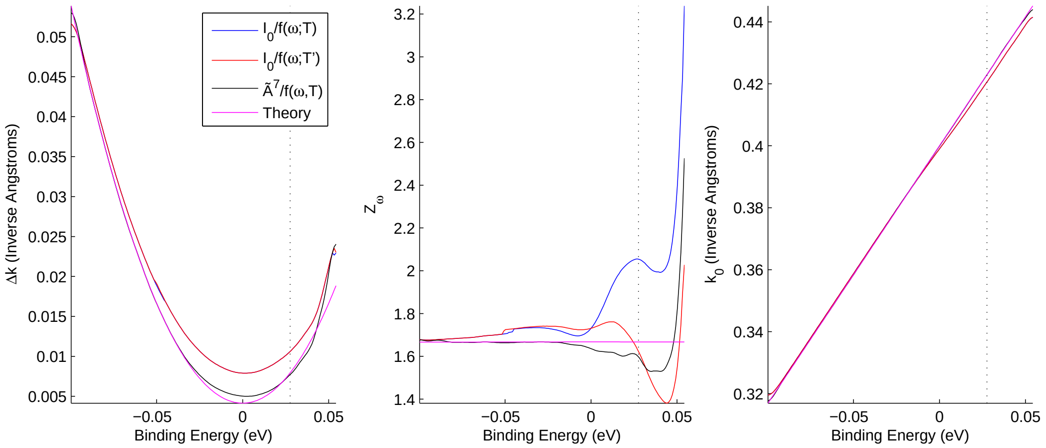


Figure 6.6: a) Lorentzian FWHM $\Delta k(\omega)$, b) spectral weight Z_ω and c) MDC peak positions $k_0(\omega)$. In all panels, the magenta line is the fit to theory, Fig 6.3a), the blue line is the fit to Fig. 6.5a), the red line is the fit to 6.5b) and the black line is the fit to 6.5c). The dashed line marks $4k_BT$, the energy up to which one can usually recover spectra in a real experiment.

nearly identical to each other. Again, the correct and incorrect dispersions begin to diverge below the Fermi level and continue to diverge as one goes to more positive binding energies. It's worth taking a moment to point out the value of a null result to the deconvolution procedure. That is, if one were to attempt to deconvolve a spectrum and saw no effect as a result the spectrum can be said with certainty to be a valid representation of $\tilde{A}$.

To model the opening of a superconducting gap, we use the weak coupling BCS spectral function [5]

$$A(k,\omega) = \frac{1}{\pi}\frac{u_k^2\Gamma_k(\omega)}{(\omega - E_k)^2 + \Gamma_k^2(\omega)} + \frac{1}{\pi}\frac{v_k^2\Gamma_k(\omega)}{(\omega + E_k)^2 + \Gamma_k^2(\omega)} \tag{6.22}$$

where

$$E_k = \sqrt{\epsilon_k^2 + \Delta_k^2} \tag{6.23}$$

with Δ_k the gap parameter and where the coherence factors u_k^2 and v_k^2 are given by

$$u_k^2 = \frac{1}{2}(1 + \frac{\epsilon_k}{E_k}) \qquad (6.24)$$
$$v_k^2 = \frac{1}{2}(1 - \frac{\epsilon_k}{E_k})$$

with the normalization condition

$$u_k^2 + v_k^2 = 1.$$

Notice that in the presence of a superconducting gap the spectral function becomes double valued in energy for the range of k around k_F that both u_k^2 and v_k^2 are non-zero. Excitations such as these are said to mix particle and hole states into a linear combination of fields that, taken together, form a single excitation with unit probability for observation per the normalization, Eq. 6.25. This compound excitation is called a Bogoliubov quasiparticle (BQP). One of the key facts embodied by this spectral function is that at $k = k_F$, $u_k^2 = v_k^2$. That is, the Bogoliubov quasiparticle peaks above and below E_F should have equal weight. Also notice the strict symmetry of the peak binding energies above and below E_F. This is a key prediction we shall want to check on the cuprates because other orders, such as density wave orderings, do not have to respect this binding energy symmetry.

In this simulation the dispersion, lifetime, T and R are the same here as for the simple linear band above. The gap parameter Δ_k has been set to 7 meV. (Recall the full BCS gap, peak to peak, is 2Δ.) Because at $T = 0$, in the absence of pair breaking scatterers, the gap is defined by symmetric delta function above and below E_F, we interpret our gap parameter as the peak to peak distance between two Lorentzians. Until quite recently one could raise strong and probably well justified objections to our attempting to model the opening of a superconducting gap at 80 K in a cuprate with the BCS spectral function. This is an issue quit beside whether or not pairing in high T$_c$ superconductors is due, e.g. to phonons. What's more, it has been known for some time that the EDC lineshape in the cuprates is probably not Lorentzian. Nevertheless, the literature as well as our own data appear to support the notion that, at least in the vicinity of the nodes, the superonducting gap Δ_k behaves in a manner indistinguishable from the BCS gap used here. In this respect, the d-wave nature of cuprate superconductivity only enters by our recognizing the gap amplitude Δ_k varies along the putative Fermi arc according to Eq. 2.2[47]. This BCS-like behavior even seems to extend to the coherences observed in experiment. As we'll see, it appears that the spectral function describing the opening of the superconducting gap is agnostic regarding the mechanism by which superconductivity is achieved.

For the simulations we shall proceed as before. Fig. 6.7a) shows the full spectral function. The opening of the gap causes the characteristic folding of the bands above and below the Fermi level. This is the effect of particle-hole mixing in the superconducting state. The intensity of the folded branches is gradually extinguished by the coherence factors, Eqs. 6.24. Panel b) shows the occupied part of the spectrum at 80 K. Even though the spectrum is cut off by the Fermi distribution the lowest lying states above the gap corresponding to BQP excitations are clearly visible. Notice also that the gap, as in panel

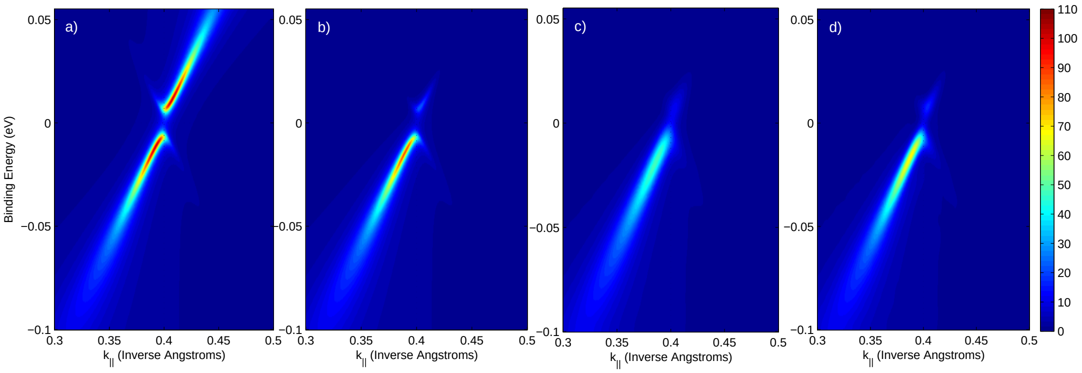


Figure 6.7: Simulations for a single band with a 7 meV BCS type superconducting gap. a) the whole spectral function, b) cut off by the Fermi distribution, c) blurred by 10 meV energy and .01 degree angular resolution and d) result of deconvolving c) with the LRM after seven iterations.

a), is not entirely empty of single particle excitations. This is due in our model to the presence of the impurity scattering term in the expression for Γ that is independent of energy and temperature, as well as the pronounced thermal broadening that occurs at this relatively high temperature. It has been shown for BCS superconductors in the strong coupling limit that such a broadening of the density of states into the gap can be attributed to excitation of single electrons by interaction with a bath of virtual phonons. Whether or not phonons are causally related to superconductivity in the cuprates, at the high temperatures at which the phenomenon is observed they will certainly make their presence felt. In any event, our simple impurity term acts as a pair breaker.

The presence of the quasiparticle peak above E_F has only recently been noticed as an obvious feature of the near nodal ARPES spectrum of the cuprates. The reason for this is illustrated by the resolution broadened model, Fig. 6.7c).

Here, as in most spectra acquired at a moderate energy resolution at "high" temperatures, the gap has been almost completely annihilated by energy resolution broadening. Further, the most delicate feature we would like to observe, the coherence suppressed band folding, is almost completely wiped out. A perusal of the literature will show that unless great care is taken to maximize resolution while maintaining adequate signal, while also using samples for which impurity scattering has been reduced to a negligible contribution, all information about the nature of excitations above the chemical potential, and indeed a great deal of information about the gap itself, is lost. The result is the appearance of a "tail" above E_F that has in the past been largely ignored. Nevertheless we are once again able to demonstrate the power of the LRM to remove the worst effects of resolution broadening. The result of seven iterations of deconvolution, $\tilde{A}^7(k,\omega)$, as applied to the image in panel c), is shown in panel d). While the overall sharpening is similar to that observed for our non-superconducting model above, when applied to the situation of a gap in the spectrum about the chemical potential the LRM shows itself to be invaluable in the recovery of the most basic features of a spectrum associated with Cooper pairing. The results of this deconvolution are summarized by the EDC's taken at k_F in Fig. 6.8 for the original, broadened and deconvolved theories as well as curves for the intermediate iterations of the deconvolution procedure. These EDC's, taken before Fermi normalization, reveal quite a bit about the efficacy of our deconvolution procedure. One of the most telling features to reveal itself in the theory is the position and intensity of the gap center. While as stated above the gap in the density of states is, for a BCS superconductor, centered at the chemical potential, once the Fermi cutoff has

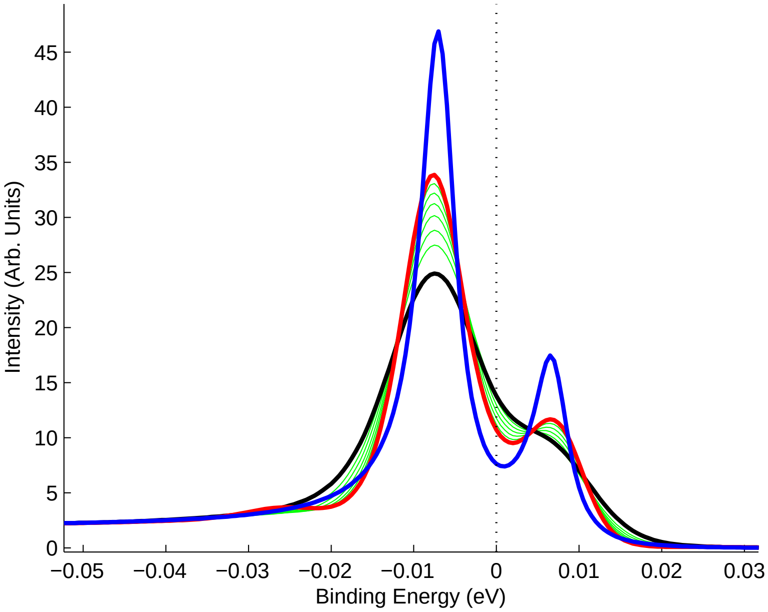


Figure 6.8: EDC's at k_F taken from the unbroadened thoery (blue curve), the broadened theory (black curve), the deconvolved spectrum (red curve) after seven iterations and intermediate results of the LRM (green curves).

been introduced this is clearly no longer the case for the *occupied* part of the spectral function. This shift in the gap minimum is a result of the redistribution of electron spectral weight from below to above the chemical potential and is a physical process. The same process causes the unnormalized binding energy of the positive binding energy peak to appear to shift thereby rendering the peak positions asymmetric in binding energy. This shift is accompanied by an induced asymmetry in the positive energy peak lineshape itself as the intensity is rapidly modulated by the Fermi function as it dies to zero. More striking is the ability of the LRM to rapidly transform the broadened theory, in which the positive energy peak appears merely as a shoulder in the spectrum, into a distribution in which the separate existence of the high energy peak is clearly delineated. Even a few iterations suffice to bring our broadened spectrum much more in line with the underlying physical picture.

The spectra generated by Fermi normalization of the broadened and deconvolved theories are shown in Fig. 6.9. Again, as for the normal state band, the improvement granted by the LRM is self evident. Note that only the deconvolved spectrum, panel c, clearly shows the presence of the folded bands and their rapid extinction due to the coherence factors. As demonstrated by the corresponding EDC's taken at k_F, Fig. 6.10, deconvolution followed by normalization to the physical Fermi function clearly gives the best result in terms of peak widths, symmetry of binding energies and position of the gap minimum. As promised, the gap minimum of the original theory falls at E_F and the peaks are symmetric in binding energy. Also, as expected, states above the chemical potential are affected much more adversely by resolution broadening than those below. It should be noted however that even the nega-

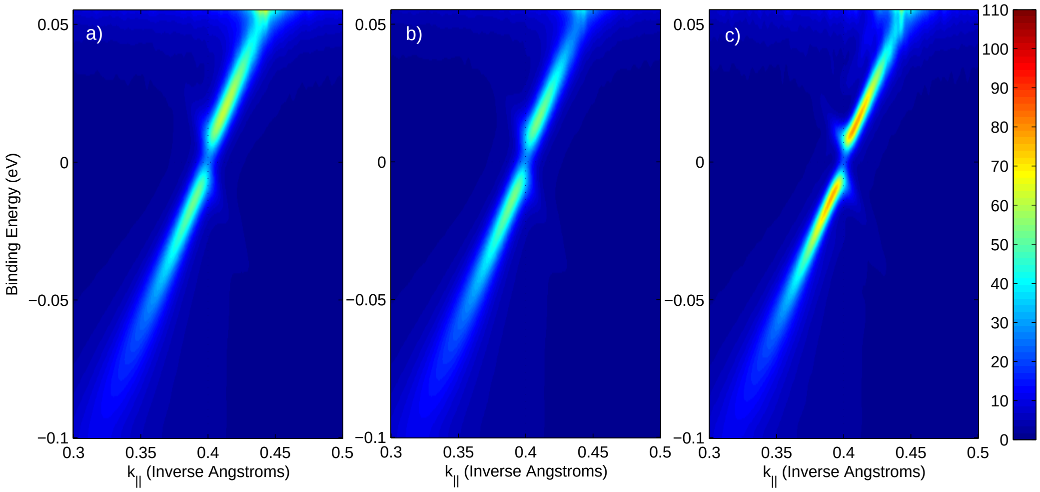


Figure 6.9: FNS resulting from a) dividing the broadened theory by the experimental Fermi function, b) dividing the broadened theory by the resolution broadened Fermi function and c) dividing the deconvolved theory, after seven iterations, by the experimental Fermi function.

tive binding energy peak is moved by broadening with the peak position, and hence the value of the gap one would measure on such data, clearly shifted.

6.5 Finer Points of Deconvolving ARPES Spectra

Having established that the best method with which to analyze resolution limited ARPES data in general, and FNS in particular, is to include Lucy-Richardson deconvolution as a standard component of the analytical toolbox it will behoove us to explore a little more deeply some of the finer points of utilizing these procedures. The most important outstanding issue for the practical application of these methods to real ARPES experiments arises from

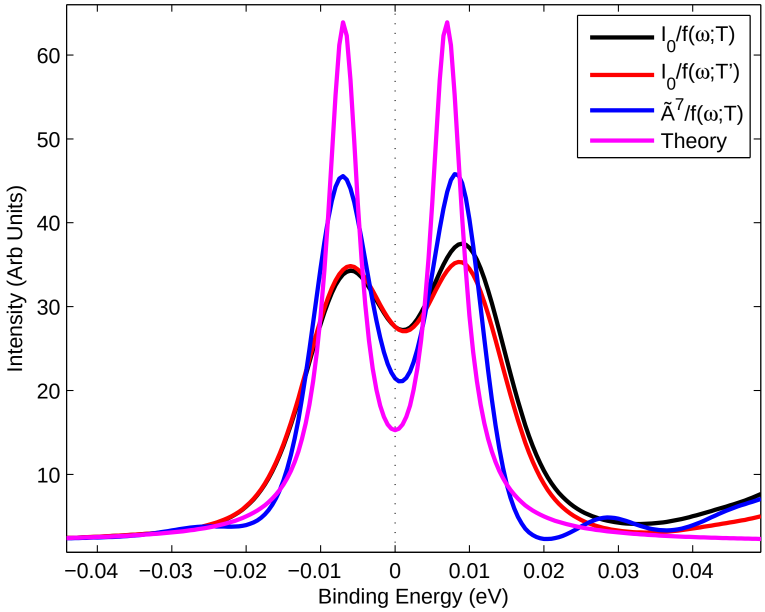


Figure 6.10: EDC's at k_F for the full theoretical spectral function (pink curve), the broadened theory divided by the experimental Fermi function and broadened Fermi function (black and red curves, respectively) and the deconvolved theory (after seven iterations) divided by the experimental Fermi function (blue curve).

two related questions. The first is, how many iterations should I use? Even in the original paper by L.B. Lucy it was admitted that one should iterate, “As often as is necessary or wise”. The second question is, if, as is evident by comparing $\tilde{A}^7$ to the original theory, one cannot fully deconvolve a spectrum due to noise effects or the introduction of unwelcome, spurious artifacts, what is the *effective* resolution I can achieve after some number of iterations? For, clearly, depending on how badly broadened the spectrum is relative to the intrinsic width of some feature, it will not always be possible to recover the true underlying spectral function in its entirety. Indeed, the presence of noise in a measurement should render that impossible even in principle because there will always be spectral weight missing.

To show how one might begin to tackle these and related questions it is instructive to carry out a far greater number of iterations than might normally be desired. In Fig. 6.11 the result of continuing to deconvolve our broadened superconducting simulation to twenty iterations is summarized. The further amplification of Fourier transform induced waves into “peaks” by the method clearly brings into sharp focus the kinds of artifacts that, in the absence of some *a priori* knowledge of the generic features of a spectrum we expect to observe, might be mistaken in a real experiment for truth. Again, this results after a few iterations from the algorithm’s attempt to “fit” relatively short wavelength features in the data, in this case induced by the broadening procedure itself. In real data, as we shall demonstrate this effect shows up in the amplification of noise in the data. On the other hand, while convergence is slow, we find the rather surprising and encouraging result that the lineshapes of the main peaks are reproduced well enough to justify our previous assertions that one can, in fact, recover something very near the true spectral function of a system from resolution broadened data. As some spectral weight has been pushed into the "artifacts", we do not expect the intensities to be as well reproduced as other factors. Evidently, so long as we are able to ignore whatever spurious features are generated by the procedure, and are able to recognize them as spurious, this method of analysis can be pushed rather far in obtaining certain quantitative information. In this sense, the LRM is very much like a traditional curve fitting procedure in which the scientist analyzing the data must inevitably make some decisions as to what model produces the best fit.

It is no sin to make some compromises and those we choose to make depends on whether we wish, per our above examples, to improve the overall picture or to render as perfectly as possible some smaller aspect of our analysis. Bearing in mind that like a conventional fitting of model to data the LRM is at heart a statistical method all we can expect to produce is a representation of the underlying physics that was *most likely* to have resulted in our data after broadening by our apparatus. This fundamental statistical uncertainty in the inversion of Eq. 6.6 is best illustrated by the images used to calculate the $\chi^2(r)$ of Eq. 6.20, Fig. 6.12. These images, the originally broadened spectrum on the left and the rebroadened, deconvolved spectrum (with $r = 20$) on the right are clearly nearly identical even though the image on the right resulted from broadening the heavily abused image in the inset of Fig. 6.11. Thus we say that while for a given image and R the application of the LRM is deterministic – meaning that it will yield the same result every time – it is not unique. Thus we find that the greater the broadening of a given feature relative to it's intrinsic width, the larger is the set of the number of possible inversions

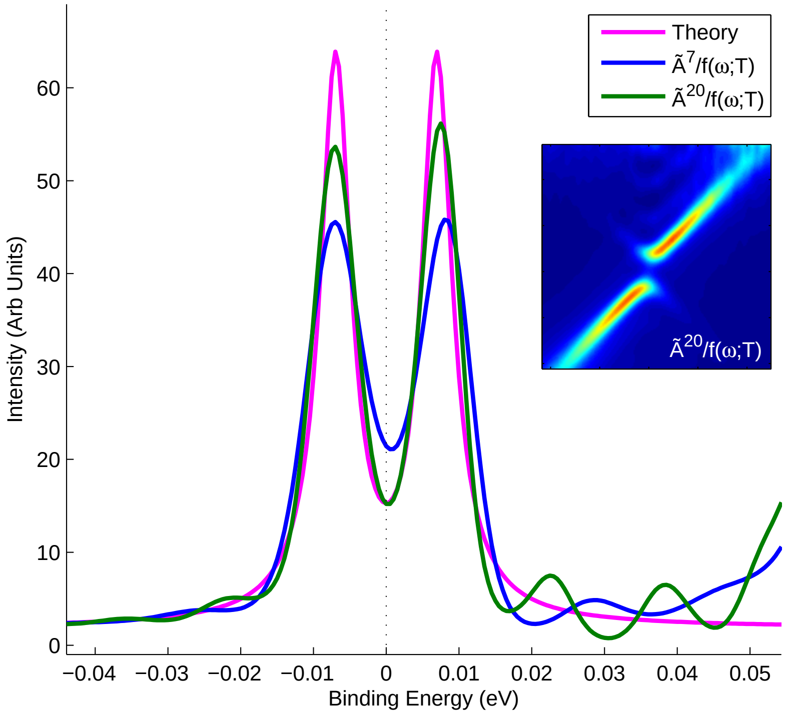


Figure 6.11: EDC's taken at k_F corresponding to the original full theoretical spectral function (pink curve) and spectra divided by the physical Fermi function after seven iterations (blue curve) and twenty iterations (green curve). The inset shows the FNS after twenty iterations.

that exist which, when re-broadened, can produce images nearly identical to that with which we started. In the worst case one might even expect to find chaotic solutions, that is, deconvolutions with an inordinately sensitive dependence upon the initial conditions. This situation would make itself plain if vastly different spectral functions resulted from negligible changes in the resolutions assumed for R. This would be an indication that the underlying features of the spectrum are much sharper than the resolution, a useful piece of information itself but not a situation the experimentalist should relish.

