(53)

Modeling the Deflector Mass Distribution in the Gravitational Lens MG J0414+0534

by

Catherine Susanne Trotter Wilson

S.B. Physics, Massachusetts Institute of Technology (1989)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1998

Author . .....
Department of Physics
19 Dec 1997

Certified by. .....
Jacqueline N. Hewitt
Associate Professor of Physics
Thesis Supervisor

Accepted by ....
George F. Koster
Professor of Physics
Chairman, Graduate Committee

Modeling the Deflector Mass Distribution in the Gravitational Lens MG J0414+0534

by

Catherine Susanne Trotter Wilson

Submitted to the Department of Physics
on 19 Dec 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

Gravitational lenses MG J0414+0534, B0218+357, B0957+561, B1422+231, and PKS B1830-211 were observed at 5 GHz with the VLBA.

MG0414+0534 revealed substructure not seen at lower resolutions: four subcomponents in each lensed image A1, A2, B, and C, with clear, consistent mapping between subcomponents. This provides more flux and position constraints for deflector modeling than other lenses provide.

Lens modeling requires both source and deflector models. By treating subcomponents as pointlike, analytic expressions are given for model source flux and position parameters, reducing the parameter search space dimensionality. Image centroid may not correspond to source centroid for non-pointlike subcomponents; a correction for this was derived.

The deflector potential was multipole-expanded in angle; each multipole component's radial dependence was Taylor-expanded in distance from the Einstein ring radius. Slightly regrouping the expansion terms, their coefficients give the multipole moments of mass exterior to, interior to, and at the ring radius. The dominant terms' coefficients parameterize 0414+0534's deflector model.

The best fit model has interior and exterior quadrupole moments, and exterior $m = 3$ and $m = 4$ multipoles. The radial profile was unconstrainable. The deflector centroid matches the optical galaxy position. Both quadrupoles are aligned with the optical isophotes. The exterior multipoles, not attributable to a single compact perturber, indicate asymmetry and boxiness in the galaxy's outer regions.

The model-predicted dimensionless B–C time delay is 1.225×10^{-11} with $+14\%/-15\%$ systematic error from choice of model form, and 1.7% statistical error. Combined with the yet-to-be-measured time delay and deflector redshift, this determines the Hubble parameter, with slight dependence on cosmology, 75 km/sec/Mpc$\times(1-\kappa)\times$34.34 days/Δt_{BC} with error from lens modeling as above, assuming Einstein-de Sitter cosmology with deflector redshift 0.5. Measuring the (B-A)/(A-C) time delay ratio may reduce lens modeling systematic error to $+0\%/-10\%$ for B–C or $+0\%/-4\%$ for B–A. The factor $1-\kappa$ arises from interaction of a possibly incorrect model value for the angularly-averaged surface mass density at the Einstein ring radius with the mass sheet degeneracy. $\kappa = 0$ for an isothermal profile.

Thesis Supervisor: Jacqueline N. Hewitt
Title: Associate Professor of Physics

Acknowledgments

I am grateful for the support of a National Science Foundation Graduate Fellowship and a Karl Taylor Compton Fellowship for the graduate study of physics at MIT, which allowed me to spend my first few years of graduate school exploring theoretical physics, and for the A. M. Lockett III Memorial Award for excellence in Theoretical Physics, which bought many of my books.

Thanks are also due to my first graduate research advisor Professor John Negele, George Ricker who gave me the opportunity to try astrophysics, and Mark Bautz who encouraged my interest in gravitational lenses.

I would also like to acknowledge my thesis advisor, Professor Jackie Hewitt, who had an abundance of interesting research projects to suggest, including the one which became this thesis, and who was always actively interested in the progress of the work. The other members of my thesis committee also provided assistance. Professor Simon Mochrie brought perspective from outside of astrophysics. Professor Paul Schechter I thank for many valuable discussions concerning the modeling of gravitational lenses.

Planning and carrying out the VLBA observation described in this thesis were done together with then graduate students Debbie Haarsma and Chris Moore.

I express my gratitude to our departmental administrator, Peggy Berkovitz, for her shielding the graduate students from the Institute's administrative hassles, and for her cheering us.

All my friends I thank for their friendship and help, both those from MIT and those from church, in particular Adam Szabo and Todd Rider who have remained fast over all these years.

On my families, the Trotters and the Wilsons, I know I can always depend.

My husband Kevin Wilson deserves special mention; he is more than a fellow graduate student, more than a best friend, in many ways the other part of me, without whom neither of us would be complete.

Finally I dedicate this thesis, and myself, to God, who made this universe in which we live and which we study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Why study gravitational lenses?

The phenomenon of gravitational lensing is an intriguing subject. The light from a background source is deflected by the gravitational potential of a foreground source resulting in a "mirage", a distorted image or images of the background object. The phenomenon is useful for a variety of astrophysical applications.

Magnification of source flux. The background object may be magnified. The apparent angular size of the background object increases, while the sky brightness remains unchanged. This increases the total flux arriving at the observer on earth. The increased flux and increased angular size allow the study of faint distant objects that otherwise would be very difficult to observe. For example, a recent interesting application was the measurement of the spectrum of a distant galaxy, revealing the redshift to be $z = 4.92$, the highest known redshift of any galaxy. The galaxy was magnified by a foreground cluster of galaxies increasing the observed flux sufficiently that the spectrum could be determined (Franx *et al.* 1997).

Dark matter. The deflection of light by the foreground object depends on all the mass, dark as well as luminous, allowing the study of the dark matter in the deflector. The amount and location of dark matter in the universe are currently

topics of great debate. For example, lensing of background galaxies is being used to map the dark matter in clusters of galaxies (Kaiser & Squires 1993).

Distance measurements and cosmological parameters. The light from the background object may reach the observer via multiple paths. The geometric figure defined by these paths samples the shape of space-time out to the redshift of the source, which can be large. If the difference between the light travel times along two paths to the source can be measured, as by seeing a flux increase or decrease in one image and days or months later the corresponding flux change in another image of a variable quasar, this time delay can be used to determine the size scale of the lens system. This determines the angular diameter distances in the lens system. It does, however, require that the gravitational potential of the deflector be adequately well known (adequately well modeled).

The relation between angular diameter distance and Doppler redshift, due to recession velocity caused by the expansion of the universe, depends on the universe's cosmology. For small redshifts the constant of proportionality between a source's distance D and its recession velocity v or consequent redshift z, due to the expansion of the universe, is simply the inverse of the Hubble parameter $D = v/H_o = cz/H_o$. For more distance objects with redshifts of order unity or greater, the relationship deviates from a straight line but the normalization is still given by H_o^{-1}, and the deviation from linearity is determined by the average mass density of the universe Ω_o and by the cosmological constant (vacuum energy density) $\Omega_{\Lambda o}$.

The Hubble parameter H_o is particularly interesting because its inverse H_o^{-1}, with dimension of time, sets the scale of the age of the universe, under the "big bang" scenario. In particular, for a Friedmann-Robertson-Walker universe with $\Omega_o = 1$ and $\Omega_{\Lambda o} = 0$ (an Einstein-de Sitter universe), the age of the universe is $t_{\text{universe}} = \frac{2}{3}H_o^{-1}$; for other mass densities Ω_o and cosmological constants $\Omega_{\Lambda o}$ the expansion history is somewhat different and the constant of proportionality between t_{universe} and H_o^{-1} differs from 2/3.

Since time delay measurements may be used to measure the effective distance to a gravitational lens, lens systems for which these time delays and redshifts are known can be used to determine the Hubble parameter. Such measurements have already been made for gravitational lenses PG B1115+080 ($H_o = 51^{+14}_{-13}$ km/sec/Mpc (Keeton & Kochanek 1997; Barkana 1997; Schechter *et al.* 1997)) and B0957+561, (between 60 and 75 km/sec/Mpc; see Haarsma (1997) for a good current discussion of the state of the B0957+561 determination of H_o). Measuring many such systems would reduce the problem that erroneous H_o values might result from poor understanding of particular lens systems.

The various other methods of measuring the Hubble parameter also leave room for improvement. Recent measurements, helpfully reviewed by Haarsma (1997), include $H_o = 54 \pm 14$ km/sec/Mpc using the Sunyaev-Zel'dovich effect (Myers *et al.* 1997), $H_o = 73 \pm 6$(statistical) ± 7(systematic) km/sec/Mpc using Type II supernovae (Schmidt *et al.* 1994), $H_o = 81 \pm 6$ km/sec/Mpc using surface brightness fluctuations of galaxies (Tonry *et al.* 1997), and $H_o = 64 \pm 6$ km/sec/Mpc using Type Ia supernovae as standard candles (Riess *et al.* 1996).

Using time delay measurements to measure the effective distance to a gravitational lens, for use in determining the Hubble parameter, the proportionality constant in the redshift-distance relation, is very appealing for several reasons: (1) The physics of the time delay phenomenon is well understood: deflection of light in a gravitational field and the slowing of time in a gravitational potential well. (2) The large distances involved, with redshifts of $\sim \frac{1}{2}$ to ~ 3, reduce the effects of peculiar velocities or bulk flows on the deduced Hubble parameter. (3) Uncertainties in the normalization of the cosmic distance scale do not affect this method, since this is a one-step distance measurement that bypasses the cosmic distance ladder. (4) Such lens-time-delay determinations of H_o are completely independent of the other techniques for the determination of the Hubble parameter, and should not be affected by the same systematic errors as

the other methods.

Measurements of H_o from multiple lens systems will serve as checks against errors due to peculiarities in individual lenses, in particular systematic error in modeling the lens's deflector gravitational potential. Measuring multiple systems also holds the possibility of constraining the average mass density of the universe Ω_o and the cosmological constant $\Omega_{\Lambda o}$, due to their effect on the redshift-distance relation at large redshifts.

The observations described in this thesis were initially motivated by the third point, in particular the determination for a large number of lenses of time delays, in quest of the Hubble parameter. In the course of observations of a number of lenses, it was found that the lensed radio source MG J0414+0534 has a great deal of detailed substructure on milliarcsecond scales (figure 1-1). This substructure makes this lens interesting for well-constrained modeling of the deflector mass distribution. Such modeling may be used to compare the dark with luminous matter in the galaxy which acts as the deflector. It also provides a good model of the deflector to await future time-delay and deflector redshift measurements of MG J0414+0534. This is important since the major source of error in the Hubble parameter measurements from gravitational lenses has been systematic error in the modeling of the deflector potential due to lack of available information on the potential.

1.2 Previous studies of MG J0414+0534

MG J0414+0534 was identified as a gravitational lens in a program, intended to find lens candidates, of Very Large Array (VLA) snapshot observations of sources from the MIT-Greenbank (MG) survey. This four-component astronomical object (A1, A2, B, C) was reported as a lens by Hewitt *et al.* (1992) who also noted that the optical counterpart was unusually red. Further observations with the VLA clearly separated components A1 and A2 and gave a flux ratio of A1/A2 $\sim$ 1.1 at radio wavelengths (Katz & Hewitt 1993).

The lensing galaxy was detected optically (I-band) by Schechter & Moore (1993). They also saw a faint object "X" to the west of component B. They noted that the A1/A2 optical flux ratio was > 2. This flux ratio was surprising since two merging images near a critical line are expected to have similar fluxes. It was also surprising because it was significantly different than the radio flux ratio. The A1/A2 optical flux ratio and galaxy detection were confirmed by the Hubble Space Telescope (HST) R-band observations of Falco (1993) and in I-band by Angonin-Willaime *et al.* (1994).

Angonin-Willaime *et al.* (1994) enumerated a plethora of possible reasons for the flux ratio discrepancy:

- extinction of A2 by the lensing galaxy
- source variability and time delay
- differential magnification or variability of A due to microlensing
- magnification gradients because the source is near a caustic of the lens, with the radio and optical flux coming from different regions of the source

Lawrence *et al.* (1995b) determined the source redshift to be $z_s = 2.639 \pm 0.002$ from an H_α spectral line. The deflector redshift is still unknown — the absorption lines seen in the spectrum are actually Fe II near the source redshift (Lawrence *et al.* 1995a). Lawrence *et al.* (1995b) argue that the deflector is very dusty, because of the unusual spectrum. The spectrum can be explained as a standard quasar spectrum with dust absorption near the expected $z_L \sim 0.5$ deflector redshift, but it cannot be explained by dust near the source. A dusty deflector has the added merit of explaining the extreme redness of the optical observations and the unexpected A1/A2 optical flux ratio — attributable to differing amounts of extinction along the two paths through the deflector.