As said earlier, $\chi^2(r)$ provides a useful quantitative check on the number

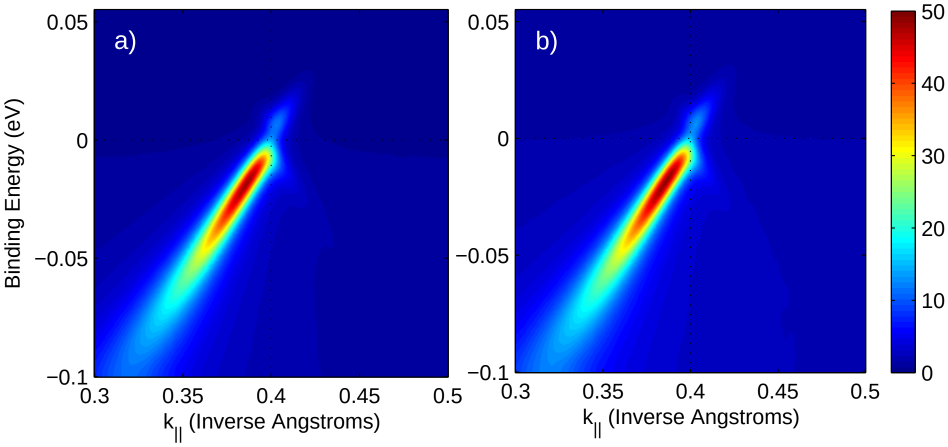


Figure 6.12: Comparison of the occupied spectral function after broadening to the re-broadened, deconvolved spectral function used to calculate $\chi^2(r)$.

of iterations we can justifiably use. $\chi^2(r)$ for the present case is plotted in Fig. 6.13. That $\chi^2(r)$ in this case tends asymptotically towards zero should not trouble since our simulations yield data that is essentially statistically "perfect". What we find in this rather typical case is that, in keeping with our above observations, the greatest improvement in the spectrum is obtained after the first few iterations, after which we run the risk of fitting noise. In this case we are certainly justified in imposing a cutoff at some number of iterations suitable for the purpose we have in mind, e.g. to improve the overall picture or to try to fit as perfectly as possible the main features of the data. Occasionally one will find that a further iteration actually makes $\chi^2(r)$ worse than $\chi^2(r-1)$, in which case the cutoff of the algorithm is imposed upon us. Such convergence is desirable for it's clarity but not necessary.

The second question, that of defining an effective resolution after having

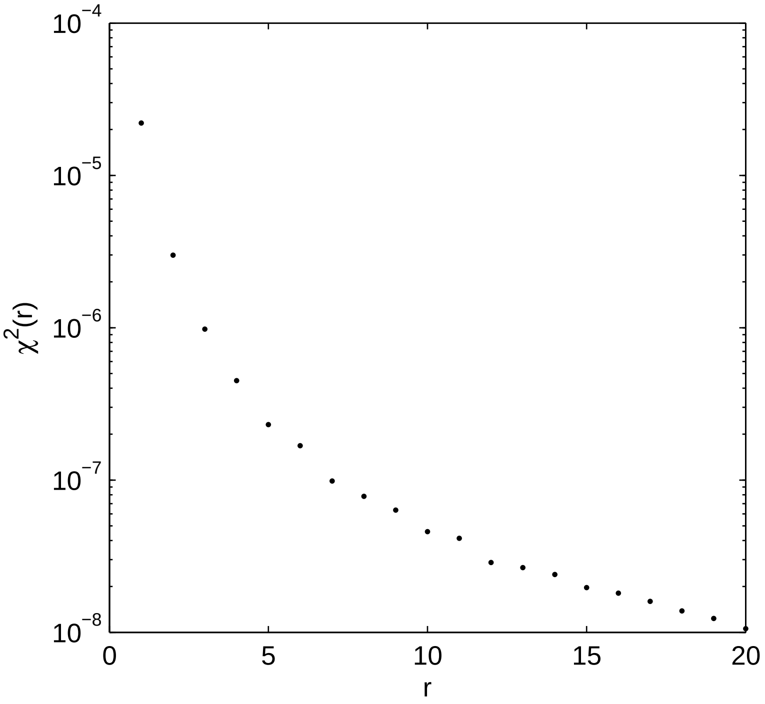


Figure 6.13: Log plot of $\chi^2(r)$.

settled on an appropriate number of iterations, is somewhat trickier to handle. Some progress can be made examining limiting cases. In the event that deconvolution produces no effect on the data it can only be said that either the resolution is better than any feature in the data and thus no computational augmentation is necessary or that any features in the data that are sharper than the resolution lie so close to each other as to be indistinguishable even to the LRM. It is not hard to imagine for instance that data taken with too few pixels, energy-per-pixel steps being too large for example, that we will learn nothing. It is thus always desirable to have some external knowledge, e.g. from theory, to justify our not observing new features. Another limiting case is to start with a delta function, equivalent to the presence of intensity at a single pixel, broaden it and then observe the the results of deconvolution for some number of iterations. By the same token one can examine the improvement of a broadened Lorentzian by fitting after each iteration with a Voigt function, though this may impose a level of physics on the problem that, at least for strongly correlated system, is not well deserved.

Ultimately we wish to impose as little external information on our analysis as possible and thus seek a standard candle that is as simple and as general as possible, that is physically grounded yet requires few assumptions. For ARPES the natural choice is the Fermi distribution itself. This is so because i) we already use the measured Fermi function of a normal metal at a known temperature to establish the energy resolution of an experimental arrangement ii) as mention before, a Fermi function convolved with a Gaussian is another Fermi function at a higher effective temperature, iii) there is a simple algebraic relationship between real and effective temperature and gaussian resolution that can be derived. Thus our program for defining the effective resolution of a deconvolved spectrum should be to generate a Fermi distribution at the effective temperature T' appropriate to our physical temperature T and measured energy resolution ΔE, deconvolve it with the number of iterations used on our spectrum of interest and then fit the result with another Fermi function so we may extract the new effective temperature, call it $T^{\aleph}$, and the effective resolution, $\Delta E^{\aleph}$. Because the Fermi-Dirac distribution is an *idea*, in the Platonic sense, we may without doing an experiment simply construct tables of $\Delta E^{\aleph}(r)$ for the combinations of T and ΔE we are interested in. An example of this procedure is given for a 50 Kelvin Fermi distribution broadened by a 15 meV energy resolution, Fig. 6.14. As expected the effects of broadening are essentially removed after a mere four iterations, yielding an effective resolution $\Delta E^{\aleph}$ of about 2 meV. After four iterations unphysical results are obtained which as usual are the result of the algorithm attempting to fit artifacts generated by itself on our too-perfect representation of the data. Examination of $\chi^2(r)$ (inset) tells us that even in the absence of physical knowledge, e.g. about the experimental temperature, little is to be gained by using more than four iterations and we are well justified in our earlier statement regarding the cutoff of the procedure as it asymptotically approaches some final value of χ^2. Similar figures can be easily produced for any combination of temperature and energy resolution.

With this guide in hand, what are we to make of the preceding results in which we applied the LRM to normal and superconducting simulated bands in which in many respects the fits got better, albeit much more gradually, as we increased the number of iterations? One aspect of this problem is related

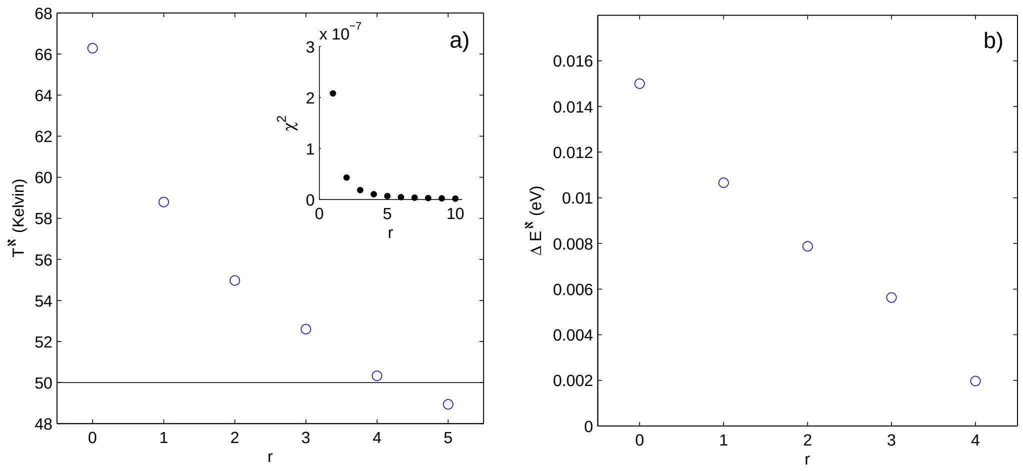


Figure 6.14: Effective temperature $T^{\aleph}$ a) and FWHM Gaussian effective resolution $\Delta E^{\aleph}$ b) fit for a Fermi distribution initialized at 50 Kelvin with an initial energy broadening (FWHM) of 15 meV. The black line in panel a) denotes the "physical" temperature of the Fermi function. The fifth iteration, going too far, yields an imaginary result for the resolution calculated for panel b) (not shown). The inset of the first panel shows $\chi^2(r)$ for the LRM in this instance out to 10 iterations on a linear scale.

entirely to the use of discrete Fourier transforms and the associated irregularities they induce as we iterate. That is, the form these artifacts take for any particular spectrum depend strongly on the details of the features contained in that spectrum and thus it is best to view our calibration of effective resolution as being essentially correct only until it converges. While in principle as we keep iterating past convergence the resolution should stay essentially zero the creeping effect of artifacts will obscure this and possibly cause problems. Thus one must be aware, at least after the first few iterations, of introducing unphysical artifacts into the analysis. It turns out, fortuitously, that the presence of some noise in real data partially ameliorates this problem.

There is however a much more serious, physics based problem to be concerned with when attempting to calibrate the effective resolution of a deconvolved spectrum. That is, in our analysis we have implicitly assumed that the Fermi function represents in some sense the sharpest possible physical feature that can be found in the data. In other words, we do not expect to find states (or bands) whose intrinsic width as measured by EDC's will be *sharper* than $k_B T$ and so, if we can deconvolve the Fermi function every other feature will have been brought into line. We have not however directly addressed the deconvolution of other line shapes, such as Lorentzians, except in the simulations above. In those cases even though the FWHM of a lineshape might be of the same order as the Fermi distribution, there are clearly parts of the function that will be narrower and thus require more iterations to fully deconvolve. Further complications with this picture are introduced when considering, for example, the shape of a band crossing the Fermi level. These problems are apparently easily dealt with by examining the internal consistency of the deconvolution and the rate at which $\chi^2(r)$ changes.

A more serious, and much more physically interesting problem, arises if we consider the possibility of states that *are* intrinsically narrower than the temperature by any definition. For the vast majority of conventional systems that such a state could exist would seem to be rather counterintuitive. Certainly this should be true for a Fermi liquid in which the electrons couple to the phonon bath with a T^2 dependence while the width of the Fermi function is approximately $4k_BT$. Even at very low temperatures one expects some broadening due, for example, to impurity scattering. Nevertheless it can be argued that we should not rule out the existence of such features *a priori* in strongly correlated or other pathological systems. Wei Ku has provided the rather elegant counter example to our expectation by considering the presence of a single impurity in an otherwise ultra pure normal system such as a metal. That impurity will acquire through Anderson localization an isotropic k dependence in its spectral function and an essentially delta function like width in energy that will not appreciably change with the bath temperature. More generally we should recognize that the temperature dependence of the width of a state, given above by $(\gamma k_BT)^2$, results from a self energy. It should not be ruled out that the form of this self energy can be drastically modified in strongly correlated systems such as the cuprates. One example of such a modification is the so-called Marginal Fermi Liquid.

Two questions about the resolution linger. First, what if the resolution we divine from experiment is for some reason wrong? Second, what if the kernel of the resolution, R, is not truly Gaussian. As to the correctness of the magnitude of resolution broadening divined from experiment, it is up to the experimenter to prove the veracity of his or her claim. Besides case by case measurements of the resolution usually obtained in the course of an experiment most ARPES systems have undergone extensive calibrations with respect to various parameter settings yielding particular energy and angular resolutions. So long as the calibrations and the spot checks are in agreement there is no problem. Further one can often reference features in particular spectra of samples to use as an internal consistency check. This is a matter of just being competent. Any intrinsic broadening e.g. due to disorder on a sample surface, should no be deconvolved if all that is accounted for in the LRM is instrumental broadening. As to the Gaussian nature of the resolution, it may well contain small contributions of higher order but in this case the operative word is small. In the many years high resolution ARPES has been practiced there has not been any obviously compelling evidence that the resolution function is not Gaussian. If this is the case then producing FNS by using Gaussian broadened Fermi functions is even worse than suggested above.

6.6 Deconvolution and Fermi Normalization of Real Data

Having explored in some detail the theoretical and technical issues involved in applying the LRM to ARPES data we are at last in a position to examine the results of a real experiment. All data presented in this chapter was acquired at the undulator beamline U13UB on the VUV ring of the NSLS. The beamline endstation was equipped with a Scienta SES-2002 electron spectrometer.

Further details of the experimental apparatus will be given in a later chapter. It suffices now to note that the beamline/spectrometer combination in use, while in principle capable of achieving better than 1 meV energy resolution, was set to an energy resolution on the order of 15 meV in order to maintain an acceptable signal-to-noise ratio. Each scan presented in this chapter still required acquisition times of five to six hours – much longer than is normal for a modern ARPES experiment – in order to accumulate enough statistics to render visible the low lying states above the Fermi level. While the data contain a wealth of information it is not our goal in this chapter to delve too deeply into the physics but rather to demonstrate the efficacy of using the LRM to enable more accurate and detailed analyses. We shall where appropriate point out the most significant features in the data.

The first example to be analyzed was acquired on a high quality single crystal sample of optimally doped BSCCO (T_C about 91 K). It was cleaved *in situ* at low temperature. The experimental temperature was 60 K and the energy and angle resolution were 16.5 meV and .1 degrees, respectively. The pressure in the endstation chamber was $5X10^{-11}$ Torr. The photon energy was 20.1 eV and was p polarized relative to the sample. The PC mode of data acquisition was used rather than the ADC mode. The cut in k space, located in the first Brillouin zone, is in the MX direction about half way between Γ and M. The misorientation of the cut with respect to the ΓM high symmetry line is responsible for the drastic change in matrix element (e.g. band intensity) across the spectrum. The spectrum displays many of the generic features often seen in this region of the Brillouin zone for BSCCO. Even in the original spectrum, Fig. 6.15a), one observes the presence of the bilayer splitting of the main BSCCO valence band. The inner parabolic band is identified as the antibonding (AB) band and the outer band as the bonding band (BB). The matrix element suppresses the BB on the left side (negative momentum) of the spectrum. The broadness of the bands at high binding energy is usually associated with its interaction with a strong bosonic mode. The sharp, flat band with zero dispersion running across the gap edge below E_F is usually associated with impurity scattering. Evidently scattered carriers are also bound into Cooper pairs. A large superconducting gap Δ of about 20 meV as measured from the AB peak is clearly visible. Thermally excited quasiparticles appear as a dim smudge traversing the Fermi level from the top of the AB band. This dim feature is the object of our study. Upon deconvolution of the spectrum, Fig. 6.15b)-e), the thermal excitation feature is seen to sharpen and ultimately to form a peak above the Fermi level separated from the negative binding energy states by a clear superconducting gap. As promised, there is no obvious sign in the data, even up to ten iterations (Fig. 6.15e), of the sorts of Fourier transform induced artifacts observed above in simulation. Further, the occupied part of the spectrum is clearly sharpened by the procedure which corrects both the observed lifetimes and dispersions. The dim band on the left hand side, in which one band is almost totally suppressed, even at this point displays the sort of folding associated with the opening of a superconducting gap in the BCS theory.

The various methods by which one can produce FNS are demonstrated in Fig. 6.16 with corresponding EDC stack plots given in Fig. 6.17. Panel a) shows the result of normalizing the raw data by a Fermi function set at the experimental temperature T. While the gap minimum remains at the Fermi

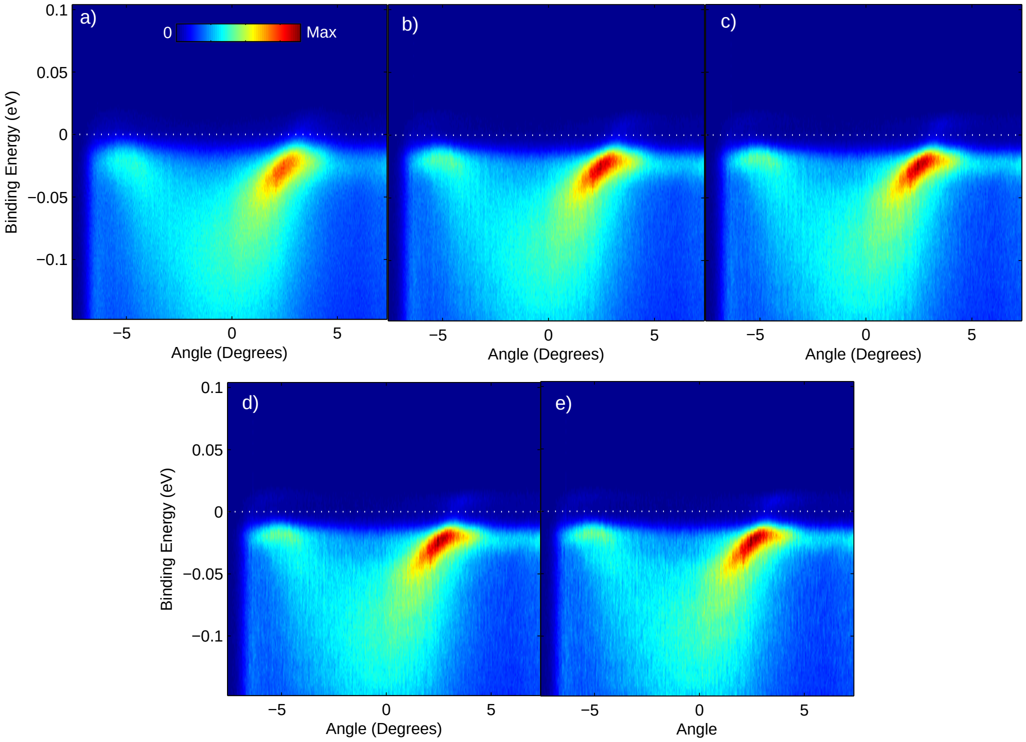


Figure 6.15: Cut in k space taken at location diagrammed in inset of panel e) on optimally doped Bi2212 at 60 K. Panel a) shows the original spectrum for which the experimental energy and angular resolution was set to 16.5 meV and $.1^o$, respectively. Panels b) through e) show the result of applying the LRM to the spectrum in panel a) after 4, 6, 8 and 10 iterations. The false color intensity scale for all panels is shown in the inset of panel a). The Fermi level is indicated by dotted lines.

level as it should for a superconductor, the intensity of thermally excited states is clearly too high relative to that of the occupied states. Naturally they are also broadened to the point where they are not clearly recognizable as bands. On the other hand, normalization by a Fermi distribution broadened to the effective temperature T' underestimates the intensity of states above E_F and does nothing to clarify their dispersion. One should also note the position of the positive high energy cutoff of the spectrum. Because quasiparticles can only be excited to 4 or 5 k_BT above the Fermi level any real spectrum, when normalized to a Fermi function, will generate a region of essentially infinite intensity values. This is due to the division of small finite numbers generated e.g. by detector noise, camera noise or high energy secondary electron emission by a number that is very close to zero. The location of this cutoff is determined by the temperature (4 or 5 k_BT) and the signal to noise of the thermally populated states. The shift of this cutoff between panels a) and b) is thus an artifact of the unphysical nature of FNS generated with resolution broadened Fermi functions. Another telling feature of panel b) is the shift of the gap minimum away from zero energy. This is a serious problem if one is attempting to, for example, find some violation of our current understanding of superconductivity in the cuprates viz a viz how far we can justify our use of the BCS picture explored above.