Witt, Mao, and Schechter (1995) examined the statistics of microlensing by individual stars in the deflector galaxy, and found that microlensing was sufficient to explain the flux ratio radio-optical discrepancy. For a quasar, radio flux normally comes from a region much larger than that of the optical flux. For small sources one

expects significant flux variations due to microlensing by individual stars. A large size emission region would wash out the effects of microlensing. They work out the probability distribution for the log of the flux ratio, $\log(\frac{A1}{A2})$, for various source sizes. The optical flux ratio seen in MG J0414+0534 would be unusually but not outrageously large in the zero-source-size limit. They note that if microlensing is the correct explanation of the ratio discrepancy then it is expected that A1/A2 will revert closer to 1 over time, and they crudely estimate the time scale for a single microlensing event to be 12 years. If A1/A2 remains fixed at the value seen in the optical, then microlensing ceases to be a good explanation.

In a further twist on the flux ratio issue, Vanderriest *et al.* (1996) report optical (R-band) and infrared (K-band) A1/A2 flux ratios of 1.3 and 1.39, for observations made in August 1994, which are closer to the radio value of the flux ratio. However, in R-band, both A1 and A2 appear extincted relative to B: (A1+A2)/B is less in R-band than in K-band, whereas in K-band it is similar to (A1+A2)/B in the radio. They put forward as plausible explanations 1) extinction coupled with a microlensing event in A1 that had ended by 1994 or 2) extinction only. However the November 1994 R-band HST image of Falco *et al.* (1997) indicates that A1/A2 is 2.5 in R-band and 2.1 in I-band, casting doubt on the observation of Vanderriest *et al.* (1996).

Observations with the refurbished HST (Falco *et al.* 1997) in November 1994 reveal an arc extending from A1 to A2 to B, in both R- and I-bands. The arc is 0.9 magnitude bluer that the images of the core, and it is uniform in color (to 0.3 magnitude) along its length. This uniformity is an argument against patchy extinction in the deflector galaxy as a source of the A1/A2 flux ratio discrepancy. They also find that the galaxy's brightness distribution is characteristic of an elliptical galaxy and is well represented by elliptical isophotes with a de Vaucouleurs profile. Their A1/A2 flux ratios, 2.50 ± 0.04 in R-band and 2.11 ± 0.01 in I-band, are consistent with both Schechter & Moore (1993) (I-band) and Angonin-Willaime *et al.* (1994) (R- and I-bands).

Further radio observations by Katz, Moore, and Hewitt (Katz *et al.* 1997) placed limits on the flux of any fifth image in the system at 1.4, 5, 8, 15, and 22 GHz. They

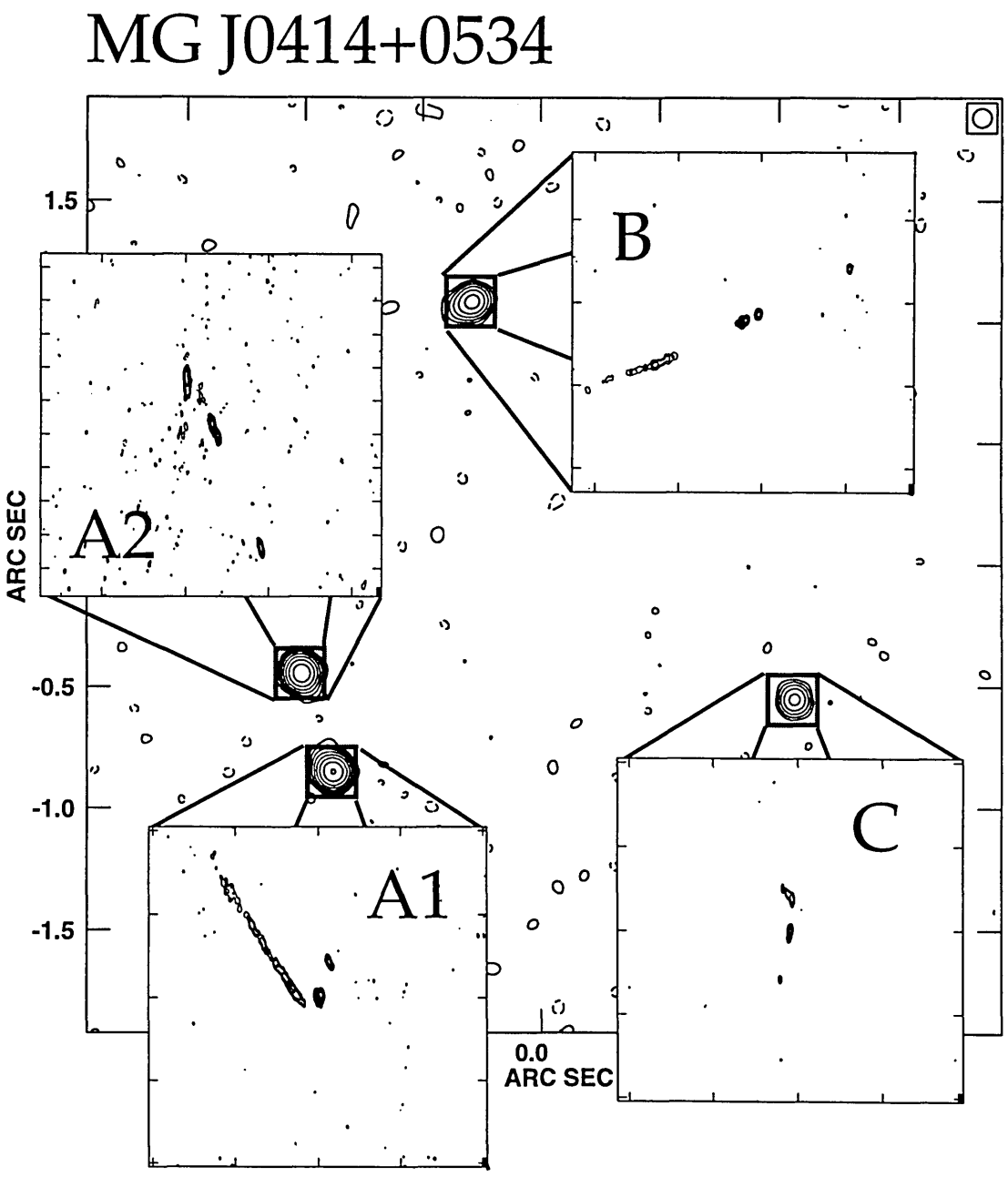

Figure 1-1: High resolution maps of MG J0414+0534 images A1, A2, B, and C from this thesis, observed at 5 GHz with the Very Long Baseline Array, are shown superimposed on the 22 GHz Very Large Array map of Katz *et al.* (1997) (kindly provided by C. Katz).

identified the background source as a GHz-peaked-spectrum (GPS) source. They found the spectral index from 5-22 GHz to be the same for all components A1, A2, B, and C: $\alpha = 0.80 \pm 0.02$, where α is defined as $S = S_o \nu^{-\alpha}$. The polarization fraction was found to be low. The A1/A2 flux ratio was measured at 5, 8, 15, and 22 GHz. At 8 GHz, the ratio was A1/A2 $= 1.114 \pm 0.002$. The error bars were somewhat larger at the other frequencies, with only slight disagreement, given the sizes of the error bars, between the flux ratios at the various frequencies.

In an attempt to determine the time delay, flux monitoring was undertaken by Moore (1996). In monitoring of MG J0414+0534 at 15 GHz with the VLA from November 1992 to May 1993, any intrinsic variability was masked by calibration errors of the order 3.4%. Moore concluded that any intrinsic variability was at most 1.7% on the time scales of interest, $\sim$ 20 days. He also monitored at 8 GHz with the VLA from March 1994 to January 1995. The RMS in the light-curves, corrected for deconvolution errors, was 1.4%, 1.5%, and 1.7% for A1+A2, B, and C respectively. This also failed to detect variability clearly attributable to source variability.

Higher resolution radio maps have shown substructure in the components A1, A2, B, and C. The Merlin 5 GHz map of Garrett *et al.* (1992) shows A1, A2, and B clearly extended, as do the 5 GHz Very Long Baseline Interferometry (VLBI) maps of Ellithorpe (1995). The European VLBI Network (EVN) 1.7 GHz map of Patnaik & Porcas (1996) shows A1, A2, and B resolved into primary and secondary components, and C extended.

In summary, issues remaining to be explained for this lens are its extreme redness, the uniform blueness of the arc relative to the redness of the core images, and the difference between the radio and optical A1/A2 flux ratios. This last can be accounted for by microlensing, but other explanations have not yet been ruled out. It remains to be determined whether object X is related either to the source or to the deflector. Observations of the time delay and of the deflector redshift are still needed.

1.3 Previous efforts to model the MG J0414+0534 deflector gravitational potential

There are two main reasons why effort is made to model the deflector gravitational potentials for gravitational lens candidates. Firstly, modeling is used as a plausibility check that a purported lens may indeed be a gravitational lens — by showing that there exists some potential that can approximately produce a particular image configuration. Secondly, modeling is done to learn about the mass distribution of the deflector either for the sake of learning about the dark matter or for the sake of calculating the time delay. To study the mass distribution or to predict the time delay one needs to know just how well the lensing constrains the deflector gravitational potential.

Kochanek (1991) studied how well the lens potentials were constrained for five different quad lens candidates including MG J0414+0534. He used a multipole expansion of the lens potential, including only the monopole and quadrupole terms. For the monopole he tried both a point mass and a singular isothermal sphere — the two extreme cases of physically reasonable radial dependencies. He considered quadrupole components due to an external mass distribution, due to a mass distribution completely contained in the region between the lens images, and a mixed case. (The external quadrupole also goes by the name of "external shear".) He found that all the quad lens candidates were moderately well fit by any monopole and quadrupole combination. The implication is that the radial dependence of the potential is very ill constrained. He noted that this is not surprising. Six relative positions from four images scarcely over-constrain the simplest non-circularly-symmetric model, which has five parameters. He made the point that there is little purpose in going to models with more complicated radial profiles if one cannot distinguish between simple models having such different radial profiles. Fluxes would provide more constraints, but optical fluxes in particular are suspect due to possible contamination from microlensing. Radio fluxes might be reasonable to include as constraints. Even so, all the constraints still fall within a very small range of radial distances from the center of

the deflector mass distribution. On the optimistic side, the lens critical radius, which is proportional to the total mass within the region bounded by the lens images, is not sensitive to the particular shape of the potential, and is very well constrained. The orientation angle of the ellipticity, too, tends to be well constrained and insensitive to the model used. The amplitude of the ellipticity depends crucially on the form of the monopole term — the shear is almost exactly twice as large in point mass models as in singular isothermal sphere models, an oddity explained by Kochanek.

Even in the initial MG J0414+0534 publication (Hewitt *et al.* 1992), a simple model for the deflector mass distribution is given. It is an elliptical potential which reduces to a singular isothermal sphere in the no-ellipticity case, very similar but not identical to Kochanek's monopole plus quadrupole models. They find that this can reasonably well reproduce the image positions and magnifications, but they did not do a quantitative analysis of the goodness of fit.

The most precise optical positions for the components are those of Falco *et al.* (1997), with 20 milliarcsecond errors on the image positions and 50 milliarcsecond errors on the galaxy position. Four different models are fitted: 1) mass follows light, 2) the elliptical potential of Hewitt *et al.* (1992), 3) a point mass with an external quadrupole (external shear), and 4) a singular isothermal sphere with an external quadrupole (external shear). Formally none of the fits were satisfactory, the mass-follows-light being definitely the worst. The other three models, with an adjustable amount of quadrupole contribution, did better. The point mass with shear was best, but was not much different in chi-squared per degree of freedom (χ^2/DOF) from the singular isothermal sphere with shear (χ^2/DOF of 4.5 and 5.5 respectively for 3 degrees of freedom). These results serve to illustrate Kochanek's (1991) point: the radial profile of the deflector's potential is ill constrained by the four-image quad lenses, even while the angle of the ellipticity and the Einstein radius are well constrained. Also the amount of shear depends on the form of the monopole term.