Panels c)-e) of Figs. 6.16 and 6.17 show the result of Fermi normalizing spectra deconvolved by the LRM after 4, 8 and 10 iterations. After only four iterations, a number which is consistent with our observations in Fig. 6.14 (which hold for the current situation) we find that the gap minimum remains at zero energy as in panel a) but that the thermally occupied band above E_F

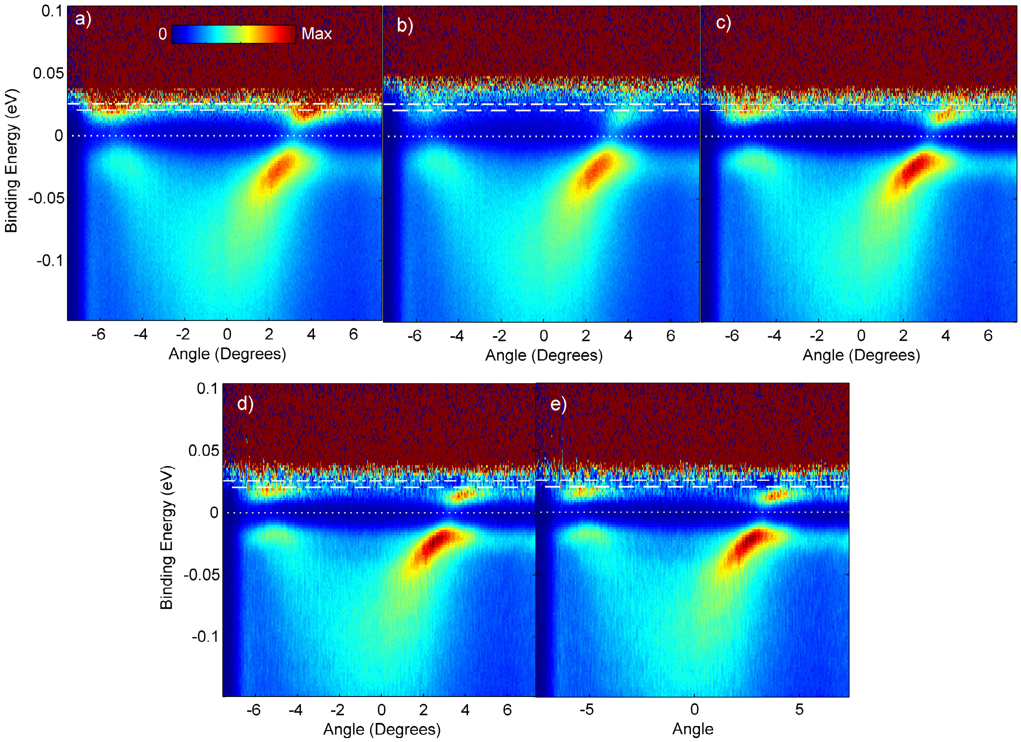


Figure 6.16: Comparison of different Fermi normalization methods applied to the data in Fig. 6.15. Panel a) results from normalizing the original data (Fig. 6.15a) to a Fermi function at the experimental temperature. Panel b) shows the result of normalizing the same spectrum to a resolution broadened Fermi function. Panels c)-e) show the result of normalizing to the experimental Fermi function after 4, 8 and 10 iterations, respectively. The Fermi level is indicated by dotted lines. The dashed lines indicate $4k_BT$ and $5k_BT$, respectively. The false color intensity scale is the same as Fig. 6.15.

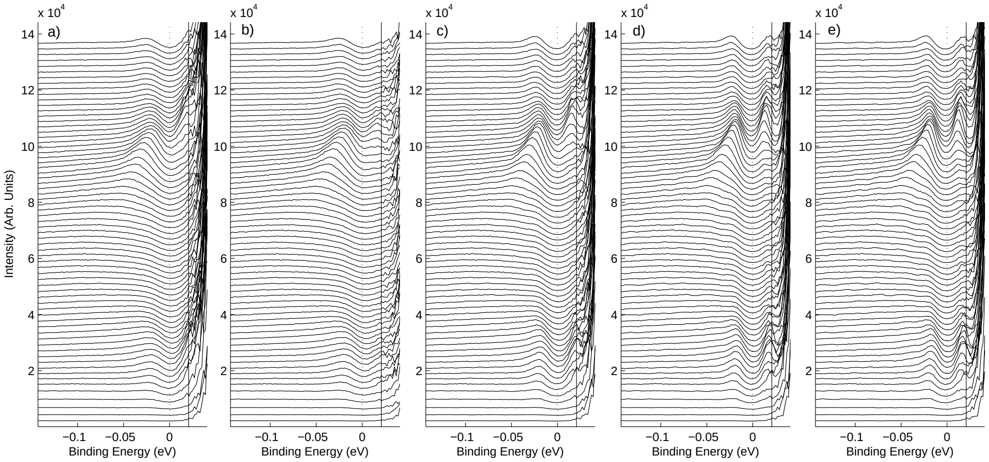


Figure 6.17: EDC stack plots generated from corresponding panels in Fig. 6.16. The Fermi level and $4k_BT$ are marked by the solid lines.

is now sharp enough to identify a dispersion and has a spectral weight as a function of momentum in line with what we expect from the coherences in Eqs. 6.24 and 6.25. Thus the LRM has allowed us to confirm the BCS-like nature of Bogoliubov quasiparticles of the superconducting state of optimally doped BSCCO below T_C. The further number of iterations in panels d) and e) of the figures show that continued application of the LRM shifts the gap minimum away from zero as would be expected for over correction of the underlying Fermi distribution in the data. On the other hand, very little happens to the occupied bands after four iterations indicating they are now essentially resolution free and are too close together to be uniquely resolved by our method. The thermally occupied bands do however increase in sharpness and change their dispersion upon further iteration. In light of the results of our simulations above of how broadened Lorentzian peaks are affected by the procedure, both above and below E_F, it is not entirely unreasonable to expect some further sharpening of peaks. We thus conclude that the thermally excited peaks probably have a lifetime much longer than might be expected based purely on coupling of the bands to the thermal bath and at least for this resolution it is not possible to achieve a unique determination of the lifetime and the Fermi normalization simultaneously for this level of broadening. This conclusion is supported by examination of $\chi^2(r)$, Fig. 6.18, in which we clearly have not yet converged after four iterations. Because we are depending upon the rectification of the Fermi distribution for our normalization procedure and because we cannot assume a particular lineshape for the bands *a priori* any further analysis should be carried out on the data sharpened with four iterations. Nevertheless further deconvolution is not useless as we can see the wall of high energy noise cut off almost exactly at $4k_BT$ in panel e).

The result of this analysis is a picture that, if not containing features as sharp as they could be in reality, is at least much better than what we started with and entirely self consistent which, more so than absolute values of intensities, is the most desirable feature to find in an ARPES experiment. It must also be recalled that the more iterations we perform, the more likely it is the procedure will begin to fit "noise" and generate instabilities in the image which will be further amplified with each iteration. Bands above E_F are, as above, particularly sensitive to this problem and so we should probably not trust further iteration anyway. To the LRM they can, after a few iterations, begin to look like noise due to their relatively low signal. In any event, four iterations is more than adequate to extract the underlying physics of the experiment which was not possible without the LRM.

We conclude our analysis of sample ARPES data with the case most like

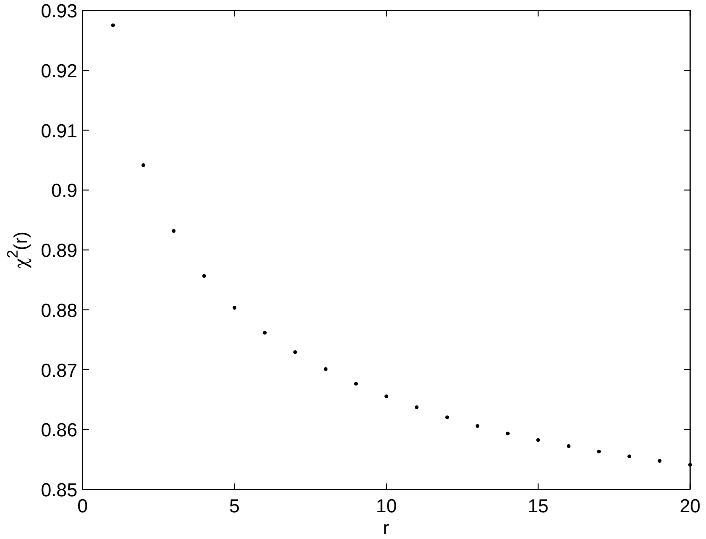


Figure 6.18: $\chi^2(r)$ for the deconvolution of Fig. 6.15.

that simulated above, that of a single band crossing the Fermi level interrupted by the opening of a superconducting gap. Without presenting the gory details of the analysis, we cite the example of the comparison between different methods of generating FNS first reported in the literature. This cut, as illustrated in the inset of Fig. 6.19, was close to the node of optimally doped BSCCO below T_C. The resolution was 15 meV and .1 degrees, respectively, and the temperature was also 60 K. The improvement with application of the LRM is self evident, as is the recovery of the BQP dispersion.

The reader will no doubt wonder now at the discrepancies between our deconvolution of real versus simulated data. The main differences between the two sets of results are a) deconvolution of real data is much less sensitive to computational artifacts such as FFT induced reflections than are the simulations and b) unlike the simulations, when deconvolving real data physically

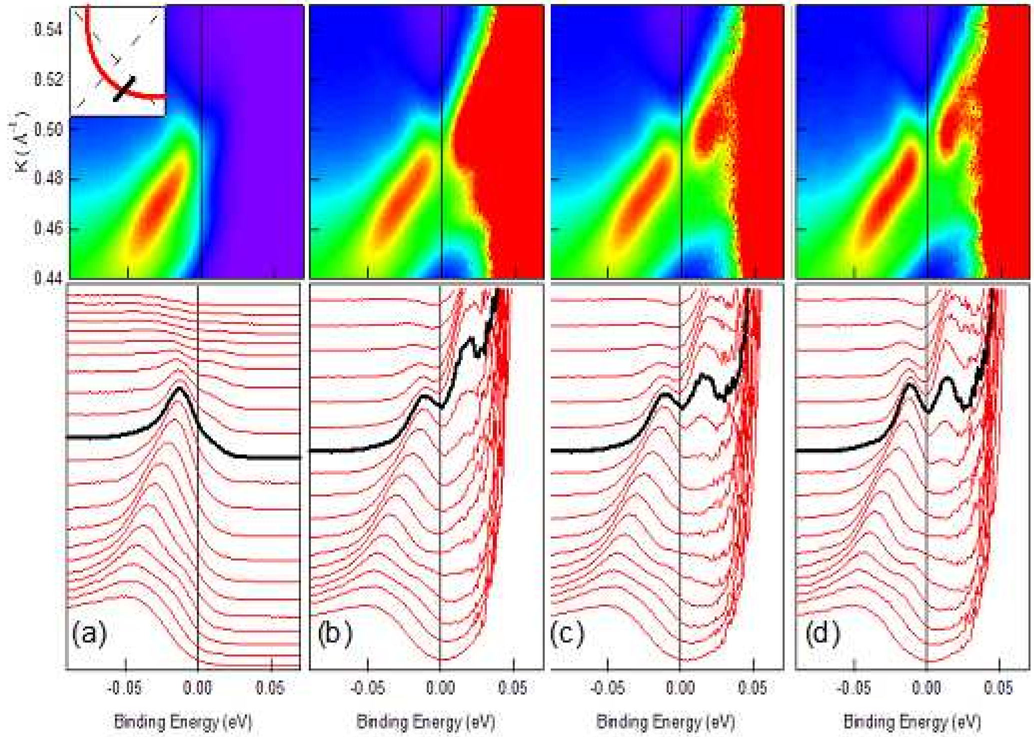


Figure 6.19: (a) the raw data before analysis, (b) the same data as in (a) after dividing by the Fermi function, (c) the spectra after dividing the data in (a) by the Fermi function broadened with a function representing the experimental resolution and (d) the deconvoluted spectra (derived from (a) using the LRM) divided by the Fermi distribution function. The solid black lines represent the EDC corresponding to the Fermi wave-vector, k_F.

consistent FNS could only be obtained up to the number of iterations consistent with deconvolution of the Fermi distribution whereas in simulations we were more or less free to iterate to our hearts' content. The first point is easily explained by the dampening effect that real noise has on FFT reflections, "softening" any sharp edges that might otherwise be found in the data.

The second point is more interesting and has several possible explanations. One possibility is that the accumulation of small experimental errors in our measurement of the energy and angular resolution as well as the kinetic energy of the Fermi level might well lead to a residual asymmetry in the FNS as well as errors associated with using a slightly incorrect resolution when deconvolving the data. These errors can easily result, for example, from small differences in sample charging and signal to noise between measurement of the Fermi distribution on the reference sample (e.g. polycrystalline gold) and the sample itself. These errors do not exist for the idealized situation embodied by our simulations. The technique might benefit in the future from a quantitative and systematic understanding of how such errors affect deconvolution of ARPES data though such an investigation is beyond the current scope of this work. Problems may also arise if the resolution function for ARPES is not truly Gaussian. In light of the apparent success in using the Gaussian kernel on our measurements such non-Gaussian contributions to the resolution will most likely appear as higher moment corrections to the Gaussian form and, once known, could easily be incorporated into the methodology. Nevertheless such higher order contributions are likely to be small and so are unlikely to effect the results greatly if we limit our iteration number to the few necessary to remove the worst effects of the broadening.

Future work can most fruitfully proceed along two tracks. First, one would like to better quantify the functional form of the resolution broadening experimentally for a range of photon sources and electron analyzer configurations. For example, the kernel likely differs somewhat when using a curved versus a straight entrance slit in the electron spectrometer. Another approach would be to extend the computational methods outlined here to a more sophisticated rendering of the problem. One direction for future work would be to attempt the implementation of so-called blind deconvolution algorithms. In those programs the resolution kernel is provided as an initial *guess* along with the data. Then the algorithm not only produces an image that was most likely to have existed prior to broadening but also a self consistent approximation to the (presumably poorly understood) kernel that caused the broadening in the first place. Ideally these experimental and computational investigations into the true resolution function of ARPES would be brought to some kind of agreement.

To summarize, we have in this chapter laid out the essential idea of the LRM and shown that because the resolution kernel of ARPES is quite simple and easily obtained to first order, it is very well suited to the improvement of raw data without any prior assumptions about the underlying physics. We went on to apply the method to some rudimentary simulations of spectral functions for both normal and superconducting states and, though sampling only a very small part of the parameter space for such work, outlined the advantages and some of the pitfalls such analysis can bring to bear on the problem of analyzing ARPES data. With these guidelines in hand we applied the method to samples of real ARPES data, by which we were able to resolve physically relevant features that were not previously discernible despite our best experimental efforts. Overall, we have shown that the LRM is at least as useful for ARPES as is it has proven itself for the analysis of medical and astronomical imagery. We have also tried to point out where null results are useful for determining whether features in various spectra are broad enough to be unaffected by finite resolution or are anomalously sharp. Finally, we pointed out where experimental errors will affect the deconvolution procedure, where we are likely to remain safe and how these issues might be addressed in future studies. We leave this chapter with the caution that as great as the LRM's potential is as a tool for improving the results of very difficult ARPES experiments, its potential abuse by the unwary or unscrupulous user is far greater. It is the great ease with which these methods can be applied that make it a dangerous tool in the wrong hands and so the results of its use, as with the use of any experimental or analytical technique, must always be met with deep thought and deeper scrutiny.

Chapter 7

Laser ARPES on Optimally Doped Bi2212

7.1 Introduction

It has been over a decade since ARPES gained the high energy and momentum resolution necessary to discern in detail the presence of mass renormalizations, or "kinks", in the electron spectral function of solids induced by electron-boson couplings or other strong interactions. The enumeration of these interactions and the identification of the Bosonic modes presumed to be responsible for them has played a critical role in our understanding of electronic structure in general and has been of particular importance in the study of correlated electron systems. This has been most notably true for the copper oxide high temperature superconductors due to the important role such couplings are thought to play in the anomalous physics of these materials. Unfortunately, because ARPES can only probe the effect such interactions have on the single electron density of states, rather than the boson spectrum itself, it has proven necessary to look to other experimental or theoretical methods to determine the boson, if any, responsible for these features. This fact has often resulted in wide disagreement as to the nature of such couplings because of the often overwhelming number of candidate bosons present at the relevant energies and momenta. As a result of these controversies the role, if any, that such interactions play in achieving high superconducting transition temperatures T_c in the cuprates has remained clouded in doubt.

Below we present evidence for an additional kink in the nodal region of the low energy ARPES spectrum of optimally doped Bi2212 at low temperature. While the existence of such a feature in the Bi2212 spectrum has been previously hinted at there has yet to appear in the literature any substantial study of its character. It will be shown below that this new kink exhibits a unique dispersion with respect to the superconducting gap in the nodal region of the Brillouin zone. That is, it always appears at 8 meV high binding energy than the superconducting gap regardless of the gap's magnitude. Because of this behavior it is possible to rule out all but a single optical phonon mode as being responsible for the observed interaction in the electron spectrum. Besides being interesting in its own right as an example of an electron-phonon coupling not previously observed by ARPES the fact that this is the only such interaction observed to date in Bi2212 that can be resolved with certainty will provide a much needed moment of clarity in the study of the physics of this highly complex system.

7.2 Laser ARPES Data and Initial Findings

The construction and operation of the laser ARPES system at the NSLS has already been explored in detail in chapter 5. The remainder of this chapter will be devoted to the presentation and explanation of the best data resulting from that experiment. We believe the result represents one of the few easily explained phenomena to be discovered in the study of the electronic structure of Bi2212. Once again discovery came as a pleasant surprise rather than as the result of a concerted effort to solve a particular problem.

When commissioning the laser ARPES system the first sample we measured was meant only to check that our experiment was functioning properly in angle resolved mode and that it had at least a respectable energy resolution considering the difficulties imposed on the project outlined in chapter 5. The “figure of merit” for high resolution ARPES, and in particular for those ARPES groups that spend most of their time thinking about high temperature superconductivity, is the nodal spectrum of optimally doped Bi2212. This is because optimally doped Bi2212 is one of the best known, most studied sample varieties in condensed matter physics and because at Brookhaven we are blessed with an essentially inexhaustible supply of ultra high quality optimally doped Bi2212 grown by Genda Gu in the neutron scattering group. These samples are world renowned for their order and their purity and represent the gold standard in floating zone crystal growth for copper oxide superconductors. Further, the optimally doped variety of Bi2212 is the most stable version of the doped system and each sample from a given grow rod, and each sample of ours acquired from different growths, are as identical as these systems can be with T_c's ranging at most between 89 K and 91 K.

Our measurement was performed at a temperature of 10 K with the same open cycle helium cryostat, electron spectrometer and vacuum as reported above. The sample was cleaved at 10 K and the Fermi level was referenced to a bulk specimen of gold in electrical contact with the Bi2212 sample. All the data reported here were acquired within two hours of cleaving *in situ*. The analyzer was set to use the 1 eV pass energy with an entrance slit of .5 mm. The results of our measurement of optimally doped Bi2212 were gratifying if not immediately appearing to be remarkable in any way. Aside from the incredibly high signal to noise ratio expected from a source as bright as the laser we only noted that spectra taken relatively far from the node appeared to display a very sharp “kink” in the dispersion close to the gap edge and, without thinking about that too much further, we assumed it was just a result of seeing the BCS-like dispersion accompanying the opening of a large superconducting energy gap. The raw spectra (converted to units of momentum and binding energy) are plotted in Figure 7.1a-c moving progressively further away from the node. The locations of these cuts in the Brillouin zone are shown in the inset of Figure 7.5a.

Accepting the energy resolution of 13 meV and angular resolution of .1 degrees these spectra were acquired at we proceeded to apply the LRM discussed earlier to the present experiment. The resulting deconvolved spectra are shown in Figure 7.1d-e. Raw and deconvolved EDC's and MDC's from the near nodal data shown in Fig. 7.1, panels a and d, are shown in Fig. 7.3. Lorentzian fits to the deconvolved distribution curves are shown in Fig. 7.4. Both fits demonstrate the the combination of laser ARPES and the LRM appear to

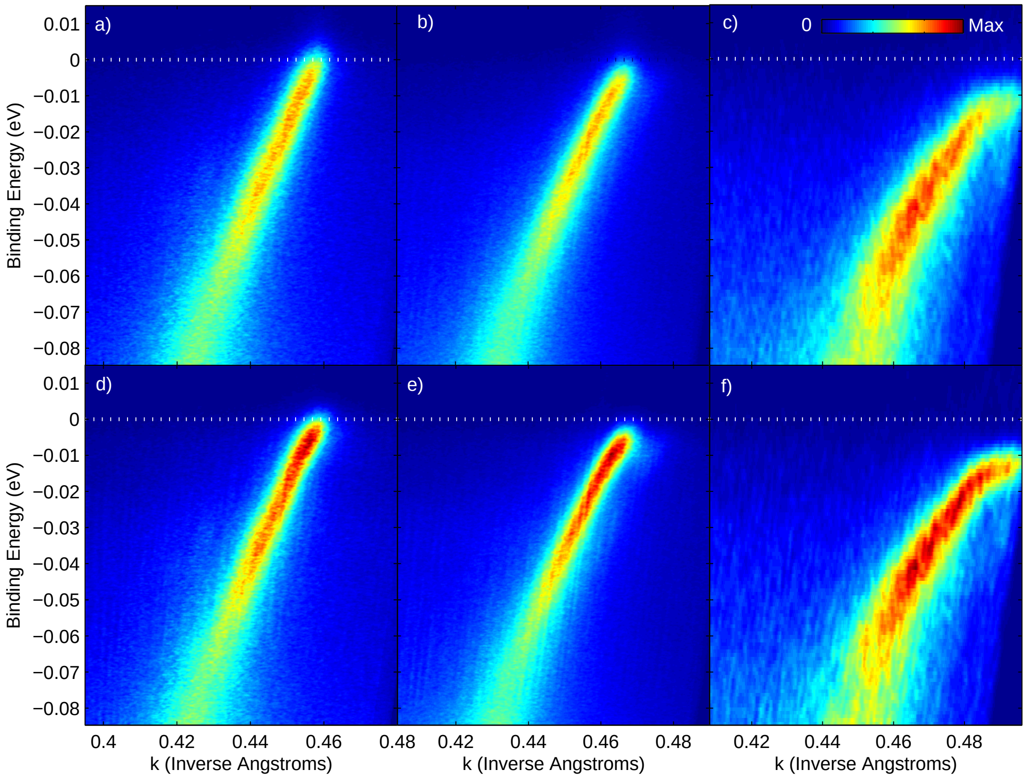


Figure 7.1: Laser ARPES spectra from the nodal region of optimally doped Bi2212 at 10 K. Panels a-c are the raw data. Panels d-f are the same spectra deconvolved using the LRM. The locations of the cuts in the Brillouin zone are diagrammed in the Figure 7.5a inset.