The radio observations of Katz *et al.* (1997) have even more precise positions than the HST optical data, with components' position errors of 10 milliarcseconds. The 15 GHz observation was modeled by Ellithorpe (Ellithorpe 1995; Ellithorpe *et al.* 1996);

this modeling has the most stringent position constraints of any published MG J0414-+0534 models. First Ellithorpe tried several different monopole terms, adding either external or mixed quadrupole terms. He also tried an elliptical potential which was not truncated at the quadrupole term. This modeling was done using the "visibility lens clean" (VLC) algorithm which "cleans" or deconvolves the synthesized antenna beam from the data at the same time as tracing the flux back through the lens. Formally all of these models were quite poor fits with $\chi^2/\mathrm{DOF} \geq 1.89$ for 64600+ degrees of freedom.

To explore a wider range of models, Ellithorpe (1995) returned to "point" modeling, using just the image positions and fluxes and the HST galaxy position as constraints. He tried a variety of models. Each model had a singular isothermal sphere monopole term and a quadrupole term. To these he added a second independent quadrupole, a second singular isothermal sphere, an octupole term (angular dependence $\cos 3(\theta - \theta_\beta)$), or a term with angular dependence $\cos 4(\theta - \theta_\delta)$. He found that to get a χ^2/DOF of under about 12 (for 3–6 degrees of freedom) he needed to break the reflection symmetry about the origin of the potential. The models that broke the reflection symmetry were those with, in addition to the monopole and quadrupole terms, either a second singular isothermal sphere or an octupole term. For the two-singular-isothermal-sphere models, this resulted in χ^2/DOF of 5.3 and 12.5 with 3 degrees of freedom, a better fit than any of the rest of the models except for the octupole; the better of the two is only excluded to $\sim$ 99.9%. For the octupole the χ^2/DOF was 1.8 with 4 degrees of freedom, which is only excluded to 88%, a fairly good fit. This best model, with a potential

$$\Phi(r,\theta) = br + \frac{1}{2}\gamma r^2 \cos(2(\theta - \theta_\gamma)) + \beta r^3 \cos(2(\theta - \theta_\beta)), \tag{1.1}$$

was used with VLC which gave only an insignificant change in the fitted parameter values, but gave a χ^2/DOF of 1.69 for 61700 degrees of freedom which is formally very bad (Ellithorpe *et al.* 1996).

Ellithorpe's (1995) 5 GHz VLBI maps were also fit to his octupole model, using

point position constraints, but the data were sufficiently poor not to shed more light on what makes a good model.

In summary, the previous MG J0414+0534 modeling work has shown us the following. The critical radius and angle of orientation of the deflector mass distribution have been well constrained. The modeling work has failed to constrain the radial dependence of the mass distribution. The amount of modeled shear (quadrupole moment) depends on the form assumed for the monopole component. To get an adequate fit to the higher resolution data (10 milliarcsecond position errors) it appears necessary to break the reflection symmetry about the origin. Visibility Lens Clean can demonstrate that a model is formally a poor fit, even if the model adequately satisfies the "point" modeling constraints. Unfortunately, VLC is computationally expensive, which makes it difficult to use to explore wide ranges of models.

In this thesis, high resolution Very Long Baseline Array (VLBA) maps of MG J0414+0534 showing substructure to each of the four images will be used for modeling.

1.4 Overview of this thesis

Chapter 2 introduces some basic concepts of Very Long Baseline Interferometry (VLBI) essential to the reduction of VLBI data to form maps of radio sources.

Chapter 3 describes Very Long Baseline Array (VLBA) 5 GHz observations of seven gravitational lenses, B0218+357, MG J0414+0534, B0957+561, MG J1131+0456, B1422+231, PKS B1830-211, and B2016+112, and gives the details of the data reduction procedure. It presents maps of the five lenses for which maps were successfully made, B0218+357, MG J0414+0534, B0957+561, B1422+231, and PKS B1830-211. These observations were made as part of a study to determine the feasibility of using the VLBA for monitoring the fluxes of these lenses in order to determine their time delays. However, the unusually complex structure of MG J0414+0534 looked so promising that the decision was made to concentrate on modeling the MG J0414+0534 deflector gravitational potential. (See figure 1-1 or 3-6.)

Chapter 4 is an introduction to the theory of gravitational lensing. This chapter

introduces the basic lens equations and concepts in the thin lens approximation, including the two-dimensional lens potential, the differential arrival time delay between images of a background source, and lensing degeneracies — transformations of the two-dimensional lens potential that do not affect the lens image sky brightness distribution. This chapter discusses what observables are available to use in modeling the gravitational potential of the deflector, and mentions some theoretical limits on what can be learned about the potential from lens modeling. It is also shown how the potential, together with time-delay and redshift measurements, can give the distance scale to the lens system — either the angular diameter distance to the lens D_L or the effective lens distance $D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$ depending on what assumptions are made and what observational information about the potential is used. An explanation is given of how the measurement of D_L or $D_{\rm eff}$ for a single lens may be used to determine the Hubble parameter H_o, and how the measurement of D_L or $D_{\rm eff}$ for many lenses with differing source and deflector redshifts may potentially be used to determine the average mass density of the universe Ω_o and the cosmological constant $\Omega_{\Lambda o}$.

Chapter 5 describes in practical detail how features in the lensed emission's surface brightness distribution may be used to constrain models of the two-dimensional lens potential. Formulas are given for "point" modeling, treating each lensed object of a point source, including suitable approximations and analytic tricks to reduce the computation involved in the minimization of the chi-squared. Correction terms are given for use in the case that the lensed objects are slightly extended rather than pointlike, again with attention given to computational efficiency. Using the extents of extended images to provide additional constraints is briefly discussed.

Chapter 6 discusses the selection of a suitable model form with which to parameterize the lens potential. It commends as a suitable model for ring and quad lenses, such as MG J0414+0534, a Taylor expansion in distance from the Einstein ring radius of terms in a multipole expansion in angle of the lens potential (suggested by Kochanek (1991)). Formulas are given for each term in the expansion and for their derivatives as needed in lens modeling. The coefficients of the terms are related to the features of the surface mass density that drive each term, and their effects on

image positions are discussed. This model is compared with other models that have been used for lens modeling.

Chapter 7 presents the results of modeling the deflector potential of the MG J0414+0534 system. Particular attention is given to the degree to which the radial profile of the deflector is constrained. Model predicted dimensionless time delays are computed, for comparison with yet-to-be-measured MG J0414+0534 time delays, for the determination of the Hubble parameter. Attention is drawn to the systematic errors that could result from an improper choice of the form for the deflector potential.

Chapter 8 presents the conclusions of this thesis, summarizing important results from all the chapters. The new 5 GHz VLBA observation of MG J0414+0534 is compared with previous radio maps and with the optical HST maps of Falco *et al.* (1997). The reader is reminded of the advances in modeling techniques and model functional forms presented in this thesis. The best MG J0414+0534 model is presented again, comparing the model lens potential with the optical observations of the lens galaxy. The significance of the modeling result is discussed in reference to its promise for precise determination of the Hubble parameter.

Chapter 2

Very Long Baseline Interferometry

This chapter introduces basic concepts of interferometry, outlining how it is possible to make maps of astronomical sources using Very Long Baseline Interferometry (VLBI). In particular it discusses the reasons for the various transformations and data processing steps that are done to interferometric data in the course of turning electromagnetic waves received at the antenna into maps of the sky brightness distribution.

The two-element interferometer is introduced first, as it is the simplest interferometer and illustrates difficulties that must be overcome to do VLBI. Delay tracking, phase tracking, and correlation are discussed; these are done by the Very Long Baseline Array (VLBA) correlator before the "observer" sees the data. The purpose and effect of fringe fitting are described. Also shown is the manner in which source structure manifests itself in the residual fringe rates and delays. This has implications for the extent to which the data can be averaged. The signals arising from each region of the source, each with its own amplitude, relative phase, residual fringe rate, and residual delay, add together forming the visibility signal (determining amplitude and phase) from which interferometric maps are made. This mapping process is discussed, including Fourier transformation to form "dirty" maps, deconvolution, and hybrid mapping.

VLBI observations, using the VLBA, were made of the gravitational lenses B0218+357, MG J0414+0534, B0957+561, MG J1131+0456, B1422+231, PKS B1830-211,

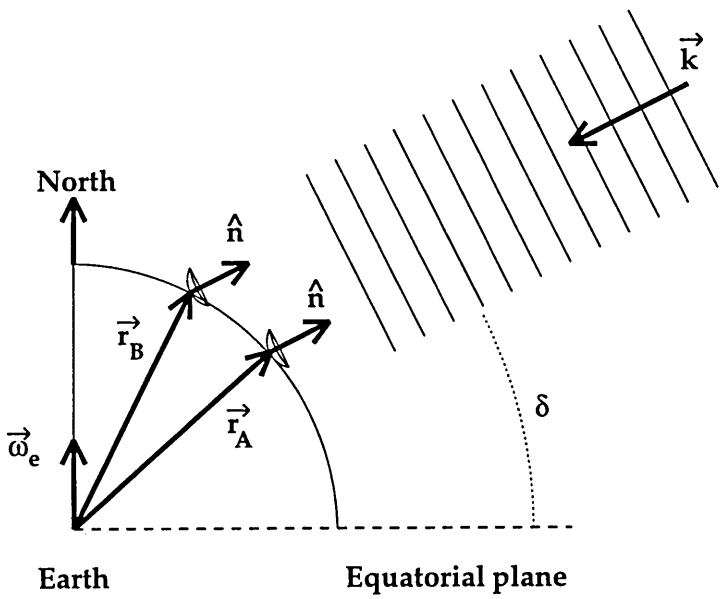

Figure 2-1: The two element interferometer. Electromagnetic radiation originating from a direction $\hat{n}$ is detected by antennas located at $\vec{r}_A$ and $\vec{r}_B$. See text for explanation.

and B2016+112. The observations and the details of the data reduction are described in chapter 3. This chapter provides the conceptual background for understanding the data reduction procedure.

2.1 Simple connected-element interferometer

The discussion of interferometry is most easily begun by considering a simple two-antenna connected-element interferometer. Two antennas, at locations $\vec{r}_A$ and $\vec{r}_B$, each observe the same small patch of sky. (See figure 2-1.) Electromagnetic radiation (noise) originates at a particular point source location, $\hat{n}$, on the sky. The electric field vector is then a plane wave propagating from the direction of the source. Within a narrow frequency band, near frequency ν, the electric field is a nearly monochromatic plane wave whose complex amplitude changes only slowly with time. Using $\omega = 2\pi\nu$ and wave vector $\vec{k} = -\frac{\omega}{c}\hat{n}$, and referring to the complex amplitude (also called the

"envelope function") as $\vec{E}_{\rm env}$, the electric field is

$$\vec{E}(\vec{r},t) = \vec{E}_{\rm env}\left(t - \frac{\vec{k}\cdot\vec{r}}{\omega}\right) e^{i(\vec{k}\cdot\vec{r}-\omega t)}. \tag{2.1}$$

Each antenna detects this electric field wave, plus a certain amount of other noise,

$$\vec{E}(\vec{r},t) = \vec{E}_{\rm env}\left(t - \frac{\vec{k}\cdot\vec{r}}{\omega}\right) e^{i(\vec{k}\cdot\vec{r}-\omega t)} + \vec{n}_{\rm noise}. \tag{2.2}$$

The correlation of the signals from the two antennas is computed: the complex conjugate of one signal is multiplied by the other signal, and the result is averaged over some time interval. This averaging time when taking the correlation must be greater than the inverse of the bandwidth, so that unwanted noise cross terms average to zero.

$$\begin{aligned} &\left\langle \vec{E}^*(\vec{r}_A,t)\vec{E}(\vec{r}_B,t)\right\rangle \\ &= \left\langle \vec{E}^*_{\rm env}(t - \frac{\vec{k}\cdot\vec{r}_A}{\omega})\vec{E}_{\rm env}(t - \frac{\vec{k}\cdot\vec{r}_B}{\omega}) e^{i\vec{k}\cdot(\vec{r}_B-\vec{r}_A)}\right\rangle \\ &\quad + \left\langle \vec{E}^*_{\rm env}(t - \frac{\vec{k}\cdot\vec{r}_A}{\omega}) e^{-i(\vec{k}\cdot\vec{r}_A-\omega t)}\ \vec{n}_{\rm B}\right\rangle \\ &\quad + \left\langle \vec{n}^*_{\rm A}\ \vec{E}_{\rm env}(t - \frac{\vec{k}\cdot\vec{r}_B}{\omega}) e^{i(\vec{k}\cdot\vec{r}_B-\omega t)}\right\rangle \\ &\quad + \langle \vec{n}^*_{\rm A}\ \vec{n}_{\rm B}\rangle\,, \end{aligned} \tag{2.3}$$

$$\left\langle \vec{E}^*(\vec{r}_A,t)\vec{E}(\vec{r}_B,t)\right\rangle = \left\langle \vec{E}^*_{\rm env}(t - \frac{\vec{k}\cdot\vec{r}_A}{\omega})\vec{E}_{\rm env}(t - \frac{\vec{k}\cdot\vec{r}_B}{\omega}) e^{i\vec{k}\cdot(\vec{r}_B-\vec{r}_A)}\right\rangle. \tag{2.4}$$

If the distances from the two antennas to the source differ too much, then the envelope functions fail to line up and the correlated signal is zero; to prevent this, the difference in arrival times of a phase front at the two antennas must be less than the inverse of the bandwidth. For a source near zenith, the phase fronts reach the two antennas at approximately the same time, $\vec{k}\cdot\vec{r}_A \approx \vec{k}\cdot\vec{r}_B$. Therefore the arguments to $\vec{E}^*_{\rm env}$ and

$\vec{E}_{\rm env}$ are the same, and the correlated signal is approximately

$$\left\langle \vec{E}^*(\vec{r}_A,t)\vec{E}(\vec{r}_B,t)\right\rangle \approx \left\langle \vec{E}^*_{\rm env}\vec{E}_{\rm env}\right\rangle e^{i\vec{k}\cdot(\vec{r}_B-\vec{r}_A)}. \tag{2.5}$$

The signal is affected by the relative motion of the antennas as the earth rotates. Both $\vec{k}\cdot\vec{r}_A$ and $\vec{k}\cdot\vec{r}_B$ change with time. The amplitude of the correlated signal is just the flux of the source, but the phase changes with time. (The averaging time when taking the correlation must be short compared to the time over which the phase changes due to the earth's motion.) The phase of the correlated signal is

$$\phi(t) = \vec{k}\cdot(\vec{r}_B - \vec{r}_A). \tag{2.6}$$

The time dependence may be Taylor expanded, letting $\vec{r}_{Bo}$ and $\vec{r}_{Ao}$ be the antenna locations at time t_o. The change in the antenna positions with time is given by the earth's rotation; $\dot{\vec{r}}_B = \vec{\omega}_e \times \vec{r}_B$ and $\dot{\vec{r}}_A = \vec{\omega}_e \times \vec{r}_A$ where $\vec{\omega}_e$ is the earth's rotation rate. Therefore

$$\phi(t) \approx \phi(t_o) - 2\pi(t-t_o)\hat{n}\cdot\left\{\vec{\omega}_e \times \frac{(\vec{r}_{Bo}-\vec{r}_{Ao})}{\lambda}\right\} \tag{2.7}$$

Near to time t_o the phase is thus described by a phase offset $\phi(t_o)$ and a rate of change of phase with time, called the fringe rate:

$$\nu_{\rm fringe} = -\hat{n}\cdot\left\{\vec{\omega}_e \times \frac{(\vec{r}_{Bo}-\vec{r}_{Ao})}{\lambda}\right\} = |\vec{\omega}_e|\cos(\delta)\cos(h)D_{EW}/\lambda \tag{2.8}$$

This fringe rate depends on the source declination δ, the source hour-angle h, the east-west antenna separation D_{EW}, and the earth's rotation rate $\vec{\omega}_e$. These "fringes" are the signature of a source in a two-antenna connected-element interferometer, and can be used to determine the declination of the source. The fringe rate increases with baseline length. For VLBI with, for example, an antenna separation of 6000 km, a wavelength of 6 cm, and a zero declination source at transit, the fringe rate would be 7 kHz !

Several constraints have been noted for a connected element interferometer with

no adjustable delay between antennas. The averaging time must be short compared with the fringe rate $t_{\rm av} < 1/\nu_{\rm fringe}$. The averaging time must be long compared to the bandwidth in order for the noise cross terms to have no correlation: $1/\Delta\nu_{\rm bandwidth} < t_{\rm av}$. The difference in arrival times of the phase front at the two antennas must be small compared with the inverse bandwidth so that the "envelope" functions of the signals are aligned: $|\hat{n} \cdot (\vec{r}_A - \vec{r}_B)|/c < 1/\Delta\nu_{\rm bandwidth}$. Thus for the example above, attempting to do interferometry with a very long baseline, the 7 kHz fringe rate would imply bandwidth of greater than 7 kHz was needed. This in turn would imply that the distance between antennas, projected along the line of sight to the source, would have to be less than 41 km, or that the source be less than 24 arcmin from zenith. Since a source at zero declination travels 15 arcmin for every minute of time, such a source would be visible only briefly to this hypothetical connected-element very-long-baseline interferometer.

2.2 Delay tracking and phase tracking at each antenna

In Very Long Baseline Interferometry, the path-length difference from the two antennas to the source, $(\vec{r}_B - \vec{r}_A) \cdot \hat{n}$, changes rapidly in time. It is necessary to keep the envelope functions aligned in time, so that the correlated amplitude will not be washed out. Therefore one must keep track of the path-length difference and how it changes with time. The path-length difference is also used to undo the fringe rate (the phase-change-with-time) caused by the earth's motion. These processes are known as delay tracking and phase tracking. The phase and delay tracking depend on the source location. This tracking is done by the Very Long Baseline Array correlator before the observer sees the data, using the source position specified by the observer.

The tracking may be done on an antenna-by-antenna basis. A time offset may be introduced before correlation to align the envelope functions. The time is adjusted to be the time of arrival of the wave-front at a fixed point, say the earth's center. The

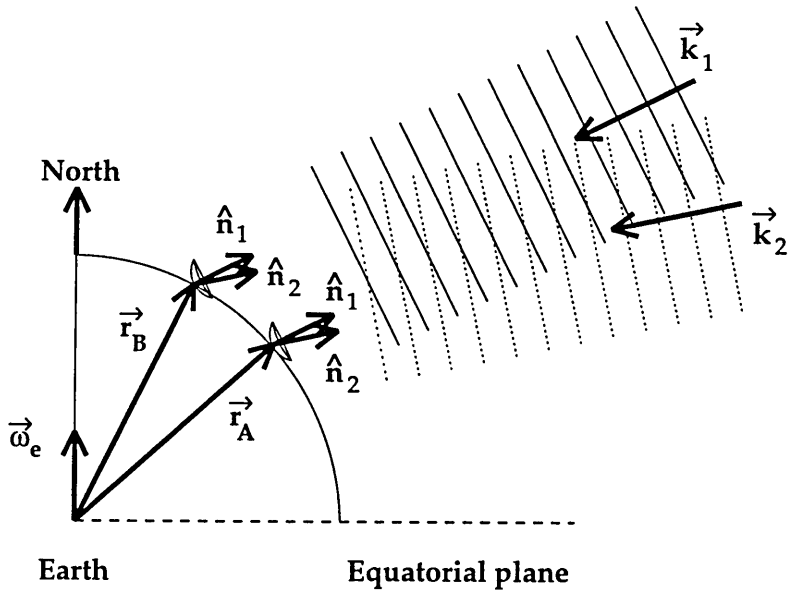

Figure 2-2: Interferometer. See text for explanation.

phase may then be corrected to account for the phase rate caused by the change in the antenna's position.

Consider the following scenario: An array of antennas observes a small region of the sky in a narrow frequency band at frequency $\nu = \omega/(2\pi)$. In that region of sky there are two point sources, in directions $\hat{n}_1$ and $\hat{n}_2$. (See figure 2-2.) The electromagnetic waves which reach the earth from these sources can be described by

$$\vec{E}(\vec{r},t) = \vec{E}_{\text{env}\,1}(t - \frac{\vec{k}_1 \cdot \vec{r}}{\omega})e^{i(\vec{k}_1 \cdot \vec{r} - \omega t)} + \vec{E}_{\text{env}\,2}(t - \frac{\vec{k}_2 \cdot \vec{r}}{\omega})e^{i(\vec{k}_2 \cdot \vec{r} - \omega t)}, \qquad (2.9)$$

where the wave vectors of the radiation are $\vec{k}_1 = -\frac{\omega}{c}\hat{n}_1$ and $\vec{k}_2 = -\frac{\omega}{c}\hat{n}_2$. The envelope functions $\vec{E}_{\text{env}\,1}$ and $\vec{E}_{\text{env}\,2}$ change only slowly with time, since the bandwidth of the detected signal is narrow.

The signal is detected at an antenna A at time t_A and location $\vec{r}_A(t_A)$

$$\vec{E}_A(t_A) = \vec{E}(\vec{r}_A, t_A) = \tag{2.10}$$
$$\vec{E}_{\mathrm{env}\,1}(t_A - \frac{\vec{k}_1 \cdot \vec{r}_A}{\omega}) e^{i(\vec{k}_1 \cdot \vec{r}_A - \omega t_A)} + \vec{E}_{\mathrm{env}\,2}(t_A - \frac{\vec{k}_2 \cdot \vec{r}_A}{\omega}) e^{i(\vec{k}_2 \cdot \vec{r}_A - \omega t_A)}.$$

The signal may then be mixed down to baseband. That is, it may be shifted from a bandpass near frequency $\nu = \omega/(2\pi)$ to a bandpass near zero frequency by multiplying by $e^{i\omega t_A}$.

$$\vec{E}_{\substack{A \\ \mathrm{BB}}}(t_A) = \vec{E}_{\mathrm{env}\,1}(t_A - \frac{\vec{k}_1 \cdot \vec{r}_A}{\omega}) e^{i\vec{k}_1 \cdot \vec{r}_A} + \vec{E}_{\mathrm{env}\,2}(t_A - \frac{\vec{k}_2 \cdot \vec{r}_A}{\omega}) e^{i\vec{k}_2 \cdot \vec{r}_A}. \tag{2.11}$$

The phase may then be corrected for the extra phase due to the displacement of the antenna from a fixed plane perpendicular to the direction to the first source $\hat{n}_1$. For the plane containing the earth's center, the signal is multiplied by $e^{-i\vec{k}_1 \cdot \vec{r}_A}$ There may be some error in the correction, $e^{-i\phi'_{\mathrm{err}\,A}}$. The phase-tracked signal is then:

$$\vec{E}_{\substack{A \\ \mathrm{PT}}}(t_A) = \vec{E}_{\mathrm{env}\,1}(t_A - \frac{\vec{k}_1 \cdot \vec{r}_A}{\omega}) e^{-i\phi'_{\mathrm{err}\,A}(t_A)} + \vec{E}_{\mathrm{env}\,2}(t_A - \frac{\vec{k}_2 \cdot \vec{r}_A}{\omega}) e^{i(\vec{k}_2 - \vec{k}_1) \cdot \vec{r}_A} e^{-i\phi'_{\mathrm{err}\,A}(t_A)}. \tag{2.12}$$

A corrected time may to used to correct for the arrival time at the earth's center of a wave front from the source at $\hat{n}_1$. This arrival time is $t = t_A - \frac{(\vec{k}_1 \cdot \vec{r}_A)}{\omega}$. There may be some error $t_{\mathrm{err}\,A}$ in this correction. The delay-tracked signal is simply the phase-tracked signal at a shifted time, $\vec{E}_{\substack{A \\ \mathrm{PT,DT}}}(t) = \vec{E}_{\substack{A \\ \mathrm{PT}}}(t_A)$ where $t_A = t + \frac{(\vec{k}_1 \cdot \vec{r}_A)}{\omega} - t_{\mathrm{err}\,A}$. It is convenient to rename the phase-tracking error: $\phi_{\mathrm{err}\,A}(t) \equiv \phi'_{\mathrm{err}\,A}(t_A)$. Then

$$\begin{aligned} \vec{E}_{\substack{A \\ \mathrm{PT,DT}}}(t) &= \vec{E}_{\mathrm{env}\,1}(t - t_{\mathrm{err}\,A}) e^{-i\phi_{\mathrm{err}\,A}(t)} \\ &+ \vec{E}_{\mathrm{env}\,2}(t - t_{\mathrm{err}\,A} - \frac{(\vec{k}_2 - \vec{k}_1) \cdot \vec{r}_A}{\omega}) e^{i(\vec{k}_2 - \vec{k}_1) \cdot \vec{r}_A} e^{-i\phi_{\mathrm{err}\,A}(t)}. \end{aligned} \tag{2.13}$$

This delay-tracked and phase-tracked signal may then be correlated with the signals from the other antennas.