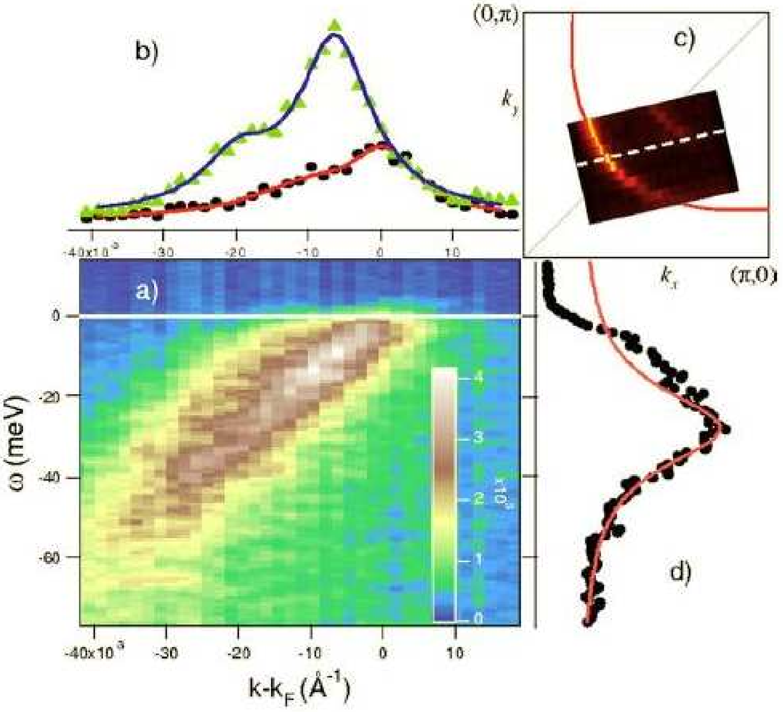


Figure 7.2: Near nodal spectrum of optimally doped Bi2212 at 8 K taken at the VUV ring of the NSLS with a photon energy of 16.7 eV. The orientation of the cut in the Brillouin zone is shown by the dashed line in panel c). b) MDC's at $\omega = 0$ meV (circles) and -10 meV (triangles). The peaks are fit by two Lorentzians. d) EDC at $k - k_F$= -.015 Å^1 with a single Lorentzian fit to the high binding energy (bonding band) peak. [3]

provide access to the intrinsic energy and momentum line shapes of Bi2212. The improvement in momentum resolution over what can be achieved using synchrotron radiation as a photoexcitation source is evident from a comparison to the spectrum, EDC's and MDC's shown in Fig. 7.2 which was acquired at an energy resolution of about 5 meV at a photon energy of 16.7 eV [3] on another optimally doped Bi2212 sample. The data were acquired with the same endstation, at beamline U13UB of the NSLS, as was used for the laser ARPES experiment. At that photon energy the momentum resolution is roughly three times less than is achieved with the laser for the same angular resolution. The drastic change in transition matrix element between the two photon energies and the drastically improved signal to noise achievable with the laser is also clear.

To be fair we were a bit shocked to find that when this was done an entirely new feature in the very low energy spectrum of Bi2212 appeared to emerge. That is, there appeared to be a very small "kink" in the dispersion setting in well short of the superconducting gap energy. Recall that in Chapters 2 and 3 we briefly touched on the issue of how an electron-boson interaction is observed by ARPES in the single particle density of states. ARPES measures the single particle spectral function $A(k,\omega)$ which is related to the many body electron Greens function by the relation $A(k,\omega) = \frac{1}{\pi}|\operatorname{Im} G(k,\omega)|$ where in the presence of an interactions $G^{-1}(k,\omega) = \omega - \varepsilon_k - \Sigma(k,\omega)$ with $\Sigma(k,\omega) = \operatorname{Re}\Sigma(k,\omega) + \imath \operatorname{Im}\Sigma(k,\omega)$ the real and imaginary parts, respectively, of the electron self energy. In Chapter 6 we modeled several spectral functions based on such Greens functions in which the self energy was taken to be that of a Fermi liquid. In those cases (ignoring the presence of a superconducting gap) the

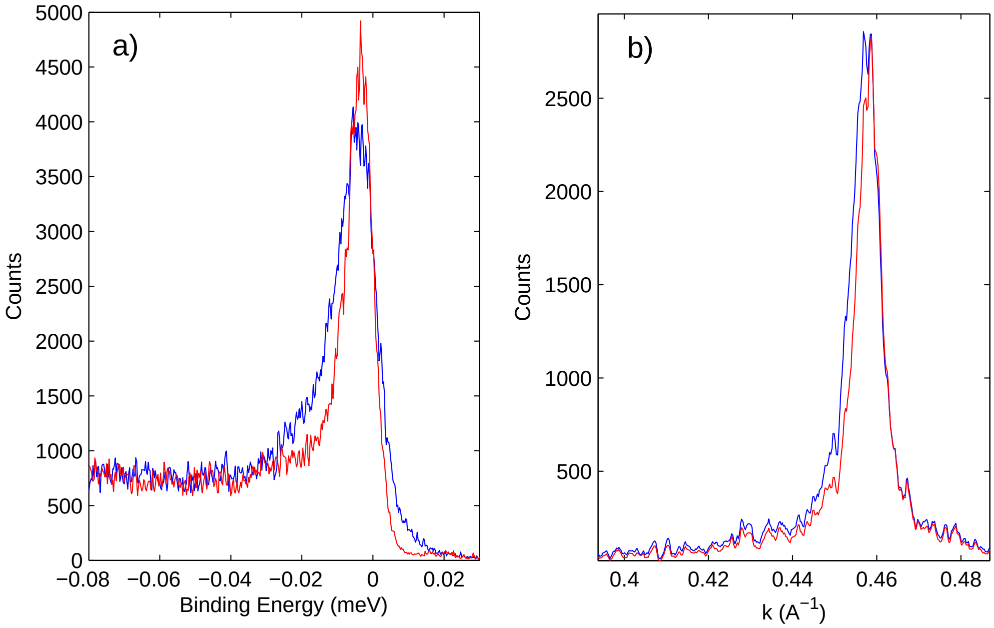


Figure 7.3: Panel a) EDC's and panel b) MDC's taken at the k_F and E_F, respectively. The blue curves are from the raw data of Fig. 7.1a and the red curves from the deconvolved data of Fig. 7.1d. All cuts are taken along a single column or row of channels demonstrating the superior signal to noise that can be achieved with a laser. Note that in addition to sharpening the peaks the peak positions themselves are corrected by application of the LRM.

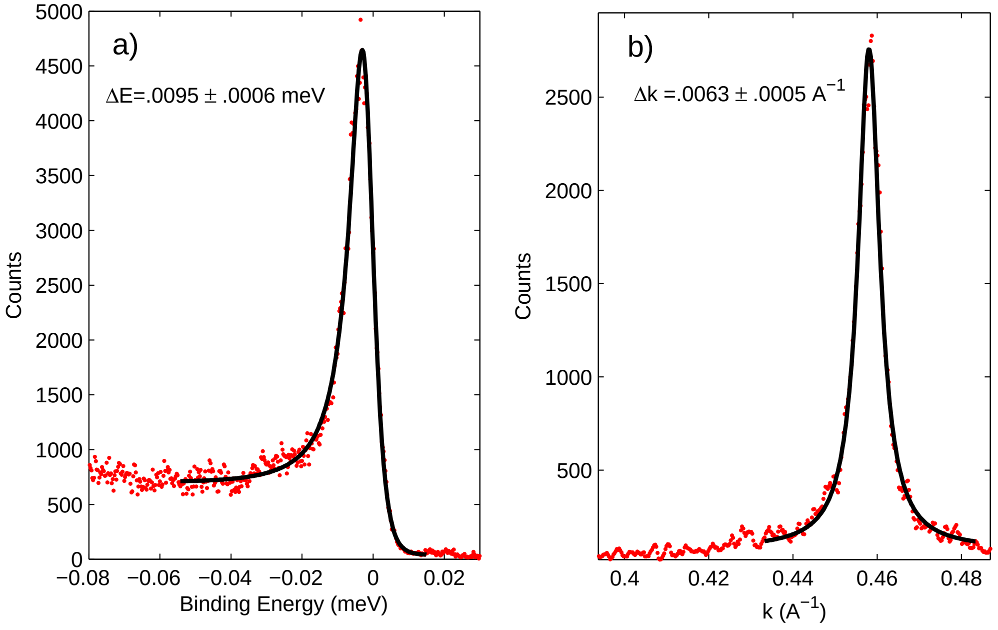


Figure 7.4: Fits to the deconvolved EDC (panel a) and MDC (panel b) shown in Fig. 7.3. Both fits are performed using Lorentzians on a linear background. The EDC fit is cut off by the Fermi-Dirac distribution. Peak widths ΔE and Δk are shown in the figure.

entire spectrum was taken to derive from a single self energy function extending over the full energy and momentum ranges of the bands under investigation. When considering Landau quasiparticles this self energy reflects the dressing of the bare electron mass by electron-electron interactions in such a way that bare electron states can be mapped one-to-one onto quasiparticle states with the renormalized effective mass m^*.

Taking the physics a step further one can also consider the effect upon the spectral function of the presence of other interactions beyond the electron-electron interaction. These effects are derived from the presence of a bosonic mode with which the electrons interact. Such an interaction will change the dressing of the electron, i.e. it will be further renormalized. In the language of field theory this renormalization takes the form of an additional self energy term in the electron Greens function that sets in at a particular energy and momentum corresponding to that of the bosonic mode. It can be shown in a generic quantum field theory that when considering multiple renormalization effects, e.g. from a mass renormalization, vertex renormalization, etc., to first order the self energies derived from the fields add linearly in the inverse Greens function. Taking $G_0(k, \omega)$ to be the Greens function of a bare band we find that the overall Greens function of a system including interactions is $G_{int}^{-1}(k, \omega) = G_0^{-1} + \Sigma_1(k, \omega) + \Sigma_2(k, \omega) + ...$ where G_0 is taken to be the bare Greens function plus contributions from various interactions. This relationship can be trivially though usefully rearranged to display how the effect of multiple interactions stack;

$$G_{int}^{-1}(k,\omega) = G_0^{-1} + \sum_{n=1}^{\infty} \Sigma'(k,\omega) + \imath \sum_{n=1}^{\infty} \Sigma''(k,\omega). \tag{7.1}$$

The simplest addition to the single particle self energy one can conceive of, after the inclusion of an energy independent impurity scattering, is a plain vanilla electron-phonon interaction. This interaction represents the ability of an electron with sufficient energy and appropriate momentum to emit and absorb phonons by scattering off the ionic lattice of the crystal. The availability of this interaction channel renormalizes the electron mass from m to some effective mass m^* characterized by a coupling constant $m^*/m = 1 + \lambda$. A change in the relationship between electron momentum and energy, i.e. the dispersion relation, causes the band velocity to also become renormalized and so introduces a "kink" in the dispersion around the phonon mode energy. At energies below the kink an electron does not possess sufficient energy to excite the phonon mode and so has a longer lifetime (sharper Lorentzian width). The change in dispersion in this region is most pronounced. Above this energy, when the electron can easily gain or lose energy through the new interaction channel, the lifetime decreases – reflecting the increased scattering – and the velocity changes to eventually reflect the bare mass. The presence of the additional self energy renormalizes the electron lifetime and binding energy, the former deriving from the imaginary part of the self energy and the latter from the real part, the two parts being causally related by Kramers-Kronig transformations. Thus, to prove that the feature we are here discussing is truly a "kink" in the sense that it is caused by the onset of an electron-boson coupling we are required to observe not only a change in the dispersion, which if the coupling is small can be quite difficult to detect, but also a change in the scattering rate above and below the mode energy. For reasons explained below, this is not such a simple task in our case.

7.3 Extraction of Quantitative Information from the Data

As we've said often throughout this work the low energy nodal spectrum of optimally doped Bi2212 is one of the most highly examined, best known results to come out of ARPES on high temperature superconductors. That *yet another kink* in the spectrum, and at such low energy, could have gone unnoticed for so long is striking. One has every reason to ask how that could be. The answer, both from this data and from a reexamination of the literature, is fairly straightforward and a good lesson in the assumptions that go into examining any type of experimental data. Firstly, as can be seen in the raw data, even a very small amount of resolution broadening almost completely wipes out this feature. It can only be seen in the raw data shown in panel c because of the kinematic effects of using a very low energy photoexcitation and because of the odd angle relative to the normal of the Fermi surface at which it was acquired. Additionally, ARPES at 6 eV has what is in this case a distinct advantage over higher photon energy data in the form of an extreme matrix element effect. (The ARPES matrix element was discussed at length in Chapter 3.) It has been observed in previous ARPES experiments on Bi2212 using excitation energies less than 8 eV that for reasons not entirely clear the antibonding band is entirely suppressed by transition matrix element effects[20]. In our own spectra it is obvious that only one band can be seen. This is in stark contrast to a spectrum shown in Figure 7.2. In this case it was clear that even at the node there is a small but observable bilayer splitting in the band. In such a circumstance it is almost impossible to tease out such a fine feature as the one here presented from even the best data in a case when you observe *two* bands lying almost but not quite on top of each other. That this is the case for almost all ARPES spectra acquired with synchrotron radiation as presented in the literature will clearly hide the existence of this very low energy kink.

This is not to say the “little kink” was unobservable. The data set from which Figure 7.2 was taken did yield a hint of the little kink but it was difficult to verify and impossible to track any of its systematics at the time. The possible existence of such a feature has even been indicated recently in the literature from the results of an experiment performed with a 7 eV laser. Nevertheless as that study made clear the reflexive use of our favorite technique for analyzing nodal ARPES spectra, the use of MDC’s and band velocities to determine the band dispersion and imaginary part of the self energy, rather than the direct EDC method, also obscures the result. Recall that $2\,\mathrm{Im}\,\Sigma = \hbar v \Delta k$ where Δk is the MDC width at some energy and v is the band velocity at that energy. This formulation only works in the approximation that the band has linear dispersion and a high velocity to begin with. We can see that in the event the band velocity is changing very rapidly as a function of energy in our region of interest it will be almost impossible to determine v with any kind of certainty and thus impossible to determine $\mathrm{Im}\,\Sigma$ from MDC's alone. This is the case for our little kink. To make matters worse the MDC analysis fails utterly in the vicinity of the superconducting gap because not only is the velocity changing rapidly there but it is also changing its overall character from linear to parabolic and thus exiting the high velocity regime for a very low velocity regime in which MDC's are not clearly meaningful.

In order to study this feature in the spectrum and determine if it really is due to a kink and thus due to a previously unidentified interaction of the low energy states of Bi2212 with some bosonic mode (more about which we will discuss below) it was necessary to examine the EDC's of the spectra directly. The challenges of this approach in the cuprates are manifold. First, it is widely assumed that the "Lorentzian" peaks that make up the EDC's of a band such as this are asymmetric in energy, the asymmetry presumably deriving from many body effects not captured in the lowest order approximation of the lineshape being due to lifetime effects. Second, the form of the background, here assumed to be linear, derives theoretically in part from the incoherent part of the electron spectral function. To say it is poorly understood would be an understatement. Finally, because there is no theory of high temperature superconductivity there is no function known to us *a priori* with which to fit spectra such as these in which a gap is present. All these complications aside it did prove possible to fit EDC's of the spectrum with simple Lorentzians on a linear background provided we cut off the fit at the gap edge. Besides providing good fits it also avoided the complications associated with the gap profile and the Fermi function at zero energy. The one caveat with this procedure involves the identification of the very small gap in the spectrum acquired

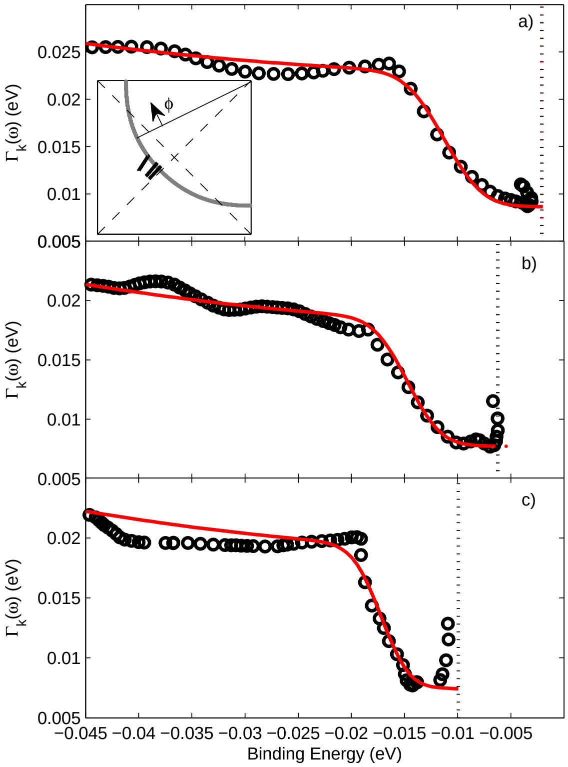


Figure 7.5: Imaginary part of the self energy as a function of binding energy extracted from the deconvolved spectra of Figure 7.1d-f. The scattering rate $\Gamma_k(\omega) = 2\,\mathrm{Im}\,\Sigma_k(\omega)$. The dotted vertical lines denote the gap position corrected for kinematic and resolution effects. Double valued points in the vicinity of the gap result from the parabolic dispersion along the gap edge upon crossing k_F.

closest to the node. Simulations such as those in Chapter 6 indicate that peak positions observed near the Fermi level are affected by energy resolution and experimental temperature broadening. The results of applying this procedure for acquiring the imaginary part of the self energy as a function of binding energy from the deconvolved spectra are plotted in Figure 7.5a-c.

Out of what appeared to be a rather complicated set of spectra we have been able to extract a shockingly simple structure in the very low energy scattering rate. Each plot of $\Gamma_k(\omega)$ can be broken down into three distinct regions. At the lowest binding energies we observed in all cases a sudden increase in the scattering rate upon approaching the gap edge, which is here defined as the point with the smallest binding energy, followed by a gradual *decrease* in the scattering rate and folding back of the dispersion. While at first glance this behavior appears odd it is nothing more than a manifestation in scattering rates and dispersions of the formation of Bogoliubov quasiparticles in the superconducting state. It is essentially no different from the coherence effects observed in the synchrotron experiment but here shown at a much higher level of detail concomitant with the increased momentum resolution imparted by the use of a very low energy photoexcitation source. The end of this "tail" is cut off where the coherence factor has suppressed the folded band intensity to the point it can no longer be reliably fit in the manner described above. While the physics of Cooper pairing is not the object of our study here the extent to which the back folded part of the Bogoliubov dispersion reaches in binding energy gives a good indication of how deep in energy the effects of pairing reach when considering a frequency dependent gap function $\Delta_k(\omega)$. The importance of this point will become clear momentarily.

The rest of the structure in $\Gamma_k(\omega)$ is divided neatly between the very low scattering rate observed immediately before the effects of the superconducting gap set in and the steep rise and subsequent saturation of the scattering rate at higher binding energy. It is this step-like profile in the imaginary part of the self energy that we are here most interested in because it is the classic signature of an electron-boson interaction. That this step in the scattering rate occurs where the kink is observed in the spectrum is a strong indication that the little kink does in fact reflect such a coupling. The most likely alternative scenario for the appearance of such features, that the kink and the step in the scattering rate are somehow related to their proximity to the superconducting gap, can be effectively ruled out by their considerable separation from the binding energies at which pairing effects clearly dominate. This separation of energy scales is clearly demarcated by the limited extent in binding energy of the Bogoliubov like dispersion pointed out above. In addition we did not observe such a feature emerging from the simulations of Chapter 6 in which we introduced a superconducting gap into our linear spectra.

Having established that the kink is most likely due to an electron-boson coupling we can turn our attention to the systematics of the effect. One of the foremost problems in ARPES on the cuprates has involved the examination of systematics e.g. with doping, temperature, momentum and energy of the other known kinks in the system. The kink at 70 meV in the nodal region, the "classic kink", for example exhibits only a weak temperature dependence[17], a strong doping dependence and evolves in energy and momentum in a rather complicated and not entirely clear way as one moves around the Fermi surface. The combination of these observations together with the abundance of available bosonic modes of the appropriate energy scale that may be responsible for its formation have made the nature of the classic kink a subject of intense debate for over a decade now.