2.3 Correlation

The delay-tracked and phase-tracked signal may be split into narrow frequency channels (also called spectral points), each of bandwidth $\Delta\nu$ centered at an observation frequency $\nu + \nu_{\rm chan}$, where ν is the nominal observation frequency used in mixing to baseband and $\nu_{\rm chan}$ is the offset corresponding to a particular channel. Note that $\nu_{\rm chan}$ is the frequency after baseband conversion.

$$\begin{aligned} & \vec{E}_{\substack{A \\ \rm PT,DT}}(t) \\ & = \textstyle\sum_{\nu_{\rm chan}} \ \vec{E}_{\substack{A,\nu_{\rm chan} \\ \rm PT,DT}}(t) \\ & \approx \textstyle\sum_{\nu_{\rm chan}} \ \Big(\vec{E}_{\rm env\,1,\nu_{\rm chan}}(t - t_{{\rm err}\,A})e^{-i\phi_{{\rm err}\,A}(t)} \\ & \qquad + \vec{E}_{\rm env\,2,\nu_{\rm chan}}(t - t_{{\rm err}\,A} - (\vec{k}_2 - \vec{k}_1)\cdot\vec{r}_A/\omega)e^{i(\vec{k}_2-\vec{k}_1)\cdot\vec{r}_A}e^{-i\phi_{{\rm err}\,A}(t)}\Big)\,. \end{aligned} \tag{2.14}$$

The signal in each spectral channel is then cross-correlated with the signal in the same spectral channel from the other antennas. The resulting cross-correlated signal is called the "visibility" on a baseline between antennas A and B,

$$V_{AB}(t,\nu_{\rm chan}) = \left\langle \vec{E}^{*}_{\substack{A,\nu_{\rm chan} \\ \rm PT,DT}}(t)\vec{E}_{\substack{B,\nu_{\rm chan} \\ \rm PT,DT}}(t)\right\rangle . \tag{2.15}$$

The angle brackets denote time averaging over some appropriate interval. This expression (equation 2.15) contains cross terms $\left\langle \vec{E}_{\rm env\,1,\nu_{\rm chan}}\vec{E}_{\rm env\,2,\nu_{\rm chan}}\right\rangle$. These cross terms correlate to zero, since the flux from the two sources is uncorrelated.[1] The averaging time for the correlation $T_{\rm av}$ is taken to be long compared with the variation in the envelope function $\vec{E}_{\rm env\,,\nu_{\rm chan}}$ in this narrow frequency channel, $T_{\rm av} > 1/\Delta\nu$. This is not a problem for 1 MHz frequency channels, as $1/\Delta\nu = 1\ \mu$s. The averaging time must be short with respect to any fringe rate caused by having errors in phase tracking or by having flux from a second region $\hat{n}_2$ away from the tracking center. These require-

[1] If the radio sources are independent of each other then the noise they emit is presumed uncorrelated. If the radio sources are two gravitationally lensed images of a single object, then their fluxes are correlated with a time delay of typically days or months. On the time scales relevant to VLBI observations, their fluxes may be considered uncorrelated with each other.

ments can be met, and they will be discussed in more detail below. The correlated signal is

$$
\begin{aligned}
V_{AB}(t,\nu_{\rm chan}) = & \qquad (2.16)\\
& \left\langle \vec{E}^{*}_{{\rm env}\,1,\nu_{\rm chan}}(t-t_{{\rm err}\,A})\vec{E}_{{\rm env}\,1,\nu_{\rm chan}}(t-t_{{\rm err}\,B})\right\rangle e^{-i(\phi_{{\rm err}\,B}(t)-\phi_{{\rm err}\,A}(t))}\\
+ & \left\langle \vec{E}^{*}_{{\rm env}\,2,\nu_{\rm chan}}(t-t_{{\rm err}\,A}-\frac{(\vec{k}_2-\vec{k}_1)\cdot\vec{r}_A}{\omega})\vec{E}_{{\rm env}\,2,\nu_{\rm chan}}(t-t_{{\rm err}\,B}-\frac{(\vec{k}_2-\vec{k}_1)\cdot\vec{r}_B}{\omega})\right\rangle\\
& \times e^{i(\vec{k}_2-\vec{k}_1)\cdot(\vec{r}_B-\vec{r}_A)}e^{-i(\phi_{{\rm err}\,B}(t)-\phi_{{\rm err}\,A}(t))}.
\end{aligned}
$$

We must calculate the expectation value of the envelope function, multiplied by itself, with a time offset $\tau_{\rm off}$. This is simply the autocorrelation function of the envelope function at a delay $\tau_{\rm off}$. The autocorrelation function is the Fourier transform of the power density spectrum. For continuum sources the power density spectrum is flat across the pass band, which in this case extends from $\nu_{\rm chan}-\Delta\nu/2$ to $\nu_{\rm chan}+\Delta\nu/2$. This power density spectrum may be thought of as a boxcar of width $\Delta\nu$ convolved with a delta function at $\nu_{\rm chan}$. So its Fourier transform, the autocorrelation function, is a sinc function with first zeros at delay $t_{\rm zero} = 1/\Delta\nu$ multiplied by a phase-wrap $e^{i2\pi\nu_{\rm chan}\tau}$. The delays $\tau_{\rm off}$ are assumed to be much smaller than the width of the sinc function, $\tau_{\rm off} \ll t_{\rm zero} = 1/\Delta\nu$ — this can be assured by taking $\Delta\nu$ sufficiently small, i.e. by using enough spectral channels so that each spectral channel is sufficiently narrow. The amplitude of the autocorrelation function at delay $\tau_{\rm off}$ is then the same as that at zero delay. This is simply the power in the signal. Let S be the spectral flux density of the source in energy per unit area per unit time per unit bandwidth. Ignoring the antenna gain we get

$$
\left\langle \vec{E}^{*}_{{\rm env}\,1,\nu_{\rm chan}}(t')\vec{E}_{{\rm env}\,1,\nu_{\rm chan}}(t'-\tau_{\rm off})\right\rangle = \Delta\nu S_1 e^{i2\pi\tau_{\rm off}\nu_{\rm chan}}. \qquad (2.17)
$$

This S is a tensor quantity because the electric field is polarized. Using equation 2.17 to simplify equation 2.16,

$$V_{AB}(t,\nu_{\rm chan}) = \left(\Delta\nu S_1 + \Delta\nu S_2 e^{i2\pi\nu_{\rm chan}\frac{(\vec{k}_2-\vec{k}_1)\cdot(\vec{r}_B-\vec{r}_A)}{\omega}} e^{i(\vec{k}_2-\vec{k}_1)\cdot(\vec{r}_B-\vec{r}_A)}\right) \times e^{-i(\phi_{{\rm err}\,B}(t)-\phi_{{\rm err}\,A}(t))} e^{2\pi i\nu_{\rm chan}(t_{{\rm err}\,B}-t_{{\rm err}\,A})}. \tag{2.18}$$

2.4 "Fringe fitting" to remove residual delay and rate errors

Consider the source at the phase-tracking center $\hat{n}_1$. Its contribution to the visibility signal (equation 2.18) is just the source flux S_1 with a residual phase

$$\phi_{\rm resid} = -(\phi_{{\rm err}\,B}(t) - \phi_{{\rm err}\,A}(t)) + 2\pi\nu_{\rm chan}(t_{{\rm err}\,B} - t_{{\rm err}\,A}). \tag{2.19}$$

This residual phase changes with spectral channel,

$$\frac{\partial\phi_{\rm resid}}{\partial\nu_{\rm chan}} = 2\pi(t_{{\rm err}\,B} - t_{{\rm err}\,A}) \tag{2.20}$$

The residual phase also changes with time. Since

$$\phi_{{\rm err}\,B}(t) - \phi_{{\rm err}\,A}(t) \approx \phi_{{\rm err}\,B}(t_o) - \phi_{{\rm err}\,A}(t_o) + 2\pi(t-t_o)(\nu_{{\rm err}\,B} - \nu_{{\rm err}\,A}) \tag{2.21}$$

then

$$\frac{\partial\phi_{\rm resid}}{\partial t} = 2\pi(\nu_{{\rm err}\,A} - \nu_{{\rm err}\,B}). \tag{2.22}$$

This residual fringe rate, $\nu_{{\rm err}\,B} - \nu_{{\rm err}\,A}$, and delay, $t_{{\rm err}\,B} - t_{{\rm err}\,A}$, can arise for a variety of reasons. A poorly known or erroneous source position may be used for the delay and rate tracking. Alternately, there may be poorly known antenna positions or a poor earth-motion model. Yet another possibility is that there may be a clock

offset at one of the sites.

It is normally desirable to average the data both in frequency and in time before processing, so as to reduce the data volume and the subsequent computational resources. In order to do this, one wants to correct for these residual delays, so that the phases of the visibility data may be aligned across frequency channels, and for these residual rates, so that the phases of the visibility data may be aligned in time. Then the data may be averaged without loss of signal amplitude.

For a well-designed interferometer, using a well-known source position, these residual delays and rates change only slowly with time. The visibility data on a single baseline for some time interval (many minutes) may be fit to find the residual delay and rate. Actually the visibility data are Fourier transformed: time to fringe-rate $t \Longleftrightarrow \nu_{err}$, and frequency channel to delay, $\nu_{\mathrm{chan}} \Longleftrightarrow t_{err}$, and the peak in fringe-rate vs. delay found. This is done for each baseline, and the result is solved for one delay and one rate per antenna. These delays and rates are then used to correct the antenna phases. This process is known as "fringe fitting".

Typically, at least with the VLBA, the fringe rates are small, and the delays change little over the course of a day. However, there may be a large, constant, delay due to clock offsets. This "global delay solution" is found first, by looking at a bright point-like calibrator source, and then applied to all sources. Then the smaller residual delays and the fringe rates are found by fringe fitting separately to each scan of each astronomical target, that is, to each uninterrupted interval of on-source time.

Consider a pair of point sources close to each other, both within the primary antenna beam. The visibility signals from the two sources are additive. Each contribution to the visibility amplitude has its own residual fringe rate and its own residual delay. When searching in the fringe-rate – delay plane, there will be one peak for each source in the field of view. The peak from the brightest source is normally the one found in this fringe fitting process. One can fit for and remove the fringe rate of only one of the sources, typically the dominant source. This corrects the fringe rate of any other source to some extent. However, each secondary source will still have some residual fringe rate and delay that depends on its distance from the first source.

After fringe fitting is done to correct for the fringe rates and delays of a source at $\hat{n}_1$, the phases will have been aligned in time and frequency for a point source at the fringe-fitting center $\hat{n}_1$. The corrected visibility data are then

$$V_{AB}(t, \nu_{\text{chan}}) = \Delta\nu S_1 + \Delta\nu S_2 e^{i2\pi\nu_{\text{chan}} \frac{(\vec{k}_2 - \vec{k}_1)\cdot(\vec{r}_B - \vec{r}_A)}{\omega}} e^{i(\vec{k}_2 - \vec{k}_1)\cdot(\vec{r}_B - \vec{r}_A)}. \tag{2.23}$$

2.5 Residual fringe rates and delays due to source structure

Even after fringe fitting there are residual phase changes with time and phase changes with frequency in the portion of the signal due to flux coming from sources away from the fringe-fitting center. This puts limits on the amount of time and spectral averaging that can be done, but it also gives information on the positions of secondary sources.

2.5.1 Limits on time and spectral averaging

If a source is truly a single point, then the visibility amplitude is constant, independent of antenna locations. Fringe fitting may be used to remove any residual rate-of-change of phase with time or frequency channel due to tracking errors. The amount of time averaging is limited only by the coherence time of the atmosphere. All the spectral channels may be averaged together.

However, if there is an additional source location some distance away from the tracking center, it will still have some residual phase change with spectral channel, and some residual phase change with time. The dependence of the phase on time is found by Taylor expansion,

$$e^{i(\vec{k}_2 - \vec{k}_1)\cdot(\vec{r}_B - \vec{r}_A)} \approx e^{i(\vec{k}_2 - \vec{k}_1)\cdot(\vec{r}_{Bo} - \vec{r}_{Ao})} e^{-2\pi i(t - t_0)\nu_{\text{fringe,resid}}}, \tag{2.24}$$

where

$$\nu_{\text{fringe,resid}} = (\hat{n}_2 - \hat{n}_1) \cdot \vec{\omega}_e \times \frac{(\vec{r}_{Bo} - \vec{r}_{Ao})}{\lambda_{\nu_{\text{chan}}}}. \tag{2.25}$$

The further the source from the tracking center in any particular direction, the faster the phase change with time. For a fixed averaging time, a larger fringe rate means more amplitude loss in the signal from that source. Therefore if one is willing to tolerate a specified amount of loss in the amplitude of a source's signal on a given baseline, then the amount of time averaging that can be done decreases with increasing distance of the source from the tracking center. The reduction in a visibility amplitude due to time averaging is

$$R'_a \approx \frac{\sin(\pi \tau_{\rm av} \nu_{\rm fringe,resid})}{(\pi \tau_{\rm av} \nu_{\rm fringe,resid})}. \tag{2.26}$$

This is approximately

$$R'_a \approx \frac{\sin(\pi \omega_e \tau_{\rm av} r D_{EW}/\lambda)}{(\pi \omega_e \tau_{\rm av} r D_{EW}/\lambda)} \tag{2.27}$$

(equation 6.67 of Thompson *et al.* (1994)), where D_{EW} is east-west baseline distance, λ is the wavelength, r is the distance from the tracking center, ω_e is the earth rotation angular velocity, and $\tau_{\rm av}$ is the averaging time.