The little kink on the other hand shows a very clear momentum dependence so far as we were able to measure it. It turns out that the binding energy of the center of the step feature in the scattering rate always appears 8 meV below the superconducting gap Δ. The momentum dependence of this feature is summarized in Figure 7.6 in which we plot Δ_k, the mode energy Ω_k and the difference between the two, $\delta_k = |\Omega_k - \Delta_k|$ as a function of Fermi surface angle ϕ. Note that the value of the gap for the cut shown in Figure 7.1c and f is corrected to reflect the value of the gap at the kink momentum rather than the gap momentum because of the large angle between that cut and the Fermi surface normal. The figure clearly shows how both the gap and mode energies evolve around the Fermi surface in the nodal region while their separation in energy remains a constant. Such a clear and unusual relationship between the gap and mode energies has profound implications for the type of boson that can be responsible for the little kink that we shall divulge in the following section.

7.4 The Identification of a Boson

When confronted with evidence for a new electron-boson coupling in any system the next step is clearly to attempt to identify the bosonic mode responsible. In Bi2212 the task is a bit more fraught than normal because of the abundance of both phonons and various sorts of magnetic excitations as well as the distinct possibility that more exotic phenomena may be relevant. This has certainly been the case for the classic kink as well as the recently identified "big kink" at 350 meV [19] which, depending whom you ask, is either a clear cut case of an electron-magnon interaction, a spurious matrix element effect or derivative of a more complicated band structure effect. Regardless, one must at least try. In this case the unusual nature of the mode's dispersion relative to the superconducting gap provides the decisive clue that has allowed

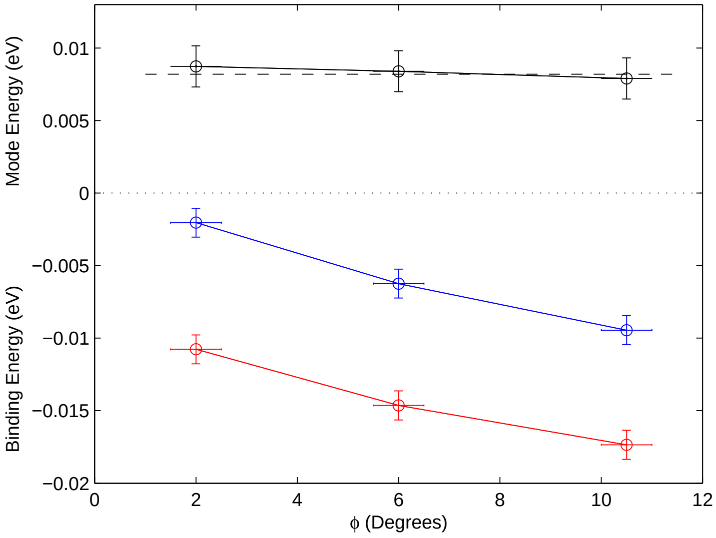


Figure 7.6: Important energy scales extracted from the self energies of Figure 7.5 as a function of Fermi surface angle ϕ. The red points are Ω_k, the blue points are Δ_k and the black points are $\delta_k = |\Omega_k - \Delta_k|$. Lines connecting points are a guide to the eye.

us to for the first time unambiguously assign a particular bosonic mode to an interaction in Bi2212 observed by ARPES.

As it turns out, the fact that δ_k is a constant, i.e. that the apparent mode energy is renormalized in the ARPES spectrum by the opening of a superconducting gap[48], implies a coupling of the low energy electrons in the system to a q 0 bosonic mode. Physically this phenomenon derives from the fact that a photohole injected into the system at a momentum k and binding energy ω greater than δ_k can, in the presence of a zone center mode, only be annihilated by an electron decaying in a near vertical transition preserving momentum. Because in the superconducting state all electrons above the gap energy are paired, the lowest energy electron that can transition in this manner will always originate from the vicinity of the gap edge, in this case 8 meV away, and so as one scans around the Fermi surface the mode energy will always appear at the same binding energy *relative* to the gap. In this picture the broadness of the step in the scattering rate at the mode energy derives from the broadness of the gap edge states, i.e. because the density of states at the gap edge is not a δ function in real materials, as well as the finite width of the phonon density of states and the residual energy resolution which is here about 5 meV.

In the energy range of interest here there is actually a relative paucity of zone center bosonic modes to be found in Bi2212. Certainly there is no magnetic mode of this form known to exist in Bi2212 at 8 meV. The most obvious q=0 bosonic excitations to be found in any crystal are optical phonons, a.k.a. Einstein modes. These observations narrow the range of possibilities considerably and make the identification of our boson a relatively straightforward process of examining the extensive literature already available concerning the measurement and assignment by model calculation of the optical phonon spectrum in Bi2212. The primary experimental techniques available for these measurements are IR spectroscopy and Raman spectroscopy which probe complementary and mutually exclusive sets of optical phonons that are for obvious reasons said to be either IR active or Raman active. An extensive listing of such measurements as well as some of the model calculations that went with them are given in the citations[4][49][50][51]. The result of this archaeological expedition was the identification of a strong A_g or A_{1g} Raman active mode [52]in Bi2212 clocking in between 58 cm^{-1} and 65 cm^{-1}; that is between 7.19 and 8.06 meV, respectively. (In the parlance of optics units of cm^{-1} are referred to as "wave numbers" denoted by the symbol $\overline{\nu} = 1/\lambda$. The conversion to units of energy meaningful in the context of ARPES is done by $E = hc\overline{\nu}$.)

Because these numbers resulted from experiments performed on a range of "optimally doped" Bi2212 samples of varying and perhaps unknown quality, at various temperatures and in various Raman scattering geometries, we found it efficacious to attempt our own measurement of the Raman active modes in optimally doped Bi2212. Our experiment was performed at room temperature on a high quality single crystal sample of optimally doped Bi2212 from the same rod growth as the sample used in our laser ARPES study. The sample was cleaved with Scotch tape just before the experiment to reveal a fresh, mirror perfect surface. The surface quality was checked by simultaneous *in situ* optical microscopy, revealing a featureless, flat cleaved surface. The experiment was performed in a Witek confocal Raman microspectrometer at the Brookhaven Center for Functional Nanomaterials. The spot size examined

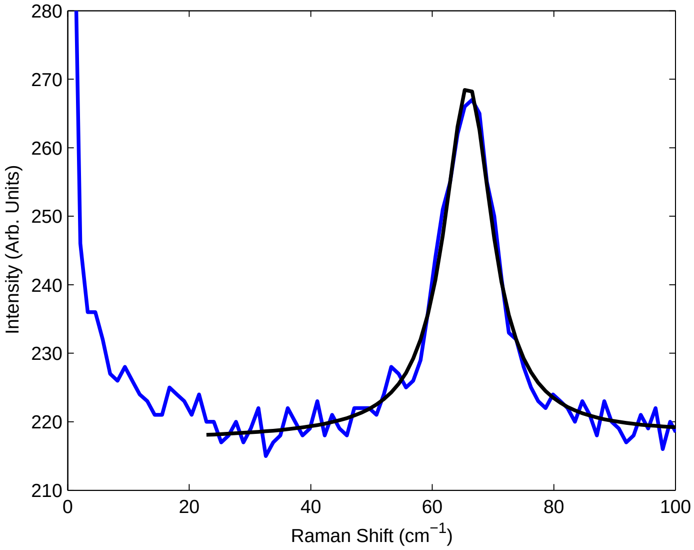


Figure 7.7: Confocal Raman spectrum of Bi2212 at room temperature using a 532 nm excitation source.

was less than one micron and a 532 nm laser excitation source was used. The result of this experiment is plotted in Figure 7.7. Because of the extremely high $f/\sharp$ of confocal Raman there was no well defined polarization vector in the experiment. This essentially results in almost all the Raman active phonons in the system being observed simultaneously, producing a relatively featureless high energy response in the Raman shift. The lower energy part of the spectrum, below about 150 cm^{-1}, is suppressed by an edge filter placed between the sample and the analyzing spectrometer so that only the strongest phonon line is observed. The Raman active phonon mode observed on our sample was centered at 66.5 cm^{-1} (8.25 meV) which is consistent both with the literature and with the mode energy Ω_0 observed in the ARPES data.

7.5 Self Consistency and the Coupling Constant

As we mentioned above the real and imaginary parts of the self energy ought to be causally related to each other via Kramers-Kronig relations. We have also pointed out that the real part of the self energy for Bi2212 is difficult to determine directly from experiment because it represents the change in the electronic band dispersion from that of the bare band induced by the interaction. If one can postulate a reasonable bare band from, say, a tight binding model then it is a trivial matter to measure the difference in energy between the observed and bare band, that difference being the real part of the self energy due to the electron-boson interaction. Unfortunately the bare band is a poorly defined concept in general and almost meaningless in Bi2212 because it is a doped Mott insulator. Also, as is evidenced by the deconvolved spectra, the change in the real part of the self energy induced by this coupling yields a much weaker signal than the change in the imaginary part.

Our identification of the culprit Boson as a particular optical phonon mode allows some rare insight into this problem. To first order we can represent the imaginary part of the self energy as a step function with $\Gamma(\omega < \Omega) = 0$ and $\Gamma(\omega > \Omega) = \gamma$ where γ is the energy step measured in ARPES. In this case the Migdal equation[53]

$$\Gamma(\omega) = 2\pi \int_0^\omega \alpha^2 F(\omega')[2n(\omega') + f(\omega' + \omega) + f(\omega' - \omega)]d\omega', \tag{7.2}$$

where $n(\omega')$ and $f(\omega' \pm \omega)$ are the Bose-Einstein and Fermi-Dirac distributions, respectively, can be easily inverted to find that $\alpha^2F(\omega) = A\delta(\omega - \Omega)$, a delta function at the mode energy. Thus,

$$\Gamma(\omega) = 2\pi A\theta(\omega - \Omega) \tag{7.3}$$

and

$$A = \frac{\gamma}{2\pi} \tag{7.4}$$

where γ is the energy step in $\operatorname{Im}\Sigma$ ignoring other contributions, e.g. impurity scattering. With this result in hand it is a trivial matter to determine the mass enhancement factor λ,

$$\lambda = 2\int_0^\infty \frac{\alpha^2F(\omega')}{\omega'}d\omega' = \frac{2A}{\Omega} = \frac{\gamma}{\pi\Omega}. \tag{7.5}$$

where γ is the measured step energy $(E_{int} -' E'_{bare})$. For example, with $\gamma = .013$ eV and $\Omega = .008$ eV, consistent with the above data, we find $\lambda = .6$ putting this interaction squarely into the weak coupling regime. There exists a simple analytic form of the real part of the self energy in the event the imaginary part $\Gamma/2$ described above has the step function form[54]. That is,

$$\operatorname{Re}\Sigma(\omega) = \frac{\lambda\Omega}{2}\ln\frac{\Omega - \omega}{\Omega + \omega}. \tag{7.6}$$

The relationship between $\alpha^2F(\omega)$, $2\operatorname{Im}\Sigma(\omega)$ and $\operatorname{Re}\Sigma(\omega)$ is shown in Figure 7.8. The singularity in the real part of the self energy evident in Eq. 7.6 at the mode energy is screened by both the finite temperature and finite mode width and dispersion present in real materials. This problem does not arise in

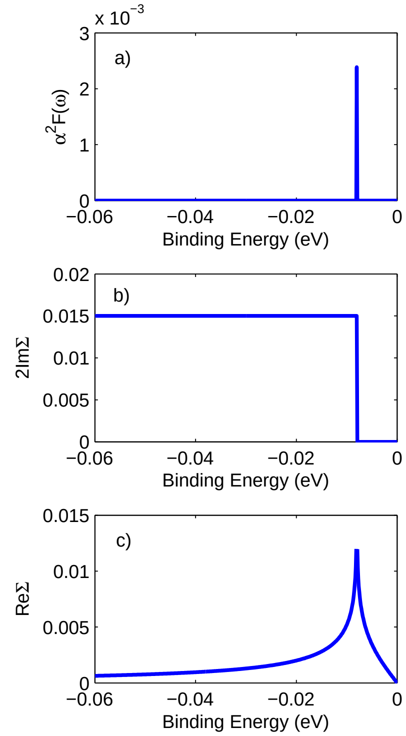


Figure 7.8: Analytical approximation of electron-phonon coupling to an Einstein mode; a) $\alpha^2F(\omega)$ as a delta function at Ω from Eq. 7.4, b) $2\,\mathrm{Im}\,\Sigma(\omega)$ calculated from Eq. 7.5, c) $\mathrm{Re}\,\Sigma(\omega)$ calculated from Eq. 7.6

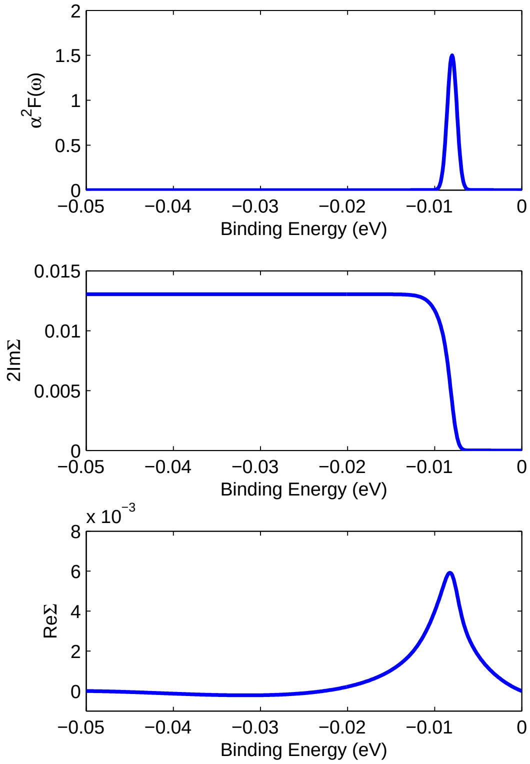


Figure 7.9: Numerical calculation of electron-phonon coupling to an Einstein mode; a) $\alpha^2 F(\omega)$ with $\alpha^2 = 1.5$ and a finite width as derived from the Raman spectrum of Figure 7.7, b) $2\,\mathrm{Im}\,\Sigma(\omega)$ calculated from Eq. 7.2 using the function in panel a), c) $\mathrm{Re}\,\Sigma(\omega)$ calculated from Eq. 7.7 with step cutoff at 70 meV.

the fully numerical calculation presented below.

We can improve our estimate of λ and $\mathrm{Re}\,\Sigma$ in this case by the following procedure. While it is in general quite difficult to invert the ARPES scattering rate to get $\alpha^2 F(\omega)$ in the event a complicated *set* of relevant phonons or other bosons is present, in this case it turns out to be relatively straightforward to do so by hand. Using the Raman data of Figure 7.7 we conclude that this phonon mode has a FWHM of 10 cm^{-1} (1 meV). We therefore construct an Eliashberg function with $F(\omega)$ a Gaussian of FWHM 1 meV centered at $\Omega = 8$ meV. Solving Eq. 7.2 we find that $\alpha^2 \cong 1.5$ comes closest to capturing the observed $2\,\mathrm{Im}\,\Sigma$ nearest the node; the model yields a broadened step in the scattering rate with $\Gamma(\omega)$ reaching a saturation value of $\Gamma = 13.5$ meV. Solving Eq. 7.5 yields a coupling constant (at T=0) of .52, close to that of the ideal Einstein mode but softened somewhat by the nonsingular nature of the physical interaction. The real part of the self energy is determined numerically by the Kramers-Kronig transform

$$\mathrm{Re}\,\Sigma(\omega) = \wp \frac{2}{\pi} \int_0^\infty \frac{\omega' \,\mathrm{Im}\,\Sigma(\omega')}{\omega^2 - \omega'^2 + \imath\epsilon} d\omega' \tag{7.7}$$

where the $+\imath\epsilon$ prescription (taking the + at positive frequencies) has been used to shift the pole at $\omega = \omega'$ an infinitesimal distance onto the complex plane and $\wp$ denotes that the principal part of the integral is to be taken. While the real part of the self energy calculated with Eq. 7.7 depends upon the manner and energy in which the integral over the imaginary part is assumed to be cut off, the overall scale of the renormalization is affected only weakly by these choices. In this case we cut off the mode at 70 meV – the energy of the "classic" kink – because the effects of the interaction become minimal at this point. In Figure 7.9 we plot the $\alpha^2 F(\omega)$, $2\,\mathrm{Im}\,\Sigma(\omega)$ and $\mathrm{Re}\,\Sigma(\omega)$ for the above parameters for comparison with Figure 7.8.

The result of this exercise shows that the qualitative effect of introducing a finite temperature and mode width into the problem are a slight broadening of the step in the scattering rate and an accompanying softening on the peak in the real part of the self energy. Further, the singularity in the real part of the self energy that was present in the idealized case has been removed without artifice. The major difference between the idealized and full calculations lies in the behavior of the real part of the self energy for $|\omega| > |\Omega|$. While for the simplified Einstein case the real part is always positive it tends to zero at high energy for the full calculation and, depending on the manner and energy at which the integral in Eq. 7.7 is cut off and the strength of the interaction, can even take on negative values for some range of energies past the mode.

To make a realistic comparison of the theory to the experimental results it is necessary to make a few phenomenological additions to the imaginary part of the self energy. Recall that the full scattering rate is given by a sum of terms

$$\Gamma_{tot} = \Gamma_{e-e} + \Gamma_{imp} + \Gamma_{e-ph} \tag{7.8}$$

where Γ_{e-e} accounts for scattering due to the electron-electron interaction, Γ_{e-ph} is the electron-phonon term calculated above and Γ_{imp} is a constant impurity scattering term on the order of a few meV. Γ_{imp} is supplied from the data by the narrowest value observed near the Fermi level in a given spectrum and is added by hand. Γ_{e-e} is not well understood in the cuprates. In the normal state it is taken to be of the marginal Fermi liquid form in which $\Gamma \propto max(\omega; T)$, i.e. to be linear in the binding energy or temperature, whichever scale is greater. On the other hand the scattering rate in the superconducting state is taken to be either cubic in ω or of the quadratic Fermi liquid form. In practice it is almost impossible to tell the difference between these forms over the range of energies that are sampled here. Because we observe that most points in $2\,\mathrm{Im}\,\Sigma$ are technically Landau quasiparticles, meaning they have a lifetime less than or equal to their binding energy, we find it expedient to model Γ_{e-e} up to the binding energy examined on the Fermi liquid form;

$$\Gamma_{e-e} = 2\beta((\pi k_B T)^2 + \omega^2) \tag{7.9}$$

where $2\beta = \pi U^2/2W^3$, U being the on sight Coulomb energy and W the bandwidth, which is also the electron kinetic energy. Following Comanac *et al* [55] we use $U = .9 \times 4.4$ eV and $W = 2.5$ eV so that 2β 1.58. This parametrization of the electron-electron term has the benefit in this case of being independent of the present measurement. Finally, we shift the step position by $\Omega_0 + \Delta_k$ and broaden it by the residual energy resolution of 5 meV. No attempt is made to account for the presence of the superconducting gap, which does not affect the kink here, or the presence of the kink at 70 meV which if due to an electron-boson coupling would affect the dispersion over the whole energy range here examined. The impurity term aside the only parameter in the fit is now α^2. The results of the model calculations are lain over the data in Figure 7.5. We found the data to be well modeled by choosing α^2's of 1.5, 1.1 and 1.2, respectively, for the total scattering rates plotted in the figure. Because an α^2 of 1.1 yields $\lambda = .38$ for a mode at 8 meV the data suggests the coupling strength diminishes as one moves away from the node though clearly a more systematic study of this effect would be desirable. In any event, the value of λ closest to the node is in very good agreement with the mass enhancement due to electron-phonon interactions calculated from first principles (for LSCO) in reference [56].

As a final check on the validity of our approach to modeling the data we

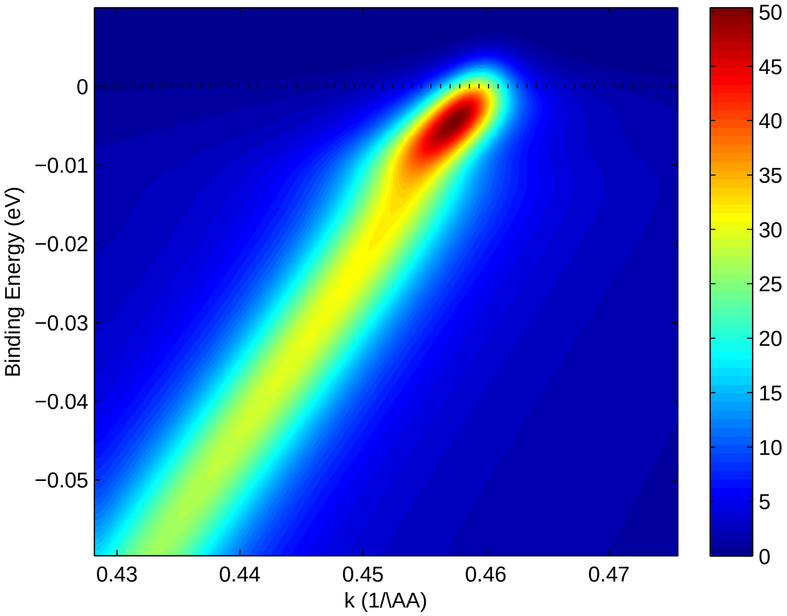


Figure 7.10: Simulation of nodal BSCCO ARPES spectrum at 10 K using the experimentally derived Einstein phonon spectrum.

plot a simulation similar in spirit to those of Chapter 6 using a linear band at the node. We take the bare band velocity to be 2.2 eVÅ, the scattering rate to be the unbroadened calculation for Γ_{tot} with $\alpha^2 = 1.25$, $\Omega_0 = 8$ meV, $\Gamma_{imp} = 8$ meV, the real part of the self energy calculated from Eq. 7.7 and then broaden it by the experimental resolution. As before we model the EDC line shape as a simple Lorentzian. The result, representative of what one expects at the node, is shown in Figure 7.10. The result bears a reassuring resemblance to the data. Qualitatively the most pronounced aspect of the model spectrum is the manner in which the little kink manifests itself much more strongly in the scattering rate than it does in the real part of the self energy. This behavior provides a good explanation for why even if we were to postulate a bare band dispersion and attempt to extract the real part of the self energy directly from the data an accurate accounting would be extremely difficult. This difficulty is enhanced in the real system by the 70 meV kink, the presence of which was not included in our model. Because the change in the real part of the self energy associated with the 70 meV dispersion anomaly must set in and increase rapidly at much lower binding energies what we observe as a very slight kink in our model will tend to appear as a "bend" in the dispersion thus making it even more difficult to observe directly than our simple model here suggests.

7.6 Conclusion of the Phonon Story

In this section we have presented ARPES data showing evidence of a small electron mass renormalization whose binding energy scales with the superconducting gap at low temperature such that $\omega_k - \Delta_k \simeq 8$ meV. This scaling is best explained by the mass enhancement originating from an electron-phonon coupling involving a q=0 optical phonon mode of A_{1g} symmetry as deduced from model calculations in the literature. There are has been previous theoretical work suggesting such a coupling is possible in the Bi based cuprates [57]. We were able to correlate this energy scale in the ARPES data to a particular Raman active mode observed previously in the literature as well as in our own Raman measurement. Further, by combining the behavior of the imaginary part of the self energy observed in ARPES with the measured phonon spectrum for the culprit mode we were able to deduce the Eliashberg function $\alpha^2F(\omega)$, the mass enhancement factor λ and subsequently provide an estimate for the self consistent real part of the self energy. Simulation of the nodal ARPES spectrum with these parameters as well as some additional phenomenological and theoretical input produced results in good agreement with the measurements.

The correlation of any bosonic mode with a coupling observed by ARPES is always circumstantial because photoemission by itself can only observe the single electron density of states. In Bi2212 such correlations have been highly controversial because at the energy scales at which they are observed there generally exists a plethora of candidate modes, any of which alone or in concert with others may yield such mass enhancements. In this case on the other hand the identification is relatively clear due to the very particular form of the modal dispersion and because of the small number of modes in the correct energy range and of the correct momentum. We are thus led for the first time to be able to unambiguously identify an interaction in Bi2212.

While this phenomenon is itself unlikely to be directly related to high temperature superconductivity in Bi2212 it does give us some insight into the more exotic physics observed in this system. First, the observation that there exists a “universal nodal Fermi velocity” in the cuprates [58] is certainly incorrect. Because the little kink involves the relative oscillations of the two copper oxide planes in the bilayer Bi2212 system we do not expect it to appear in the single layer compounds, e.g. Bi2201, and they will thus have a different nodal Fermi velocity. Looking to other bilayer systems it is clear that if such a coupling is present it will appear at a different energy, if at all, because the low energy of the mode in Bi2212 is tied to the great mass of the out of plane Bismuth atom in the unit cell. We thus conclude not surprisingly that the nodal Fermi velocity should be similar in all cuprates because it originates from nearly identical copper oxide planes but that the details of any given system might yield very different velocities at E_F in the superconducting state depending on whether or not such low energy couplings in the system are present.

More importantly this finding will necessarily deepen the story behind the famous 70 meV kink first observed in Bi2212 and later in many other cuprates. The observation that the band structure of Bi2212 is affected by a renormalization that sets in prior to the 70 meV kink implies that the coupling strength associated with this interaction as derived from any measure of the real part of the self energy will have to be revised downward. This revision will in turn affect the identification of the mode or modes, if any, responsible for that interaction. It is hoped this observation will make that problem, still unresolved after more than a decade, more tractable.

Finally, this coupling can provide some insight into the unique nature of the quasi 2D electronic states that reside in the copper oxygen planes of Bi2212. In Figure 7.11 we diagram the relative motions of the atoms in the Bi2212 unit cell provided by calculations of the A_{1g} phonon eigenmodes in reference [49]. The 8 meV Einstein mode, as well as those of similar energy, is dominated by the c axis motion of the Bismuth atoms as well as the collective motion of the copper ions in the planes, apical oxygen atoms (which in the figure reside above or below the planar Copper atoms) and the Strontium atoms. This calculation provides an intuitively meaningful physical picture for why this c axis mode should couple to electrons residing in the copper oxygen planes with an essentially 2D character.

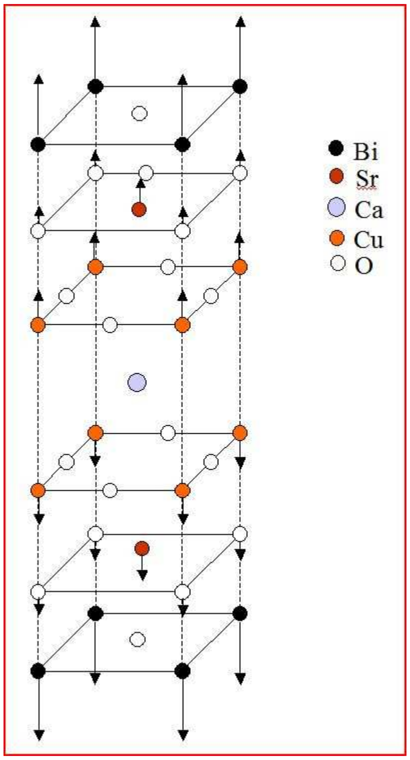


Figure 7.11: Schematic of the Bi2212 primitive unit cell atomic motions associated with the 8 meV A_{1g} optical phonon mode. The arrow lengths depict the relative magnitudes of displacement suffered by various atoms. Note that in the Bi2212 primitive unit cell the out of plane apical oxygens are located directly above (below) the copper atoms of the upper (lower) copper-oxygen planes. This figure is adapted from Reference [4].

Evidently there are several factors at work here. Determining the relative importance of these mechanisms is beyond the scope of this work. For now it suffices to point out that firstly this mode involves the motion of the planar copper atoms relative to stationary planar oxygen atoms. This relative motion will effectively alter how the Cu ions screen electrons in the planar $pd\sigma$ orbitals. Secondly, this is the only A_{1g} mode in Bi2212 except for one observed, curiously, at 77 meV that involves motion of the apical oxygen atoms. The possible relevance of this fact stems from the observation that the apical oxygen atoms of all the cuprates are thought to play an important role in mediating the next-nearest neighbor oxygen-oxygen hopping energy t' in the plane. Because the nearest neighbor hopping term t is strongly suppressed in the cuprates by correlations between nearest neighbors the t' term plays an outsized role in the planar conduction. Thus a phonon involving the apical oxygens may well alter the electron-ion screening in the plane sufficiently to induce the mass renormalization observed by ARPES. Finally, the movement of the copper atoms produces in this mode a relative motion of the two Copper Oxygen planes that can be pictured as two hands clapping. Because the nature of interlayer conductivity in the cuprates remains poorly understood we can only speculate that this will have some effect on the planar electrons though it does suggest that one should not find a similar interaction arising in the single layer cuprates. Confirmation or refutation of any or all these ideas will have to await future experimental and theoretical examination. Finally, because the temperature scale θ_E associated with this mode is coincidentally very close to T_c ($\Omega/k_B = 92.8$ K while $T_c = 91$ K) we expect this feature to be unobservable above T_c. That effect is not profound but rather due to the loss of coherence and subsequent broadening of states above T_c, making this mode, at such low energy, totally unobservable in the scattering rates. Further, the broadening of the Fermi distribution as T is raised will, in analogy to a resolution broadening, alter the dispersion of the very low energy electronic states in such a way as to cause the change in the real part of the self energy to become unobservable.

Chapter 8

Unoccupied Electronic States of Underdoped and Optimally Doped Bi2212

8.1 Methodology and Expectations

The experiment examined here was based upon the systematic study of low lying thermally populated excited states in optimally and underdoped Bi2212. While many justifications and reasons might be given for undertaking such an exercise perhaps the most important at the time was that the unoccupied states of Bi2212 represented almost entirely unexplored territory for photoemission on the cuprates. Though difficult, such an experiment opened new and ultimately unforseen possibilities in the study of the electronic structure of these materials since by this time the occupied electronic structure had seemingly been well characterized and did not appear to offer much in the way of new insight into high temperature superconductivity. Ultimately this study led not only to the discovery of new physics in these materials but also out of necessity to the development of the LRM for applications in ARPES.

The methodology of our study has already been laid out at some length in Chapter 6, namely that of observing unoccupied states at "elevated" temperatures by normalizing LRM resolution corrected spectra to the Fermi-Dirac distribution $f(\omega, T)$. The result was called a Fermi normalized spectrum or FNS and revealed the nature of unoccupied states up to $4k_BT$ above the Fermi level. As alluded to earlier, the LRM correction was devised in response to the observation that other methods of normalizing ARPES spectra acquired with an energy resolution even modestly greater than the natural widths of the intrinsic features of a spectrum often yielded unphysical results. Additionally, because the signal to noise ratio (SNR) of an ARPES spectrum is approximately Gaussian in the number of counts detected at the analyzer for a given channel on the microchannel plate (MCP), and because the number of counts above the Fermi level relative to the occupied side of a spectrum decays almost exponentially with increasing energy, the integration time needed to get a reliable picture of the unoccupied states was an order of magnitude greater than what is typical. For this study, each spectrum required between six and twelve hours of continuous data accumulation. The acquisition of enough spectra for an FNS study of a given sample at a time when our group was rather small in number thus required us to work alternating twelve hour shifts over the course of several days. As a result this time was also sufficient to complete the reading of an entire Thomas Pynchon novel dealing largely with the symbolic and metaphorical ramifications of birefringence and wave-particle duality.

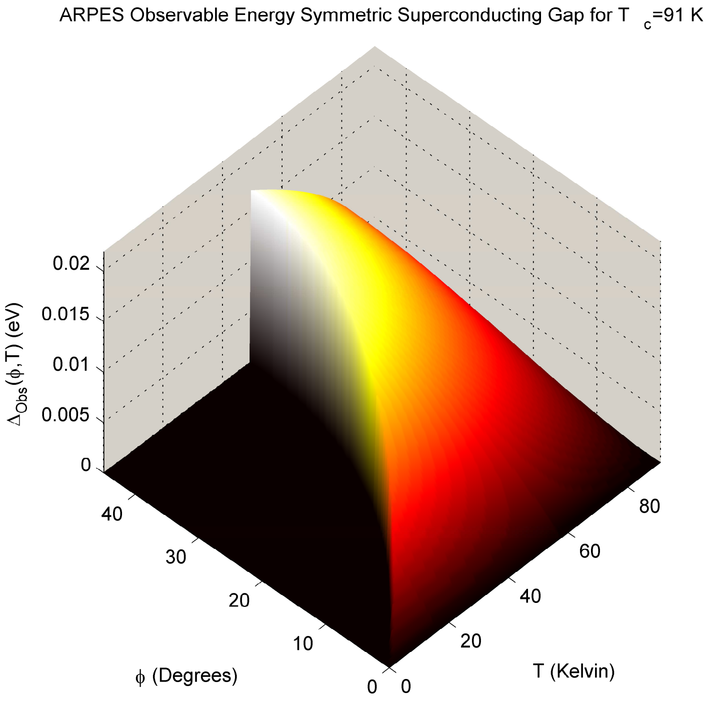


Figure 8.1: The d-wave weak coupling BCS gap Δ as a function of temperature T and Fermi surface angle ϕ for a superconductor with $T_c = 91$ K. Regions of zero height on the z axis denote Fermi surface angles for which unoccupied Bogoliubov quasiparticles are unobservable by the thermal population technique. As one raises the temperature from absolute zero unoccupied states of the d-wave gap further from the node can be observed. For optimal doping one can observe low lying unoccupied states across the whole Brillouin zone for $T \gtrsim 50$ K.

In our examination of the superconducting state the original object was to observe the presence of Bogoliubov quasiparticles on both the occupied and unoccupied side of the Fermi level simultaneously. The form these spectra were assumed to take has already been reviewed in Chapter 6 and at the time we were aware of only one other ARPES experiment in which this issue had been explored[59] though a similar study appeared in the literature concurrent with our own experiments [60]. As a rough guide of where in the Brillouin zone to look for an energy symmetric superconducting gap, at least for optimal and overdoped cuprates, it is useful to plot the gap as a function of both Fermi surface angle ϕ (defined in Chapter 2) and T, keeping only those values of the gap at or below the cutoff temperature for observability, $E_{Obs} = 4k_BT$. This is done in Figure 8.1 using the d-wave form for the angular dependence of the gap $\Delta(\phi, T) = \Delta(T)\sin(2\phi)$ with the approximate BCS expression for the gap magnitude as a function of temperature $\Delta(T) = 3.2k_BT_c\sqrt{1 - T/T_c}$. Though only approximate and somewhat optimistic because of the intervention of the pseudogap phenomenology it is close enough for optimally doped samples to guide an investigation of this sort in the nodal region of the Brillouin zone.

In addition to the observation of the Bogoliubov quasiparticle dispersion expected to be associated with the opening of the superconducting gap it was also decided that a few other features merited attention. As a check on the efficacy of our methodology for generating accurate FNS an examination of the nodal state was required. That is because the node should exhibit no energy gap either above or below T_c, regardless of doping, and so should cross the Fermi level in a manner very similar to that of the model linear band examined in Chapter 6. Similarly, it was expected that an examination of the unoccupied part of the electron spectral function in the normal pseudogap state in the vicinity of the node, i.e. along the so-called Fermi arc, should yield similar results. As shall be seen below we were pleasantly surprised to find that this is most definitely *not* the case for the underdoped samples that were examined. Finally, because the pseudogap state remains one of the most baffling phenomena in solid state physics, it was of paramount importance to examine the unknown status of its unoccupied side using FNS in the normal state as well as its evolution below T_c. This experiment was motivated in part by the time resolved optical reflectance mentioned above as well as various measurements showing the existence of a strong diamagnetic response for underdoped cuprates well above T_c, but below T^*, indicative of a strong phase incoherent pairing persisting well into the normal state, i.e. pre-formed pairs. Taken altogether it was hoped that these measurements could shed some light on the long known particle-hole asymmetry in cuprate tunneling specta indicating a much larger density of electronic states below the chemical potential than above it.

8.2 Result of the Examination of Unoccupied States in Optimally and Under Doped Bi2212

Our study focused on two dopings of Bi2212; the optimally doped samples with a T_c of 91 K and doping fraction of .16 and heavily underdoped samples with a T_c of 65 K and a doping between .12 to .13. T_c was measured for all samples by the bulk magnetic susceptibility technique in a Magnetic Properties

Measurement System (MPMS) using SQUID magnetometry. All samples were fabricated by the traveling solvent floating zone technique known to yield the highest quality of single crystal cuprates. Underdoping of samples was achieved by annealing in a vacuum at 550 C for 48 hours. Samples were oriented in a Laue diffractometer prior to mounting. All measurements were carried out on beamline U13UB with a Scienta SES-2002 photoelectron spectrometer. The overall energy and angular resolution of the spectrometer/beamline system was 15 meV and .1 degrees, respectively. The energy resolution as well as the position of the Fermi level were calibrated on a bulk sample of gold in electrical contact with the Bi2212 samples. Samples were cleaved *in situ* by the top post method with a vacuum chamber base pressure of 5×10^{-11} Torr. All measurements were carried out within 24 to 48 hours of cleaving a given crystal.

Though optimally doped Bi2212 is well known to have a small pseudogap with a T^* of about 140 K it essentially served as the control for the experiment because the sample quality is the most consistent of any doping of Bi2212 and because the high T_c and relatively sharp, bright states, especially below T_c, make it easier to observe the details of low lying thermal excitations. It is also a useful check because it is by far the best studied doping of the Bi2212 family, at least from the point of view of ARPES. Figure 8.2 shows a typical result for both raw and Fermi normalized spectra taken on optimally doped Bi2212 at 80 K in the nodal region. The location of the cut in k space is shown in the inset of panel a. A “bump” in the spectrum above E_F is clearly visible in panel a, as well as in the corresponding EDC stack plot in panel c. Panel b shows the FNS generated by normalizing the spectrum in panel a to the experimental

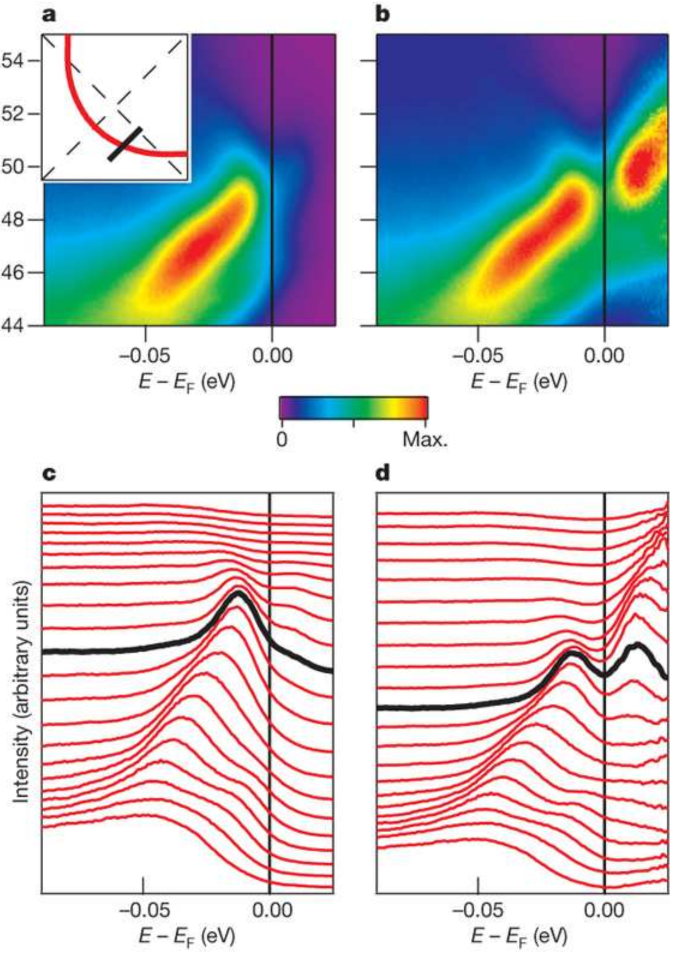


Figure 8.2: Spectra from optimally doped Bi2212 below T$_c$ treated by the LRM a) before and b) after Fermi normalization. Panels c) and d) show EDC's corresponding to panels a) and b), respectively. The location of the cut through the Brillouin zone is shown in the inset of panel a).

temperature Fermi distribution. The spectrum in panel a, from which panel b was derived, has already been corrected for resolution effects using the LRM. The overall result depicted in panel b and in the corresponding EDC stack plot in panel d shows a remarkably conventional BCS-like Bogoliubov quasiparticle dispersion. The peak-to-peak energy gap 2Δ is symmetrical about the chemical potential and the coherent weights of the occupied and unoccupied bands evolve as a function of momentum in accordance with the coherence factors u_k^2 and v_k^2 of a simple weak coupling pairing model. Because superconductivity in even the optimally doped cuprates is not believed to originate from the BCS mechanism such behavior for the Fermion spectrum in the presence of a superconducting gap should be interpreted as a generic result of the formation of a phase coherent Cooper pair condensate. In this sense, the superconducting gap in optimally doped Bi2212 in the nodal region appears to be eminently conventional and in striking accord with the results of the simple simulations shown in Chapter 6.

An examination of the FNS for a $T_c = 65$K underdoped sample in the superconducting state shows similar results. Spectra acquired at the node for both the OP 91 and UD 65 samples below T_c, Figure 8.3a and c, show the unbroken persistence of the single particle spectrum above E_F. Away from the node, both dopings display a particle-hole symmetric superconducting gap as shown in panels b and d of the same figure. It should be noted that the spectrum shown in panel d of Figure 8.3 was acquired at a point along the Fermi surface that is off of the Fermi arc at T_c, or in other words at a point in k space that is already pseudogapped before the sample goes superconducting. This highly nontrivial fact will be explored more fully below though for the

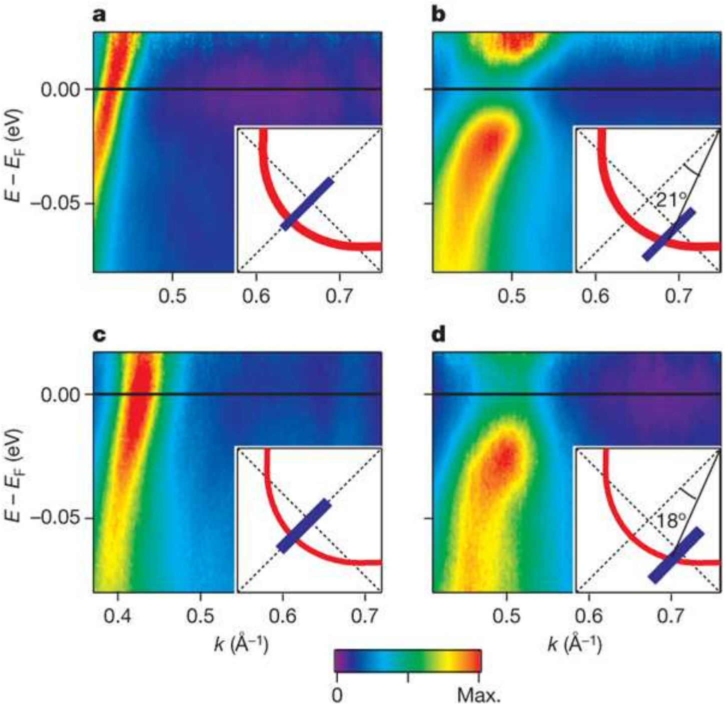


Figure 8.3: FNS of optimally doped (panels a) and b)) and T$_c$= 65 K underdoped (panels c) and d)) Bi2212 below T$_c$. Panels a) and c) show spectra acquired at the nodes. Panels b) and d) show spectra acquired away from the nodes at Fermi surface angles ϕ shown in the insets.