Similar considerations apply to bandwidth averaging. The dependence of the phase on frequency channel is

$$e^{-i2\pi \nu_{\rm chan} t_{\rm delay,resid}} \tag{2.28}$$

where

$$t_{\rm delay,resid} = (\hat{n}_2 - \hat{n}_1) \cdot \frac{(\vec{r}_B - \vec{r}_A)}{c}. \tag{2.29}$$

The further the source from the tracking center in any particular direction, the faster the phase change with channel. If one is willing to tolerate a certain specified amount of amplitude loss for a source on a given baseline, then the amount of spectral averaging that can be done decreases with increasing distance of the source from the tracking center. Due to bandwidth averaging, the visibility amplitude of a continuum signal is reduced by a factor

$$R'_b \approx \frac{\sin(\pi \Delta\nu_{\rm av} t_{\rm delay,resid})}{(\pi \Delta\nu_{\rm av} t_{\rm delay,resid})}. \tag{2.30}$$

This is approximately

$$R_b' \approx \frac{\sin(\pi \Delta\nu_{\rm av} r D/c)}{(\pi \Delta\nu_{\rm av} r D/c)} \tag{2.31}$$

(equation 6.59 of Thompson *et al.* (1994)), where D is the baseline length, r is the distance from the tracking center, and $\Delta\nu_{\rm av}$ is the bandwidth over which the averaging is done.

For the flux reduction due to bandwidth smearing and that due to time-average smearing to be roughly comparable in amplitude, the following relation must hold:

$$\tau_{\rm av} = \frac{\Delta\nu_{\rm av}}{\nu_0} \frac{1}{\omega_e}. \tag{2.32}$$

This relation between averaging time and averaging bandwidth is pointed out by Bridle (1989) in his equation 24-A4.

The effect on maps will be discussed below in section 2.9

2.5.2 Information on source positions

The residual fringe rates provide information on the location of sources of flux away from the tracking center. Let $\lambda_{\nu_{\rm chan}}$ be the wavelength corresponding to the center of the spectral channel, then $\lambda_{\nu_{\rm chan}} = c/(\nu + \nu_{\rm chan})$. (Recall that $\nu_{\rm chan}$ is the spectral channel's offset in frequency from the observation frequency ν.) Then the visibility may be written as

$$\begin{aligned} V_{AB}(t, \nu_{\rm chan}) &= \Delta\nu S_1 + \Delta\nu S_2 e^{i(\vec{k}_2 - \vec{k}_1)\cdot(\vec{r}_B - \vec{r}_A)\frac{\nu+\nu_{\rm chan}}{\nu}} \\ &= \Delta\nu S_1 + \Delta\nu S_2 e^{-2\pi i(\hat{n}_2 - \hat{n}_1)\cdot\frac{(\vec{r}_B - \vec{r}_A)}{\lambda_{\nu_{\rm chan}}}}. \end{aligned} \tag{2.33}$$

The time dependence of the visibility may be shown by Taylor expanding the dependence on time of the baseline $\vec{r}_B - \vec{r}_A$.

$$V_{AB}(t, \nu_{\rm chan}) = \Delta\nu S_1 + \Delta\nu S_2 e^{-2\pi i(\hat{n}_2 - \hat{n}_1)\cdot\frac{(\vec{r}_{Bo} - \vec{r}_{Ao})}{\lambda_{\nu_{\rm chan}}}} e^{-2\pi i(t - t_o)\nu_{{\rm fringe}\,AB}}, \tag{2.34}$$

where the fringe rate is

$$\nu_{\text{fringe}\,AB} = (\hat{n}_2 - \hat{n}_1) \cdot \vec{\omega}_e \times \frac{(\vec{r}_{Bo} - \vec{r}_{Ao})}{\lambda_{\nu_{\text{chan}}}}. \tag{2.35}$$

Any source at a location $\hat{n}_2$ displaced from the tracking center $\hat{n}_1$ has a residual fringe rate on each baseline. If one Fourier transforms from time to fringe rate, $t \Longleftrightarrow \nu_{\text{fringe}\,AB}$, there should be a spike for each source, and the strength of the spike should give an idea of the strength of the source. Such "fringe-rate plots" are an easy but useful diagnostic of the quality of the observation, to see if the signal to noise ratio was high. Single components appear as single spikes. Multiple-component lenses appear as multiple spikes, to some extent confused with each other.

Notice that the fringe rate for a source depends on the displacement of the source from the fringe-fitting center, $(\hat{n}_2 - \hat{n}_1) = \psi_x\hat{x} + \psi_y\hat{y}$, and on the baseline vector $\vec{b}_{AB} = \frac{(\vec{r}_B - \vec{r}_A)}{\lambda}$:

$$\nu_{\text{fringe}\,AB} = (\psi_x\hat{x} + \psi_y\hat{y}) \cdot \vec{\omega}_e \times \vec{b}_{AB}. \tag{2.36}$$

A constant value of $\nu_{\text{fringe}\,AB}$ constrains the source location, $\psi_x\hat{x} + \psi_y\hat{y}$, to lie on a line on the sky. Different baselines constrain the source to different lines on the sky. The intersection of these lines gives the source location. Plots of such lines are known as "fringe-rate maps". (Fringe-rate mapping is discussed in section 12.5 of Thompson *et al.* (1994).) For astronomical sources with multiple distinct sources, fringe rate maps are useful for locating the approximate locations of the sources. Mapping, which will be described below, can only be done in fairly small fields, due to limited computational resources. It is good to know in advance where to put those fields relative to the fringe-fitting center.[2]

[2]In general, one will know from lower resolution observations of the source, for example with the Very Large Array (VLA), what components to expect. Occasionally, fringe fitting (see section 2.4) will find and remove the fringe-rate effects of a different component than that expected to be dominant. If the fringe-fitting center is actually at component "b" and one assumes that it is at component "a" then map field locations chosen based on the knowledge of the relative component locations will fall in the wrong places. Fringe-rate maps, which are computationally very quick, let one verify approximate component locations and avoid such possible problems.

2.6 Mapping with an interferometer

Consider again the visibility signal,

$$V_{AB}(t, \nu_{\text{chan}}) = \Delta\nu S_1 + \Delta\nu S_2 e^{-2\pi i(\hat{n}_2 - \hat{n}_1)\cdot\frac{(\vec{r}_B - \vec{r}_A)}{\lambda_{\nu_{\text{chan}}}}}. \tag{2.37}$$

Note that its only dependence on time, on frequency channel, and on antenna is through the baseline vector:

$$\vec{b}(A, B, t, \nu_{\text{chan}}) = \frac{(\vec{r}_B - \vec{r}_A)}{\lambda_{\nu_{\text{chan}}}}. \tag{2.38}$$

Then we may write

$$V(\vec{b}) = \Delta\nu S_1 + \Delta\nu S_2 e^{-2\pi i(\hat{n}_2 - \hat{n}_1)\cdot\vec{b}}. \tag{2.39}$$

This may be readily generalized to many sources. As before all the cross terms between sources correlate to zero, so that

$$V(\vec{b}) = \Delta\nu \sum_k S_k e^{-2\pi i(\hat{n}_k - \hat{n}_1)\cdot\vec{b}}. \tag{2.40}$$

For a continuum of sky brightness, the individual source fluxes S_k may replaced by the sky brightness $I(\hat{n})$, and the sum turned into an integral:

$$\sum_k S_k f(\hat{n}_k) \longrightarrow \int d\Omega I(\hat{n}) f(\hat{n}). \tag{2.41}$$

Equation 2.40 may be re-written as

$$V(\vec{b}) = \Delta\nu \int d\Omega I(\hat{n}) e^{-2\pi i(\hat{n} - \hat{n}_1)\cdot\vec{b}}. \tag{2.42}$$

For clarity in understanding the approximations that will be made, it is convenient to introduce a coordinate system, $\hat{u}, \hat{v}, \hat{w}$, aligned with East, with North, and with the direction to the tracking center. (See figure 2-3.) The basis vector $\hat{w}$ is chosen to point towards the tracking center $\hat{n}_1$, so that $\hat{w} = \hat{n}_1$. The basis vector $\hat{u}$ points east, and lies in the equatorial plane of the earth. The third basis vector $\hat{v}$ is chosen

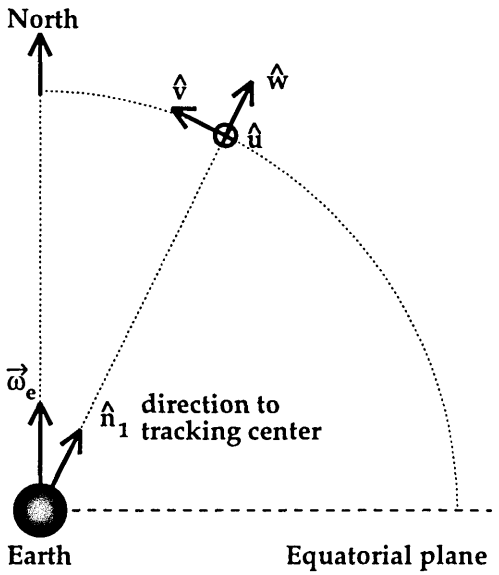

Figure 2-3: The $\hat{u}, \hat{v}, \hat{w}$ coordinate system

to make a right-handed orthonormal coordinate system, $\hat{v} = \hat{w} \times \hat{u}$. Note that $\hat{u}$ and $\hat{v}$ define a plane perpendicular to the direction to the source. They point eastward and northward, respectively, in that plane.

All the quantities in equation 2.42 may be re-written in this $\hat{u}, \hat{v}, \hat{w}$ coordinate system. The components of the baseline vector $\vec{b}$ are given the names u, v, and w: $u = \hat{u} \cdot \vec{b}$, $v = \hat{v} \cdot \vec{b}$, and $w = \hat{w} \cdot \vec{b}$, so that $\vec{b} = u\hat{u} + v\hat{v} + w\hat{w}$. Since the direction to the tracking center, $\hat{n}_1$, was used to define the coordinate system, $\hat{n}_1 = \hat{w}$ by construction. The direction to an arbitrary location on the celestial sphere, $\hat{n}$, is specified by the direction cosines l and m, where $l = \hat{u} \cdot \hat{n}$ and $m = \hat{v} \cdot \hat{n}$. Only two numbers are needed to specify $\hat{n}$, since $\hat{n}$ is a unit vector, but note that each (l, m) corresponds to two different unit vectors $\hat{n}$, one on the same side of the earth at the tracking center $\hat{n}_1$, and one on the opposite side: $\hat{n} = l\hat{u} + m\hat{v} \pm \hat{w}\sqrt{1 - l^2 - m^2}$. The angular area element $d\Omega$, representing the surface area of a unit sphere at location $\hat{n}$, must also be written in terms of the direction cosines l and m. The algebra is done in the footnote

below, giving[3] $d\Omega = \frac{dl\ dm}{\sqrt{1-l^2-m^2}}$. Therefore equation 2.42 may be re-written as

$$V(u,v,w) = \tag{2.43}$$
$$\Delta\nu \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dl\ dm}{\sqrt{1-l^2-m^2}} I(l\hat{u}+m\hat{v}+\hat{w}\sqrt{1-l^2-m^2}) e^{-2\pi i(ul+vm+w(-1+\sqrt{1-l^2-m^2}))}$$
$$+\ \Delta\nu \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dl\ dm}{\sqrt{1-l^2-m^2}} I(l\hat{u}+m\hat{v}-\hat{w}\sqrt{1-l^2-m^2}) e^{-2\pi i(ul+vm+w(-1-\sqrt{1-l^2-m^2}))}.$$

The interferometer antennas are directional and do not detect flux from directions more than 90° away from the tracking center. Therefore for these directions, $I(l\hat{u} + m\hat{v} - \hat{w}\sqrt{1-l^2-m^2}) = 0$. Then without risk of ambiguity we may write $I(l,m) \equiv I(l\hat{u} + m\hat{v} + \hat{w}\sqrt{1-l^2-m^2})$. Therefore equation 2.42 may be re-written as

$$V(u,v,w) = \Delta\nu \int\int \frac{dl\ dm}{\sqrt{1-l^2-m^2}} I(l,m) e^{-2\pi i(ul+vm+w(-1+\sqrt{1-l^2-m^2}))}. \tag{2.44}$$

(equation 2-21 of Thompson (1989)).

When the flux originates in only a limited region of the sky, the direction cosines l and m are approximately the distance, in radians, from the tracking center: $l = \hat{u}\cdot\hat{n} \approx \psi_x$ and $m = \hat{v}\cdot\hat{n} \approx \psi_y$. The surface area element $\frac{dl\ dm}{\sqrt{1-l^2-m^2}}$ may be approximated by simply $dl\ dm$ with less than 1% error for $\frac{1}{\sqrt{1-l^2-m^2}} < 1.01$ or $\sqrt{1-l^2-m^2} < 0.14$ radians = 8 degrees. Since this is much larger than the primary beam of the VLBA interferometer antennas, this approximation can be made.

The visibility $V(u,v,w)$ depends on the coordinate w only through the phase term

---

[3] Any integral over the surface of a sphere may be written as an integral over a three-dimensional volume, using a delta function to force the contribution of the integrand to be zero except on the surface of the sphere, $I = \int d\Omega f(\hat{n}) = \int dV \delta(|\vec{n}|-1) f(\vec{n})$. Let the components of $\vec{n}$ be represented by (l,m,n), then $I = \int_{-\infty}^{\infty} dl \int_{-\infty}^{\infty} dm \int_{-\infty}^{\infty} dn \delta(\sqrt{l^2+m^2+n^2}-1) f(l,m,n)$. For any function $g(x)$ the delta function $\delta(g(x))$ may be written as a sum of terms, one term for each zero of $g(x)$. Let x_i be the zeros of $g(x)$, then $\delta(g(x)) = \sum_i \frac{\delta(x-x_i)}{|g'(x_i)|}$. Therefore $\delta(\sqrt{l^2+m^2+n^2}-1) = \frac{\delta(n-\sqrt{1-l^2-m^2})}{\sqrt{1-l^2-m^2}} + \frac{\delta(n+\sqrt{1-l^2-m^2})}{\sqrt{1-l^2-m^2}}$. The delta function may be used to turn the triple integral I into a double integral.

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dl\ dm}{\sqrt{1-l^2-m^2}} f(l,m,\sqrt{1-l^2-m^2}) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dl\ dm}{\sqrt{1-l^2-m^2}} f(l,m,-\sqrt{1-l^2-m^2}).$$

Comparing this with $I = \int d\Omega f(\hat{n})$, one identifies $d\Omega = \frac{dl\ dm}{\sqrt{1-l^2-m^2}}$ and notes that each (l,m) corresponds to two unit vectors, $\hat{n} = (l,m,\sqrt{1-l^2-m^2})$ and $\hat{n} = (l,m,-\sqrt{1-l^2-m^2})$.

$e^{-2\pi iw(-1+\sqrt{1-l^2-m^2})}$. Under certain conditions this phase term is close to unity and may be ignored. An interferometer antenna with a narrow beam detects flux from only a small patch of sky. Therefore $I(l,m)$ is non-zero only for direction cosines l and m which are small. Therefore it is useful to do the expansion

$$e^{-2\pi iw(-1+\sqrt{1-l^2-m^2})} = e^{\pi iw(l^2+m^2+\mathcal{O}((l^2+m^2)^2))} \tag{2.45}$$

It can readily be seen that this phase is negligible for $w(l^2+m^2) \ll 1$. Note that $w = \hat{w} \cdot \vec{b} = \hat{n}_1 \cdot (\vec{r}_B - \vec{r}_A)/\lambda_{\nu_{\text{chan}}}$ is the projection of the baseline vector along the line of sight to the source, as measured in wavelengths. The phase contributed by the w-term in equation 2.44 is negligible if the extent of the source's flux (in radians) $\sqrt{l^2+m^2}$ as measured relative to the tracking center is less than $1/\sqrt{w}$. More quantitatively, requiring that $\text{Re}(e^{\pi iw(l^2+m^2)}) > 0.99$ requires that $w(l^2+m^2) < 0.045$, that is $\sqrt{l^2+m^2} < \sqrt{0.045}\sqrt{\frac{\lambda_{\nu_{\text{chan}}}}{|\hat{n}_1 \cdot (\vec{r}_B - \vec{r}_A)|}}$. For a 6000 km baseline and observations at 6 cm wavelength, this requirement is $\sqrt{l^2+m^2} < \sqrt{0.045}\sqrt{\frac{6cm}{6000km}} = 2.1 \times 10^{-5}$radians = 4.4 arcsec. It can be seen that this w-dependent phase term is negligible for sources of size 4 arcsec and smaller.

Using these approximations, equation 2.44 may be re-written as

$$V(u,v) = \Delta\nu \int dl\, dm\, I(l,m) e^{-2\pi i(ul+vm)} \tag{2.46}$$

where $V(u,v,w)$ has been re-written as $V(u,v)$ since the right-hand side of the expression no longer has any dependence on w.

For astronomical sources with flux originating in only a limited region of sky, the direction cosines may be thought of as angles on the sky: $l = \hat{u} \cdot \hat{n} \approx \psi_x$ and $m = \hat{v} \cdot \hat{n} \approx \psi_y$. For such small angles it is more intuitive to write equation 2.46 in terms of these angles

$$V(u,v) = \Delta\nu \int d\psi_x d\psi_y I(\psi_x, \psi_y) e^{-2\pi i(u\psi_x + v\psi_y)} \tag{2.47}$$

(Clark 1989)[4]

This relation is the basis for all mapping done using interferometry. It is the Fourier transform from (ψ_x, ψ_y) to (u,v). Inverting the Fourier transform and ignoring the absolute normalization, one obtains an expression for the sky brightness,

$$I(\psi_x, \psi_y) = \int du\, dv\, V(u,v) e^{2\pi i(\psi_x u + \psi_y v)}. \tag{2.48}$$

Thus, conceptually, it is very simple to make a map of the source. Simply measure $V(u,v)$ for every possible baseline (u,v), then Fourier transform $V(u,v)$. Various practical difficulties are discussed below.

2.7 Limited UV coverage, dirty maps, deconvolution

There are many complications to making interferometer maps. One complication is that the visibilities $V(u,v)$ are measured only on a finite number of baselines, since an interferometer has a finite number of antennas. Each baseline traces out an arc in the u-v plane as the earth rotates, thus increasing the coverage of the u-v plane.

[4]Starting from equation 2.42 one may reach equation 2.47 in one step if one does not insist on quantifying how small the flux-emitting region must be in order for the approximation to be valid. Using interferometer antennas with a narrow beam, flux is detected from only a small patch of sky. The displacement of $\hat{n}$ from the pointing center $\hat{n}_1$ lies in a plane, $\hat{n} - \hat{n}_1 = \psi_x \hat{u} + \psi_y \hat{v}$. Therefore the component of the baseline vector along the line of site to the source becomes irrelevant. Only relevant is the projection of the baseline vector onto a plane perpendicular to the direction to the pointing center $\hat{n}_1$. This plane is called the "u-v plane". Define $u = \hat{u} \cdot \vec{b}$ and $v = \hat{v} \cdot \vec{b}$. Then equation 2.47 follows immediately.

Still, the visibility has only been measured on a very small portion of the u-v plane. In practice one constructs the "Direct Fourier Transform", $I^D(\psi_x, \psi_y)$. This treats $V(u,v)$ as zero where it is not measured. The sampling is denoted by a sampling function $S(u,v)$, a sum of (possibly weighted) delta functions at the sample points. Then, (Clark 1989)

$$\begin{aligned} I^D(\psi_x, \psi_y) &= \sum_k V(u_k, v_k) e^{2\pi i(\psi_x u_k + \psi_y v_k)} \\ &= \int du\, dv\, S(u,v) V(u,v) e^{2\pi i(\psi_x u + \psi_y v)}. \end{aligned} \tag{2.49}$$

This is recognizable as a convolution of the visibility function $V(u,v)$ with another function that is called the "beam pattern" or the "dirty beam" and denoted by $B(\psi_x, \psi_y)$. Let

$$\begin{aligned} B(\psi_x, \psi_y) &= \int du\, dv\, S(u,v) e^{2\pi i(\psi_x u + \psi_y v)} \\ &= \sum_k e^{2\pi i(\psi_x u_k + \psi_y v_k)}. \end{aligned} \tag{2.50}$$

Then

$$I^D(\psi_x, \psi_y) = B(\psi_x, \psi_y) * I(\psi_x, \psi_y). \tag{2.51}$$

This direct-Fourier-transformed brightness distribution $I^D(\psi_x, \psi_y)$ is known as the "dirty" map. This dirty map is the true brightness distribution convolved with a beam pattern. Note that the beam pattern is the response to a point source, since for a point source $V_\nu(u,v) = 1$. Much effort is required to deconvolve the beam pattern from the brightness distribution.

Since information on the visibilities is just plain missing from many regions of the u-v plane, any deconvolution is not unique. Any flux distribution that would show up only in the un-sampled regions of the u-v plane is not constrained by the data and could be present in any quantity.

One has no information about the visibility amplitudes for u-v distances beyond D_{max}/λ. All very-high spatial frequency information is missing. This limits the resolution of the final map. Likewise, there is some minimum antenna separation

D_{min}, so one has no information about the visibility amplitudes for u-v distances below D_{min}/λ. All very-low spatial frequency information is missing. Extended regions of slowly changing flux would show up only in the low spatial frequencies. The interferometer cannot see them. These low-spatial-frequency modes are assumed zero. This flux does not show up in the maps, and is said to be "resolved out".

However, one can make use of some *a priori* assumptions to constrain the deconvolution process. In particular, all flux is assumed to come from a limited region of the sky.

The deconvolution method used in this thesis is the Clark-Cotton-Schwab CLEAN algorithm as implemented in the Astronomical Image Processing System (Hogbom 1974; Clark 1980; Schwab 1984). Search boxes, called "clean boxes", are set on the dirty map to bound the regions in which it is thought that there is real flux. The region of the dirty map within the clean boxes is searched to find the point with the highest amplitude. Some small fraction of the beam pattern is subtracted from that location, and the process is repeated. This results in a model of the source brightness distribution that consists of point components at map pixel locations. After cleaning has progressed for the desired number of iterations, these clean components are convolved with a Gaussian of the same width as the central lobe of the beam pattern. These residual Gaussians are added to any residual un-cleaned flux to make the "cleaned" map of the source. A helpful introduction to deconvolution is given by Cornwell & Braun (1989).

2.8 Phase calibration

Interferometric mapping is complicated by various effects that can produce phase errors in the signals at the antennas. There may be phase offsets due to the atmosphere; these remain constant only over short periods of time. (One cannot time average for longer than the atmospheric coherence time, but fringe rates from extended source structure were more of a constraint for the sources observed in this thesis.) There may also be phase offsets introduced by the hardware. The visibility phases must be

calibrated. The solution interval should be not longer than the atmospheric coherence time.

One allows for one offset per antenna per solution interval, $\phi_{err_i}(t)$. For N antennas there are N phase offsets in each solution interval. However, N antennas have $N(N-1)/2$ baselines. So there are $N(N-1)/2$ observational phases in each solution interval. The number of measured phases minus the number of unknown offsets is $N(N-3)/2$. As long as the interferometer has more than three antennas, not all the phase information is lost.

To calibrate the phases, a model is made of the source. The phase offsets (one per antenna per solution interval) which best fit the source model are then found. A map is made from the data, and a better model constructed from the map. This new model is again used to calibrate the data. A new map is made — and so on. The model improves with each iteration, until it converges to some model consistent with the data. The process of calibrating the data to a model made from the source is known as self-calibration. The whole repeated self-calibration and mapping procedure is known as hybrid-mapping. A good introduction is given by Pearson & Readhead (1984). Hybrid-mapping is helpfully discussed with reference to the VLBA by Walker (1995).