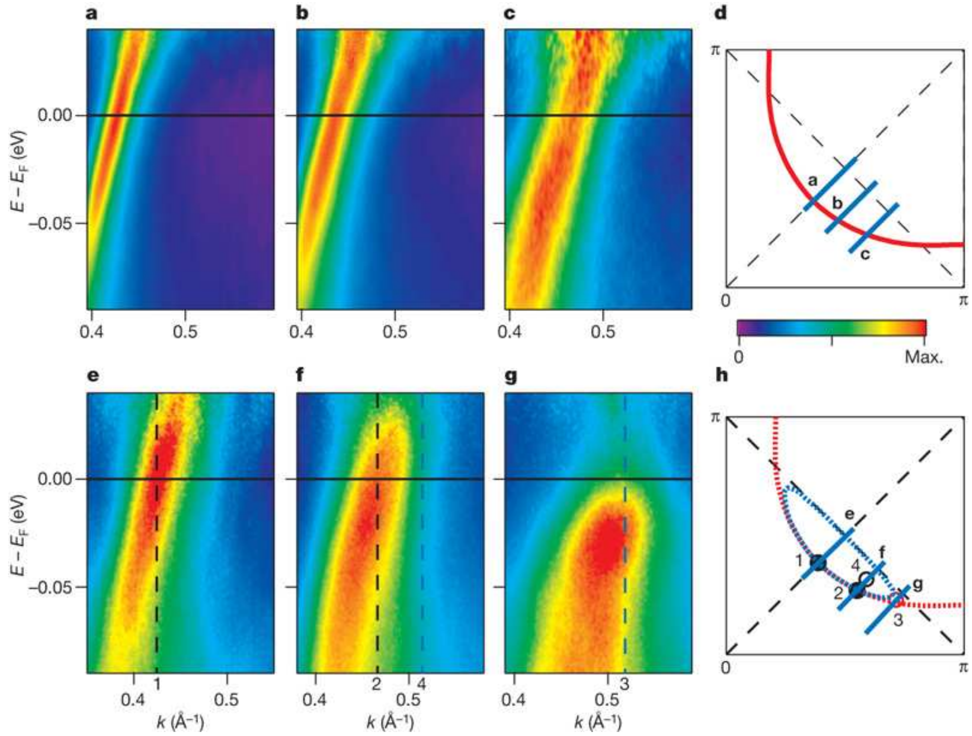


Figure 8.4: Normal state FNS for optimally doped (panels a-c) and underdoped (panels e-g) Bi2212 in the normal pseudogap state at T = 140 K. Panels d and h show the cut locations in k space for the optimal and underdoped samples, respectively. Panel h shows a small, closed Fermi pocket consistent with the underdoped FNS data. The area of this pocket is consistent with the nominal doping level of the underdoped sample, about .12. The underdoped sample displays a striking electron-hole asymmetry in the normal state.

moment it suffices to show that at least in the superconducting state the single particle spectral functions of both samples appear to be very ordinary for a d-wave superconductor. This behavior is consistent with another similar study of FNS in the superconducting state of a $T_c = 91$ K "underdoped" Bi2212 sample [60].

The first major surprise of this experiment occurred when we examined the normal state FNS of the underdoped sample for points in the Brillouin zone along and just off of the Fermi arc. Normal state FNS at 140 K for the optimally doped sample are plotted in Figure 8.4, panels a-c, and for the underdoped sample in panels e-g. The locations of the cuts in the Brillouin zone are shown in panels d and h for the optimally and underdoped samples, respectively. While the optimally doped sample shows rather conventional metallic behavior along its Fermi arc, i.e. displaying bands crossing the Fermi level unperturbed into unoccupied territory, the same is decisively not true for the underdoped sample. Upon moving away from the node, the spectral function acquires a clear particle-hole *asymmetry*. The suppression of spectral weight above E_F, with the top of the band moving to progressively lower energy as one moves away from the node, is the hallmark of single particle gap opening and centered *above* the Fermi level. That this gap is not particle-hole symmetric even when the top of the band has moved below the Fermi level (panel g) indicates that this gap cannot be associated with the formation of Cooper pairs above T_c. Further, the phenomenological description of the Fermi arc as comprising the set of single particle states in the vicinity of the node that are not pseudogapped in the normal state clearly requires some modification to reflect the fact that these states are not defined by what is left behind after the antinodal pseudogap is opened but rather by how this new, high energy gap disperses in momentum space. A full discussion of whether and how the "classic" antinodal pseudogap and this apparently new energy gap are related shall be deferred until the remainder of the data has been examined.

The data presented thus far give no indication of possible agreement with the observation of a strong diamagnetism or energy symmetric gap indicated by normal state Nernst effect or tunneling spectroscopy measurements. Be-

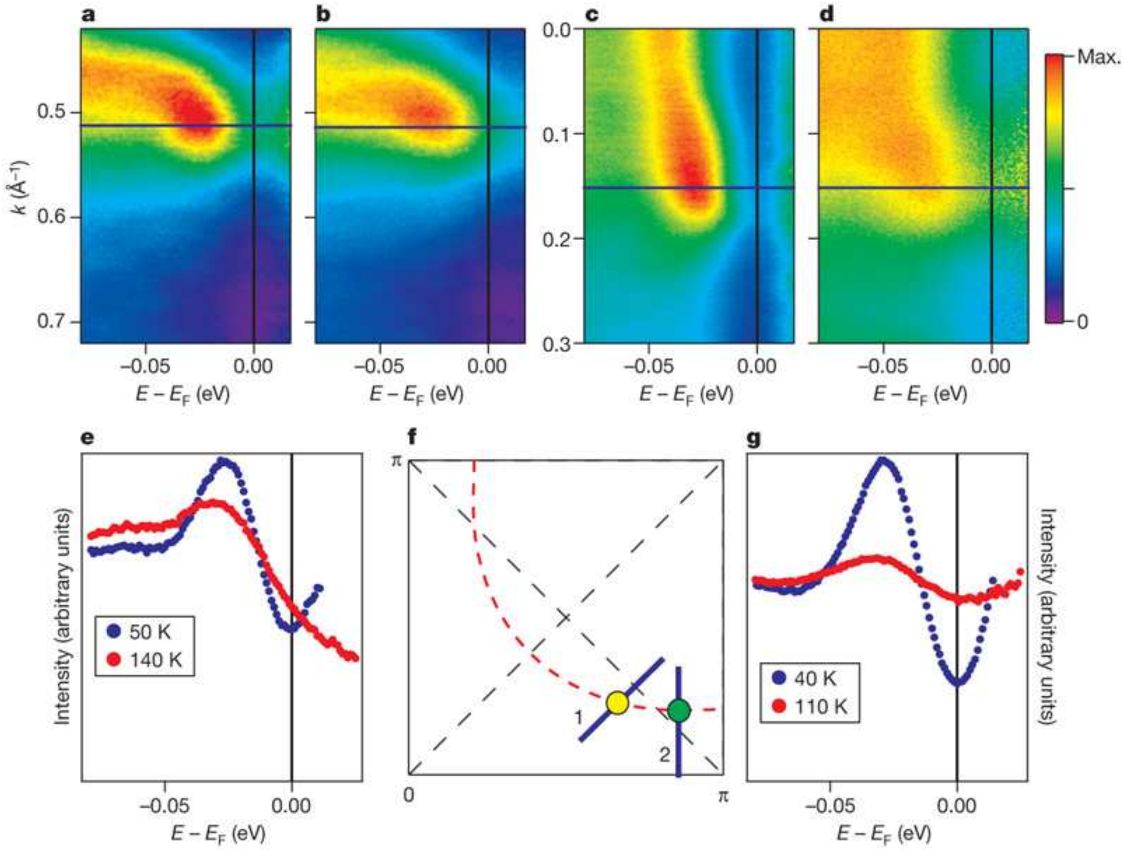


Figure 8.5: FNS from the T_c = 65 K samples a) at 50 K, b,d) at 140 K and c) 40 K. The cut locations are diagrammed in panel f. Panel e) EDC's at k_F from panels a and b just off the Fermi arc. Panel g) EDC's at k_F from panels c and d near the antinode. Nodal states display electron-hole asymmetry in the normal state with symmetry restored in the superconducting state. Antinodal states display electron-hole symmetry in both the normal pseudogap and superconducting states.

cause it is impossible for "pairing" in the particle-particle channel observed in the nodal region to generate such a signature, which has thus far been taken as evidence of the existence of so-called pre-formed Cooper pairs, one must seek remedy elsewhere. In searching for a spectroscopic signature of this phenomenon in ARPES it became evident that if only by process of elimination it was necessary to examine the unoccupied states much closer to $(\pi, 0)$ where the classic pseudogap is strongest. The result of this part of the experiment is shown in Figure 8.5. Panels a and b of Figure 8.5 show the FNS below and above T_c, respectively, for the underdoped sample a short way past the Fermi arc but still well away from the antinode. The particle-hole asymmetry of the gap in this region is is highlighted by the EDC's taken at the approximate Fermi momentum k_F. Panel c of the same figure shows an FNS acquired near $(\pi, 0)$ in the superconducting state with an energy symmetric, zero energy centered gap (panel g) as one expects for a superconductor. Extraordinarily, the FNS at the same point by the antinode *far above* T_c, panel d in the figure with accompanying EDC in panel g, also shows a particle-hole symmetric gap with the gap minimum centered at E_F. The locations of these cuts through the zone are shown in panel f.

8.3 Interpreting the Results

Because of the difficulty in acquiring this data and because of the wealth of qualitatively new information contained therein, it was very difficult to form a complete picture of what is going on here. Nevertheless a few astute observations about the data set and some support from complementary experimental techniques and theory offer a way in. In assimilating all this information it is still easiest to examine various regions of the Brillouin zone separately and to only then attempt to draw an overall conclusion.

8.3.1 The Nodal Region

In discussing the nodal region around which particle-hole asymmetry is observed it is useful to examine some explanations for the Fermi arc phenomenology that existed prior to this experiment. The Fermiology of the cuprates has long been studied using the so-called symmetrization technique rather than the Fermi normalization technique used in this study[36]. Symmetrization takes advantage of the fact that the occupied part of the ARPES spectrum goes like $A(k,\omega)^{full}f(\omega)$ and the unoccupied part goes like $A(k,\omega)^{empty}(1-f(\omega))$. If one assumes that the occupied and unoccupied parts of the spectrum are essentially identical, as was essentially the case with all of our optimally doped data, then one can approximate the full low energy spectral function by saying $A^{full} \approx A^{empty}$ so that $A(k,\omega)f(\omega)+A(k,\omega)(1-f(\omega)) = A(k,\omega)$, effectively canceling the contribution to the spectrum of the Fermi distribution. In practice this amounts to reflecting the spectrum across E_F, and possibly about k_F as well, and summing so as to acquire a full picture at both positive and negative binding energies. While a reasonable thing to do in some circumstances and clearly an easier method than generating FNS, it will obviously fail in the presence of particle-hole asymmetric states like those observed here in underdoped Bi2212. While the physical assumptions behind this procedure are now known to be grossly inadequate for studying the cuprates it nevertheless provides at least a convenient method for observing whether and where a gap has opened up in the density of states around the Fermi level.

In a world in which the Fermi arc at all dopings is assumed to look like what was seen here in only the optimally doped system it is difficult to come up with a plausible physical explanation for the lack of a Fermi surface in the normal state at suboptimal dopings. In light of the non-ARPES measurements cited above that indicated the presence of preformed pairs in the normal state the most "natural" explanation for some time was in the spirit of the Emery and Kivelson model. In this picture it was imagined that while the cuprates exhibited a strong pairing associated with the T^* energy scale, true bulk superconductivity was only achieved upon doping and lowering the temperature to the point that classical thermal phase fluctuations between pairs could be overcome. Once phase coherence was established across the sample a true condensate of strongly bound Cooper pairs could form. In this picture, the Fermi arc was the result of a loss of phase coherence around the node at elevated temperatures that subsequently caused the order parameter in this region to average to zero, thus forming arcs in the single particle density of states where the superconducting gap is weakest. The recent observation, still in doubt, that the length of this Fermi arc scales with T/T^* appeared to support an argument of this type[61]. This argument is essentially destroyed by the observation of an asymmetric gap in the underdoped material, e.g. one not centered at zero energy, evolving smoothly from the arc region to beyond the arc. This point past the arc (Figure 8.5b) would under symmetrization appear as a pairing gap but in fact is due to another interaction altogether. Thus the Fermi arc is not simply the result of a partially dephased pre-superconducting state but something else altogether.

As it turns out, a far more satisfactory and ultimately more profound explanation for why we observe node-centered arcs of Fermi surface in the normal state of underdoped cuprates has its origins in an equally old but until recently unverifiable idea. This idea, simply put, is that doped Mott insulators form an entirely new taxonomic kingdom of materials on a par with metals and semiconductors. Undoped semiconductors have a band gap and no Fermi surface, both qualities corresponding to the absence of free carriers in the system at low temperatures. Metals on the other hand are characterized by a well defined Fermi surface, the area of which in the Brillouin zone is directly proportional to the number of free carriers in the system, which is in turn proportional to the filling fraction of the bands crossing the Fermi level and forming the Fermi surface. This is the Luttinger sum rule. In the language of many body theory it is said that the Fermi surface occurs at the zero energy poles of the electron Greens function.

The undoped Mott insulator defies these expectations. If one were to simply count the filling of these systems from an examination of the chemistry involved in the copper oxide plane one would arrive at the conclusion that the valence band is half filled and so there should be a metallic Fermi surface whose area totals up to half that of the full Brillouin zone. LDA certainly gives this picture and it is the LDA Fermi surface that, though wrong, is drawn in the literature and so far in this thesis as a means of at least orienting where in the zone we are. However, because of strong correlation effects arising from electron-electron Coulomb interactions then through strong superexchange coupling J it is more energetically favorable for this system to form an insulating antiferromagnet of the type depicted in Chapter 2.

Now the question arises as to what happens when you *dope* a Mott insulator, as is routinely done to achieve high superconducting transition temperatures in the cuprates, by adding extra holes or electrons. As we've said before it *appears* that we are faced with a paradox, for although the cuprates gradually begin to acquire upon doping some of the properties one associates with a metal, e.g. they conduct electricity and develop electronic bands that cross the Fermi level in places we call arcs, no normal state Fermi surface develops and so there is no obvious relationship between the valence band filling (now less than one half in the hole doped curprates) and the number of free carriers in the system. Further, there is no obvious way to relate these seemingly simple properties to the low energy electronic states of the system that in a real metal would contain almost all the relevant information we seek.

Early in the theoretical exploration of the cuprates it was suggested that many of the normal state properties of the system could be explained if rather than forming a single large open Fermi surface pocket centered at (π, π) with an area corresponding to the overall filling $1/2+\delta$ (the $1/2$ originating from the original half filled band), where δ is the doping fraction, small, closed Fermi surface pockets formed in the nodal region encompassing an area proportional to δ itself[62]. This picture makes the rather appealing physical argument that you really should always have a pocket of some kind with an area corresponding to the number of *free* carriers in the system even if the rest of the carriers of the system are frozen out of the equation do to lingering strong correlation effects. It also provides a nice way of understanding one of the fundamental problems of correlated electron systems in general, i.e. how a change of carrier concentration not only causes a shift in the chemical potential as would be the case for a metal or semiconductor, but how the overall band structure can undergo significant qualitative changes as a function of doping that have no place in traditional band structure calculations. Clearly, if correct, this statement would amount to saying the doped Mott insulators form a classification of conducting materials all their own and so of course the vast majority of bulk and even microscopic properties measured in such systems should appear "anomalous".

With the development of ARPES as a technique to directly and accurately map the Fermi surfaces of a wide range of materials off limits to traditional, more indirect measurements it would seem a trivial matter to examine the locations of electronic states around the Fermi level across the whole Brillouin zone to determine whether and how such pockets come to be. And indeed over the course of many years this measurement has been performed on a variety of cuprates at various doping levels always with the same conclusion: there are no "little pockets" to be found, only disconnected arcs of density of states at the Fermi level – Fermi arcs – in the normal pseudogap state. Even with the recent advent of de Haas-van Alphen measuremnts on pathologically clean YBCO samples in ultrahigh magnetic fields appearing to show the existence of small Fermi pockets in this state[63], the issue has remained far from clear precisely because the samples used in these measurements are considered too unrepresentative of the wider body of cuprates and because no one really knows what happens when you apply a 60 Tesla (!!!) magnetic field to a very complex material.

As the reader will no doubt have gathered by now from this lengthy discussion of cuprate pocketology and some of the figures presented above we believe the idea that small nodal Fermi pockets emerge upon doping the Mott insulator has quite a bit of merit to it after all. First, in a theoretical development that occurred prior to the experiments discussed here, it was shown that under certain circumstances the Luttinger sum rule has to be modified [9]. Specifically, it was shown that the introduction of dopants into a disordered spin liquid spiritually associated with the cuprates requires a modification of how one defines a pocket. In systems of this type it turns out that pockets are not formed by a Fermi surface as such but by a Fermi "arc" defined by the poles of the electron Greens function at zero energy that merges seamlessly with a "Luttinger arc" formed by the zeros of the Greens function at zero energy. This Fermi-Luttinger pocket, or "hot pocket", manifests itself as a Fermi arc because the spectral weight associated with the zeros of the Greens function (the Luttinger surface) is completely suppressed by coherence factors as one moves away from the Fermi part of the surface, which remains visible to singly particle probes such as photoemission.

This theory appears to explain rather conveniently how one can have a pocket only half of which is visible to a probe of electron density of states such as ARPES, and indeed one has a right to wonder how such a phenomenon could ever be observed as by definition half of what you'd look for in an experiment is invisible. The theory as presented at the time gave no obvious prescription as to what to look for in what *can* be observed to settle the issue one way or the other. Here we maintain that our measurements do in fact show an experimental signature of such a phenomenon. In the course of interpreting the data presented in Figure 3 showing electron-hole asymmetry

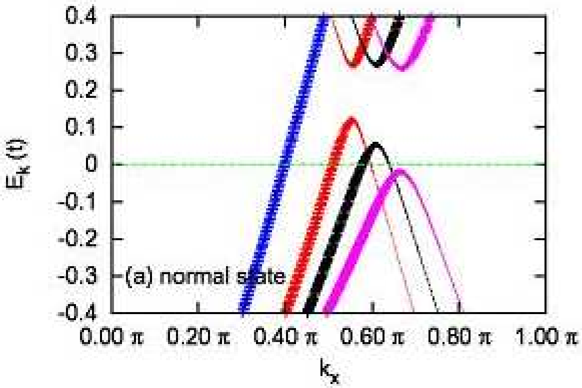


Figure 8.6: Model calculation from Yang et al. (unpublished) showing how the band structure is renormalized away from the nodes by the opening of a gap above E_F. The thickness of the bands schematically denotes the relative spectral weights of the “rising” and “falling” parts of the band arising from coherence factors similar to the BCS u_k^2 and v_k^2.

along and past the Fermi arc in the normal pseudogap state it was observed that if one were to symmetrize the FNS acquired midway along the arc in momentum, rather than energy, about the k point at which the band appears to reach the edge of a gap, the resulting spectrum would form a parabola the invisible side of which recrosses the chemical potential just shy of the antiferromagnetic zone boundary running from $(\pi, 0)$ to $(0, \pi)$. This observation of a) a dispersing energy gap opening above the Fermi level, b) loss of coherent weight for k points beyond this gap edge and c) the inference of an invisible yet mathematically necessary band forming the other side of a pocket at a meaningful symmetry point in the Brillouin zone provides strong evidence for the type of Fermi-Luttinger pocket described above and in the reference. The situation in energy-momentum space is shown in Figure 8.6, adapted from reference [62]. The zero energy surface, i.e. the form of the pocket, inferred from our measurements is shown in Figure 8.5, panel h. Further ARPES measurements around the edge of the pocket arc acquired since our study agree with and extend this picture. The area of this pocket is approximately $\delta/2$ as it should be for a bilayer system in which the doped carriers must almost equally fill the bilayer split bonding and antibonding bands. Further, this picture of a pocket existing only on one side of the magnetic zone boundary is in qualitative agreement with recent tunneling spectroscopy measurements[64].

Additionally, recent microwave conductivity measurements appear to bare out this picture. Broun *et al.* [65] observed that for YBCO with dopings close to 5 percent per copper oxygen plane that $T_c \propto (\delta - \delta_0)^{1/2}$ while the zero temperature extrapolated superfluid density $\rho_{s0} \propto \delta - \delta_0$ where δ_0 is the critical doping at which superconductivity sets in: five percent per plane. Such a dependence can be derived under the pocket picture as follows. At very low dopings we can, to first order, approximate the area of a little pocket as a circle so $A = \pi k_r^2$ where k_r is the radius of the pocket in reciprocal space. The doping this area represents is $\delta = 2\pi k_r^2 + \delta_0$ where the factor of two accounts for the fact we can put two electrons of opposite spin at each momentum as we fill the bands. It has been shown [61] that the Fermi arc length $L \propto T/T^*$ for underdoped samples above T_c and thus the arc length above T_c, $L_c = DT_c/T^*$ where D is a constant of proportionality on the order of unity. Approximating this arc length, which contains all the density of the electronic states at the Fermi level just above T_c as half the circumference of the circle enclosing the pocket we find $L_c = \pi k_r = DT_c/T^*$. Eliminating k_r between the equations for area and circumference we find

$$\delta_{plane} - \delta_0 = \frac{2}{\pi}(\frac{DT_c}{T^*})^2 \tag{8.1}$$

which is to say $T_c \propto (\delta - \delta_0)^{1/2}$.

A more honest but difficult calculation in which we take the pockets to be elliptical rather than circular also yields a quadratic dependence of the doping on T_c but with a different numerical prefactor. For an arbitrarily shaped elliptical pockets as one would have for higher dopings one can approximate the arc length by a series expansion in the semimajor axis of the ellipse or, more accurately, by incomplete elliptical integrals performed numerically. The supposed T^* dependence of the Fermi arc length, which is by no means settled at this time, is at any rate small over the narrow doping range examined in the experiment and at higher dopings may account for the deviation from the simple quadratic dependence of T_c on the doping. It has been shown that the data of Broun *et al.* is tantamount to the relationship $\rho_{s0} \propto T_c^2$ and thus proportional to the doping. If we take the density of states at the fermi level $N(0) \propto L_c$ so that effectively $T_c \propto N(0)^2$ we find that the cuprates differ drastically from BCS superconductors with a pairing energy proportional to $e^{-1/N(0)V}$ and to $N(0)V$ (where V is a sample volume) for weak and strong couplings, respectively [5].

While more measurements at other doping concentrations and temperatures are clearly required to verify the consistency of this picture for more than one setting the data is clearly consistent with a small Fermi-Luttinger pocket picture in which small pockets form on only one side of the magnetic zone boundary. The asymmetric shape of this pocket in k space, falling on only one side of the antiferromagnetic zone boundary, indicates that superconductivity cannot be caused by the presence of the well known $\boldsymbol{Q} = (\pi, \pi)$ magnetic mode which many have presumed forms the Cooper pairing "glue" of the cuprates. Pockets consistent with that picture would form symmetrically about the antiferromagnetic zone boundary. Similarly, any attempt to explain the normal pseudogap ground state as being conventionally metallic, as suggested by recent quantum oscillation experiments, will also fail.

To summarize, FNS of the nodal region strongly indicate that doped Mott insulators, of which underdoped Bi2212 is a classic example, are characterized by the opening of a gap away from the node above the chemical potential, the result of which is consistent with the formation of a very special kind of Fermi(-Luttinger) surface pocket whose area is proportional to the doping fraction δ ($\delta/2$ in the bilayer system) rather than the overall carrier concentration per plane $(1/2)(1 + \delta)$. This observation is inconsistent with the idea that the formation of incoherent Cooper pairs in the normal state above T_c originates in the nodal region of the Brillouin zone. Further measurements to pinpoint the binding energy of the upper side of this gap as well as to better understand the doping, temperature and k dependence of this phenomenon are clearly required. Because it is thought that the size of these pockets, as determined by the Fermi arc length, increases with increasing temperature as demonstrated by the T/T^{*} scaling of the Fermi arc length, reflecting the shrinking of the pseudogap, there is clearly a great deal of work that needs to be done to connect the physics of the nodal and antinodal regions into a seamless whole.

8.3.2 The Antinodal Region

As always in the cuprates the discussion must now return to the antinodal region and the physics of the classic normal state pseudogap. This discussion is made more difficult by fact that new and intriguing experimental results that appear to finally be shedding some light on the nature of this state have been entering the literature at a furious pace. For the moment the best we can do is to interpret our experiment on its own merits and then attempt to place it in the larger context of what is thought to be known about the pseudogap at the present time.

To begin with, the data presented above clearly indicate the presence of an energy symmetric gap both above and below T_c in the pseudogap region of the 65 K underdoped Bi2212 material. This characteristic symmetry of the energy gap is a necessary but not sufficient condition for the existence of Cooper pairs either above or below T_c. Further, it is not at all clear at present how one deals with the topological problem of the manner in which states forming the boundary of this symmetric gap evolve, presumably smoothly, into the asymmetrically gapped pseudogap states observed close to the Fermi pockets, as exemplified by the FNS in Figure 8.5b. That said, if we take at face value the Nernst effect measurements showing strong diamagnetism and vortex flow in the pseudogap state above T_c but below T^* as evidence for some kind of Cooper pair syrup in the normal state [8] then we are again led by process of elimination to the $(\pi, 0)$ states that are here seen to exhibit features consistent with such pairing. Further, the lack of a true condensate in the presence of this pairing is easily understood to result from the obvious lack of coherence observed in the states above T_c.

The location of these pairing states indicates they occur along the copper-oxide bonds on real space. That these pairs are highly localized on the real space lattice is seen both by their location in k space and by the fact that the electrons in these states have nearly infinite mass in that direction due to the presence of the extended van Hove singularity along the $\Gamma - M$ direction in the Brillouin zone. That Cooper pairs can be forced by real space constraints into what is effectively an insulating state has been demonstrated on an engineered system using conventional superconducting materials[66]. That such behavior might be able to arise naturally in a single crystal then should perhaps not come as too big a surprise. The lack of coherence between such isolated Cooper pairs results naturally from the high temperature of the system and accompanying classical thermally induced phase fluctuations, the extremely short coherence length for such a pair if one were to attempt to calculate it in the literal BCS fashion as $\xi_0 = \frac{\hbar v_F}{\pi \Delta_0}$ and the fact that Cooper pairs, as bosons, are subject to the fact that pair number and wavefunction phase are conjugate variables, i.e. that $\Delta N \Delta \phi \geq \hbar/2$. If we fix the number of particles by real space localization then the conjugate phase of the wavefunction is necessarily indeterminate and so no phase locking across pairs can occur.

In addition to these observations from ARPES and how they interact with the most basic level of theory perhaps the most revealing additional piece of information comes from tunneling spectroscopy. While the spectroscopic information offered by this technique is somewhat limited compared to ARPES because it measures a zone-integrated density of states modulated by some unknown tunneling matrix element, this handicap is made up for by the ability to resolve in real space where certain spectroscopic signatures originate from. Recently, it has been shown by several groups that even well into the superconducting state of the underdoped materials states that in energy correspond to the pseudogap states seen around the antinode in ARPES form a real space "checkerboard" pattern. This checkerboard is incommensurate with the atoms of the lattice (that is, it resides between atoms) and has no long range order. Interpretations of this phenomenon in the tunneling community range between simply observing that the pseudogap even below T_c corresponds to glassily ordered bound excitations in real space to their being the result of a more or less conventional charge density wave with an onset temperature of T^* and a total absence of long range order[67].

8.4 Tieing it All Together

These views represent the two main strains of belief in pseudogap physics. On one side are those who believe the pseudogap reflects the presence of an ordered ground state in competition with superconductivity, e.g. some kind of electronic density wave. On the other side are those who believe that the pseudogap, if originating from preformed pairs, causes or is at least a necessary ingredient for high temperature superconductivity in the underdoped cuprates. Our experiment points to the latter explanation. That is, the pseudogap is largely representative of the formation of some kind of quasi-ordered array of localized Cooper pairs with a pairing temperature on the order of T^*. Around T^ν they begin to acquire enough phase coherence to support localized supercurrents robust enough to screen a magnetic field and produce vortices but that are unable to form a bulk superconducting condensate across the whole sample. On the other hand there are good reasons to believe that the pseudogap pairs do not participate in the superconducting condensate that emerges, presumably from the Fermi arcs, below T_c.

This order of things in fact must be true if we make the following observation. We know that the pseudogap states remain localized even below T_c from tunneling and that both above and below T_c they exhibit an energy gap consistent with, if not proof of, pairing. Nernst effect measurements suggest pairing in the normal state though there have been plausible suggestions that such a signal could arise from the presence of a more conventional density wave. How can we tell the difference? An examination of Figure 8.5 gives the answer. The states observed closest to the antinode exhibit a drastic increase in coherence below T_c as compared with above and the states observed close to but just off of the Fermi pocket not only exhibit an onset of coherence below T_c but actually show that superconductivity forces electron-hole symmetry onto this previously asymmetric gap. The simplest explanation for these observations is that the pseudogap results from frozen pairs that acquire phase coherence, if not mobility, by sympathy with the condensate formed by the onset of true superconductivity in the nodal region. Conversely, there is no obvious reason anything at all should happen with respect to coherence in the pseudogap states were they to be indicative of the presence of an order in competition with superconductivity.

There is clearly more to this story. Experimental evidence has recently been piling up showing that the condensate below T_c originates from the Fermi arc. In this picture one recovers something at least in the same spirit of BCS in the sense that one requires a density of electronic states at the Fermi level out of which to form mobile Cooper pairs that will condense and carry a current. We have not touched at all on how this pairing might actually come to pass except to rule out some possibilities, e.g. the magnetic resonance, in the underdoped systems. We have also not so much as mentioned the overdoped cuprates which, while presumed to be much more "conventional" in nature because there is no obvious signature of a pseudogap, nevertheless defy explanation as to how they can acquire a T_c ranging up to 90 K in Bi2212. These questions lie beyond the reach of the experiments described in this thesis and so will have to await the patience of future students and post docs.

Bibliography

[1] A. Damascelli, Z. Hussain, and Z.-X. Shen, Rev. Mod. Phys. **75**, 473 (2003).

[2] M. P. Seah and W. A. Dench, Surface and Interface Analysis **1**, 2 (1979).

[3] T. Valla, T. E. Kidd, J. D. Rameau, H.-J. Noh, G. D. Gu, P. D. Johnson, H.-B. Yang, and H. Ding, Phys. Rev. B **73**, 184518 (pages 5) (2006).

[4] N. N. Kovaleva, A. V. Boris, T. Holden, C. Ulrich, B. Liang, C. T. Lin, B. Keimer, C. Bernhard, J. L. Tallon, D. Munzar, et al., Phys. Rev. B **69**, 054511 (2004).

[5] J. R. Schrieffer, *Theory of Superconductivity (Revised Printing)* (Addison-Wesley Publishing Company, 1988).

[6] J. G. Bednorz and K. A. Muller, Z. Physik, B **64**, 189 (1986).

[7] S. Hufner, M. A. Hossain, A. Damascelli, and G. A. Sawatzky, Rep. Prog. Phys. **71**, 062501 (2008).

[8] N. P. Ong and Y. Wang, Physica C: Superconductivity **408-410**, 11 (2004).

[9] R. M. Konik, T. M. Rice, and A. M. Tsvelik, Phys. Rev. Lett. **96**, 086407 (2006).

[10] A. Ino, C. Kim, M. Nakamura, T. Yoshida, T. Mizokawa, A. Fujimori, Z.-X. Shen, T. Kakeshita, H. Eisaki, and S. Uchida, Phys. Rev. B **65**, 094504 (2002).

[11] D. Parker and A. V. Balatsky, Phys. Rev. B **78**, 214502 (2008).

[12] H. Ding, T. Yokoya, J. C. Campuzano, T. Takahashi, M. Randeria, M. R. Norman, T. Mochiku, K. Kadowaki, and J. Giapintzakis, Nature (London) **51** (1996).

[13] T. Kondo, R. Khasanov, T. Takeuchi, J. Schmalian, and A. Kaminski, Nature (London) **457** (2009).

[14] Ashcroft and Mermin, *Solid State Physics* (Saunders College Publishing, 1976).

[15] T. Valla, A. V. Fedorov, P. D. Johnson, B. O. Wells, S. L. Hulbert, Q. Li, G. D. Gu, and N. Koshizuka, Science **285**, 2110 (1999).

[16] A. Lanzara, P. V. Bogdanov, X. J. Zhou, S. A. Kellar, D. L. Feng, E. D. Lu, T. Yoshida, H. Eisaki, A. Fujimori, K. Kishio, et al., Nature (London) **412**, 510 (2001).

[17] P. D. Johnson, T. Valla, A. V. Fedorov, Z. Yusof, B. O. Wells, Q. Li, A. R. Moodenbaugh, G. D. Gu, N. Koshizuka, C. Kendziora, et al., Phys. Rev. Lett. **87**, 177007 (2001).

[18] B. P. Xie, K. Yang, D. W. Shen, J. F. Zhao, H. W. Ou, J. Wei, S. Y. Gu, M. Arita, S. Qiao, H. Namatame, et al., Phys. Rev. Lett. **98**, 147001 (2007).

[19] T. Valla, T. E. Kidd, W.-G. Yin, G. D. Gu, P. D. Johnson, Z.-H. Pan, and A. V. Fedorov, Phys. Rev. Lett. **98**, 167003 (pages 4) (2007).

[20] H. Iwasawa, J. F. Douglas, K. Sato, T. Masui, Y. Yoshida, Z. Sun, H. Eisaki, H. Bando, A. Ino, M. Arita, et al., Phys. Rev. Lett. **101**, 157005 (2008).

[21] J. D. Jackson, *Classical Electrodynamics* (John Wiley and Sonse, Inc., 1999), 3rd ed.

[22] S. Hüfner, *Photoelectron Spectroscopy: Principles and Applications* (Springer, 2003), 2nd ed.

[23] A. Damascelli, PHYSICA SCRIPTA T **109**, 61 (2004).

[24] Y. Takata, Y. Kayanuma, S. Oshima, S. Tanaka, M. Yabashi, K. Tamasaku, Y. Nishino, M. Matsunami, R. Eguchi, A. Chainani, et al., Phys. Rev. Lett. **101**, 137601 (2008).

[25] S. D. Kevan, Review of Scientific Instruments **54**, 1441 (1983), URL `http://link.aip.org/link/?RSI/54/1441/1`.

[26] P. D. Johnson, J. Elec. Spec. and Related Phenomena **126**, 133 (2002).

[27] T. Kiss, F. Kanetaka, T. Yokoya, T. Shimojima, K. Kanai, S. Shin, Y. Onuki, T. Togashi, C. Zhang, C. T. Chen, et al., Phys. Rev. Lett. **94**, 057001 (2005).

[28] J. D. Koralek, J. F. Douglas, N. C. Plumb, Z. Sun, A. V. Fedorov, M. M. Murnane, H. C. Kapteyn, S. T. Cundiff, Y. Aiura, K. Oka, et al., Phys. Rev. Lett. **96**, 017005 (2006).

[29] E. D. Hansen, T. Miller, and T.-C. Chiang, Phys. Rev. Lett. **80**, 1766 (1998).

[30] R. Loudon, *The quantum theory of light* (Clarendon Press, Oxford, 1983), 2nd ed.

[31] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, Phys. Rev. Lett. **7**, 118 (1961).

[32] P. W. Milonni and J. H. Eberly, *Lasers* (John Wiley and Sons, 1988).

[33] SNLO nonlinear optics code available from A. V. Smith, AS-Photonics, Albuquerque, NM.

[34] J. Lu, G. Wang, Z. Xu, C. Chen, J. Wang, C. Zhang, and Y. Liu, Optics Communications **200**, 415 (2001).

[35] E. Kisker, E. F. Wassermann, and C. Carbone, Phys. Rev. Lett. **58**, 1784 (1987).

[36] M. R. Norman, H. Ding, M. Randeria, J. C. Campuzano, T. Yokoya, T. Takeuchi, T. Mochiku, K. Kadowaki, P. Guptasarma, and D. G. Hinks, Nature (London) **392**, 157 (1998).

[37] D. V. Evtushinsky, D. S. Inosov, V. B. Zabolotnyy, A. Koitzsch, M. Knupfer, B. Buchner, M. S. Viazovska, G. L. Sun, V. Hinkov, A. V. Boris, et al., Phys. Rev. B **79**, 054517 (2009).

[38] N. J. C. Ingle, K. M. Shen, F. Baumberger, W. Meevasana, D. H. Lu, Z.-X. Shen, A. Damascelli, S. Nakatsuji, Z. Q. Mao, Y. Maeno, et al., Phys. Rev. B **72**, 205114 (2005).

[39] J. Shi, S.-J. Tang, B. Wu, P. T. Sprunger, W. L. Yang, V. Brouet, X. J. Zhou, Z. Hussain, Z.-X. Shen, Z. Zhang, et al., Phys. Rev. Lett. **92**, 186401 (2004).

[40] X. J. Zhou, J. Shi, T. Yoshida, T. Cuk, W. L. Yang, V. Brouet, J. Nakamura, N. Mannella, S. Komiya, Y. Ando, et al., Phys. Rev. Lett. **95**, 117001 (2005).

[41] T. Valla, Phys. Rev. Lett. **96**, 119701 (2006).

[42] X. J. Zhou, J. Shi, W. L. Yang, S. Komiya, Y. Ando, W. Plummer, Z. Hussain, and Z.-X. Shen, Phys. Rev. Lett. **96**, 119702 (2006).

[43] H.-B. Yang, J. D. Rameau, P. D. Johnson, T. Valla, A. Tsvelik, and G. D. Gu, Nature (London) **456**, 77 (2008).

[44] W. H. Richardson, J. Opt. Soc. Am. **62**, 55 (1972).

[45] L. B. Lucy, Astronomical Journal **79**, 745 (1974).

[46] S. Hummel, A. Gross, and W. S. M. Werner, Surface and Interface Analysis **41**, 357 (2009).

[47] W. S. Lee, I. M. Vishik, K. Tanaka, D. H. Lu1, T. Sasagawa1, N. Nagaosa, T. P. Devereaux, Z. Hussain, and Z.-X. Shen, Nature (London) **450**, 81 (2007).

[48] A. W. Sandvik, D. J. Scalapino, and N. E. Bickers, Phys. Rev. B **69**, 094523 (2004).

[49] A. A. Tsvetkov, D. Dulić, D. van der Marel, A. Damascelli, G. A. Kaljushnaia, J. I. Gorina, N. N. Senturina, N. N. Kolesnikov, Z. F. Ren, J. H. Wang, et al., Phys. Rev. B **60**, 13196 (1999).

[50] M. Kakihana, M. Osada, M. Käll, L. Börjesson, H. Mazaki, H. Yasuoka, M. Yashima, and M. Yoshimura, Phys. Rev. B **53**, 11796 (1996).

[51] R. Liu, M. V. Klein, P. D. Han, and D. A. Payne, Phys. Rev. B **45**, 7392 (1992).

[52] O. V. Misochko and E. Y. Sherman, J. Phys.: Condens. Matter **12**, 9095 (2000).

[53] G. Grimvall, *The Electron-Phonon Interaction In Metals* (North-Holland, New York, 1981).

[54] E. Schachinger, J. J. Tu, and J. P. Carbotte, Phys. Rev. B **67**, 214508 (2003).

[55] A. Comanac, L. de' Medici, M. Capone, and A. J. Millis, Nature Physics **4**, 287 (2008).

[56] F. Giustino, M. L. Cohen, and S. G. Louie, Nature (London) **452**, 975 (2008).

[57] W. Meevasana, N. J. C. Ingle, D. H. Lu, J. R. Shi, F. Baumberger, K. M. Shen, W. S. Lee, T. Cuk, H. Eisaki, T. P. Devereaux, et al., Phys. Rev. Lett. **96**, 157003 (2006).

[58] X. J. Zhou, T. Yoshida, A. Lanzara, P. V. Bogdanov, S. A. Kellar, K. M. Shen, W. L. Yang, F. Ronning, T. Kakeshita, T. Noda, et al., Nature (London) **423**, 398 (2003).

[59] H. Matsui, T. Sato, T. Takahashi, S.-C. Wang, H.-B. Yang, H. Ding, T. Fujii, T. Watanabe, and A. Matsuda, Phys. Rev. Lett. **90**, 217002 (2003).

[60] W. S. Lee, I. M. Vishik, K. Tanaka, D. Lu, T. Sasagawa, N. Nagaosa, T. Devereaux, Z. Hussain, and Z. Shen, Nature (London) **81** (2007).

[61] A. Kanigel, M. R. Norman, M. Randeria, U. Chatterjee, S. Souma, A. Kaminski, H. M. Fretwell, S. Rosenkranz, M. Shi, T. Sato, et al., Nature Physics **2**, 447 (2006).

[62] X.-G. Wen and P. A. Lee, Phys. Rev. Lett. **80**, 2193 (1998).

[63] N. Doiron-Leyraud, C. Proust, D. LeBoeuf, J. Levallois, J.-B. Bonnemaison, R. Liang, D. A. Bonn, W. N. Hardy, and L. Taillefer, Nature (London) **447**, 565 (2007).

[64] Y. Kohsaka, C. Taylor, P. Wahl, A. Schmidt, J. Lee, K. Fujita, J. W. Alldredge, K. McElroy, J. Lee, H. Eisaki, et al., Nature (London) **454**, 1072 (2008).

[65] D. M. Broun, W. A. Huttema, P. J. Turner, S. Ozcan, B. Morgan, R. Liang, W. N. Hardy, and D. A. Bonn, Physical Review Letters **99**, 237003 (pages 4) (2007).

[66] J. Stewart, M. D., A. Yin, J. M. Xu, and J. Valles, James M., Science **318**, 1273 (2007).

[67] W. D. Wise, M. C. Boyer, K. Chatterjee, T. Kondo, T. Takeuchi, H. Ikuta, Y. Wang, and E. W. Hudson, Nature Physics **4**, 696 (2008).