2.9 Time and bandwidth averaging limits for maps

The signal reduction on a single baseline, for a source displaced from the tracking center, due to time averaging and spectral averaging, was discussed above in section 2.5.1. More important is the net effect of time averaging and spectral averaging on a map, rather than on a single baseline at a single time. The analysis becomes more complicated when one considers multiple baselines rather than single baselines. The signal reduction effect depends on the orientation of each baseline relative to the source displacement — and these orientations change over the course of an observation. At any point in time, some baselines to a given antenna are more affected than other baselines to that antenna. The effect on a map made from interferometer data has been analyzed in detail by Thompson *et al.* (1994). For a many-antenna

interferometer, using reasonable assumptions, they calculated the peak response to a point source a distance r from the origin as a fraction R_b of the response to the same source at the origin,

$$R_b = 1.064\frac{\theta_b\nu_0}{r\Delta\nu_{\rm av}}{\rm erf}(0.833\frac{{\rm r}\Delta\nu_{\rm av}}{\theta_{\rm b}\nu_0}) \tag{2.52}$$

(equation 6.64 of Thompson *et al.* (1994)). They point out that the result is not very dependent on bandpass shape. The form is simpler for a Gaussian bandpass of equivalent width $\Delta\nu_{\rm av}$,

$$R_b = \frac{1}{\sqrt{1+(0.939\frac{r_1\Delta\nu_{\rm av}}{\theta_b\nu_0})^2}} \tag{2.53}$$

(equation 6.65 of Thompson *et al.* (1994)), where θ_b is the antenna beam $\theta_b \approx \frac{c}{D_{max}\nu_0}$, ν_0 is the observation frequency, and $\Delta\nu_{\rm av}$ is the bandwidth.

2.10 The Very Long Baseline Array

The observations described in this thesis were made with the Very Long Baseline Array (VLBA) of the National Radio Astronomy Observatory (NRAO). The VLBA is an array of ten identical antennas devoted exclusively to VLBI. This makes observations much more straightforward for the observer than dealing with disparate antennas controlled by more than one observatory.

The VLBA has ten antennas, each 25 m in diameter, located on U.S. territory. The longest baseline length is 8600 km, from Mauna Kea, HI to St. Croix, VI. The VLBA has nine frequency bands for observations, between 300 MHz and 45 GHz, with system temperatures of around 30 K at 5 GHz (the frequency used in this thesis). Correlation of the observations is done at the VLBA correlator in Socorro, New Mexico.

This and more detailed information can be found in the VLBA Observational Status Summary (Wrobel 1996) and the NRAO summer school proceedings edited by Zensus *et al.* (1995).

Chapter 3

VLBA Observations and Data Reduction

Seven gravitational lenses were observed with the Very Long Baseline Array (VLBA) at two frequencies at three epochs. The original purpose of the observation was to check for the suitability of using the VLBA to monitor the fluxes of the sources — with the goal of measuring the time delays of these lenses. While reducing the 5 GHz data from the first epoch, it was found that the gravitational lens MG J0414+0534 displays substructure on milliarcsecond scales, which could be used for modeling the gravitational potential of its deflector.

3.1 Summary of observations

The gravitational lenses B0218+357, MG J0414+0534, B0957+561, MG J1131+0456, B1422+231, PKS B1830-211, and B2016+112 were observed with the Very Long Baseline Array. The observations, made 7-8 July, 15-16 September, and 25-26 November 1995, were at 6 cm and 3.6 cm wavelength. At each frequency, each source was observed for 7 or 8 scans of approximately 6 minutes each, with the total recorded time being limited by the tape allotment. To the extent possible, these scans were spaced evenly over the time that the source was up during the 20 hour observation, to cover as much of the u-v plane as possible. The calibrator sources J1310+3220 and

J1407+2827 were observed for approximately 3 minutes at each frequency, for use as fringe-finder and amplitude-check sources. 3C84 was observed 5 times at varying parallactic angle, 3 minutes at each frequency, to use for polarization calibration. The total bandwidth was 32 MHz, 16 MHz each in right and left circular polarization. Due to concern about loss of peak flux from bandwidth smearing, the observations were correlated on multiple passes, with different correlation centers.

The 6 cm data from the July observation, correlated at the location of the A or A1 image of each lens, has been reduced using the 15JUL95 and 15JAN96 versions of the Astronomical Image Processing System (AIPS) of the National Radio Astronomy Observatory (NRAO). Maps were made of the lenses B0218+357, MG J0414+0534, B0957+561, B1422+231, and PKS B1830-211. The lenses MG J1131+0456 and B2016+112 did not have high enough signal-to-noise to fringe fit and map.

3.2 Data reduction procedure for all program sources

After correlation, the resulting "visibility" data (see equation 2.18) requires further processing to produce maps of the astronomical sources. The antenna amplitudes must be calibrated. "Fringe fitting" (section 2.4) removes the residual phase rate and delay errors in the visibility data. The data are averaged in time and over frequency channels to reduce the data volume and speed further processing, but the extents of the sources limit how much averaging can be done (sections 2.5.1 and 2.9). The Fourier transformation of the visibility data produces a map of the sky brightness (section 2.6) but the beam pattern due to limited u-v coverage must be deconvolved from the map (section 2.7) and the antenna phases must be calibrated (section 2.8). The hybrid mapping process (section 2.8) in which the clean component model of the deconvolved map is used to calibrate the antenna phases, and the phase-calibrated visibility data are used to produce a deconvolved map, requires a number of iterations to converge. The clean component model from the map resulting from the hybrid

mapping is used as a more accurate model for fringe fitting, and the mapping process is repeated.

Amplitude calibration and the correction for delays due to clock offsets were done for all calibrator and program sources together. Fringe fitting, time- and frequency-averaging of the visibility data, and hybrid mapping were done for each program source separately. No amplitude self-calibration was done to the visibility data. The data reduction technique used for all the program sources is described in this section. Special treatment required for the individual gravitational lenses is described, and the resultant maps displayed, in section 3.3. An improved data reduction procedure was used to redo the reduction of the MG J0414+0534 data, and this is described in section 3.4.

Amplitude calibration

The antenna system temperatures were used for calibration. A discrepancy was noted: the left-circularly-polarized flux amplitudes were approximately 85% of the right-circularly-polarized flux amplitudes on all baselines for both amplitude-check sources J1310+3220 and J1407+2827. Such an effect can be caused by errors in the voltage thresholds in the digital samplers (National Radio Astronomy Observatory 1995), but no correction was applied for this until the second pass of data reduction (section 3.4).

Global delay solution for each antenna

Correction must be made for clock offsets at each antenna (section 2.4). A bright point-like calibrator source, a "fringe-finder source" is useful for finding this "global delay solution". Plots of the amplitude versus u-v distance revealed J1310+3220 as the most point-like of the calibrators. A single delay solution was found for each antenna, each intermediate frequency (IF), each polarization using J1310+3220 as a fringe finder source. These delay solutions were applied to all sources.

Flagging of the data

Certain data can be recognized as bad, in particular anomalies that are tied to a particular antenna over a particular time range, since signal variations due to the source should depend on the baseline length and direction, not on the antenna or the time. Such anomalies include low amplitudes at the start of scan or high amplitudes simultaneously on all baselines to a single antenna, indicating antenna problems or interference. Such data were "flagged" and thereby eliminated. It was found best to look at how the visibility amplitudes varied with baseline and time, for a few representative spectral channels. It was assumed that if one spectral channel was bad, then all spectral channels were bad.

Examination of program sources using fringe-rate plots and fringe-rate maps

Fringe-rate plots (section 2.5.2) were used to check whether the signal to noise ratio was high. These plots gave indication of how easy fringe fitting the data would be, and even whether fringe fitting should be attempted. Fringe-rate maps (section 2.5.2) were made to verify the location of the lens images or at least the brightest of the lens images.

Fringe fitting

Residual errors in the delays and rates used to track to sources must be removed (see section 2.4) leaving only the phase rates and delays due to source structure. The process of determining and removing these delay and rate errors is known as "fringe fitting".

For observations of multi-component astronomical objects, such as these multi-image lenses, each lensed image (A, B, etc.) has its own residual phase rates and delays. When searching in the delay-and-rate space for the source signal, a secondary source may confuse the search if its amplitude is close to that of the primary source.

One technique is to divide out the expected effects of the source structure using

a model of the source, as a preliminary to fringe fitting. This was tried for the first two lenses mapped, B0218+357 and PKS B1830-211, using models consisting of one point source for each lens image, constructed using the approximate positions and fluxes for the images from published maps. With PKS B1830-211, it was found that fringe fitting assuming a point source worked better than fringe fitting using a poor multi-component model. Therefore point models were used for each of the remaining lenses.

When fringe fitting a multi-component astronomical object, assuming a point-source object, one must make sure that for each solution time-interval, the same source (the primary, not the secondary or a noise spike) is found in the delay-rate search. Since each source had seven or eight scans over the 20 hour observation, it could be seen, for each individual source, how the delay and rate solutions varied with time. It was clear from the brighter sources that the delay and rate tracking errors generally changed only gradually over the multi-hour observation. Therefore the criterion that the delay and rate solutions must change smoothly with time from scan to scan could be used to ensure that solutions for the same component were being found for each scan. Various controls are available to control the fringe search process: the reference antenna, the required signal-to-noise ratio of the solutions, the delay and rate search intervals, limits on the u-v range, and the time interval over which to find a solution. When different adjustments to the fringe-search controls found different delay and rate "solutions", the set of solutions that changed most smoothly with time was assumed to correspond to a single component of the lens.

The fitted delays and fitted rates were generally quite small. Apart from PKS B1830-211, the fitted rates were under 5 mHz and the fitted delays were under 15 nanoseconds. With such good *a priori* delay and rate calibration from the VLBA, one can use quite narrow search windows when fringe fitting, thereby reducing the chance of finding noise spikes or secondary components. The one lens which had higher delays and rates was PKS B1830-211, which had been correlated at the location of the northeast image, A. Fringe fitting with a point model found delay and rate solutions for the southwest image B which had higher peak and integral flux.

Source	Correlation center	Distance to furthest component (arcsec)	Averaging Bandwidth		Averaging Time	
			For ~1% peak flux reduction (MHz)	Actual (MHz)	For ~1% peak flux reduction (s)	Actual (s)
B0218+357	A	0.33	3.3	2.0	9.1	8.0
PKS B1830-211	A	0.97	1.1	1.0	3.1	2.0
B1422+231	A	1.3	0.8	0.5	2.3	2.0
MG J0414+0534	A1	2.0	0.54	0.5	1.5	2.0
B0957+561	A	6.2	0.18	0.5	0.5	2.0

Table 3.1: Time-averaging and spectral-averaging of the visibility data of the program sources. The reductions in peak flux are given for the lens image furthest from the correlation center

Time averaging and spectral averaging

Time averaging and spectral averaging are used to reduce the data volume and therefore the computing resources needed. However, there are limits relating the extent of the sky brightness distribution to the amount of time and spectral averaging that can be done without degrading the resulting maps. (These limits are discussed in sections 2.5.1 and 2.9 of this thesis.)

The data were recorded with 2 s averaging, so the averaging time could be any multiple of 2 s. Each frequency band of 8 MHz was recorded in 16 spectral channels of 0.5 MHz each, therefore the bandwidth over which to average could be 0.5, 1, 2, 4, or 8 MHz. A formula for the reduction in peak flux for a source located some distance away from the correlation and fringe-fitting center is given in equation 2.53. The averaging time that gives a peak flux reduction similar to that caused by the bandwidth averaging is given in equation 2.32. In table 3.1 are listed the averaging intervals permitted by the requirement that neither time averaging nor spectral averaging produce more than a 1% drop in the peak flux of the component most distant from the correlation center. The actual averaging intervals used are also listed in table 3.1. The two images of B0957+561 are so widely separated that even without further averaging, B0957+561 image B suffers from bandwidth smearing.

If it were planned to map a lens using a single correlation pass, the effect of

bandwidth-smearing and time-average smearing would be less if the lens were correlated and fringe fitted at a location intermediate between the sources.

Hybrid mapping

The Fourier transformation of the visibility data to form a "dirty" map of the astronomical object is described in section 2.6 of this thesis. The necessity, due to limited u-v coverage, of deconvolving the interferometer beam pattern from the map is explained in section 2.7, and the "CLEAN" deconvolution algorithm is described. The necessity for calibration of the antenna phases is described in section 2.8, and the iterative hybrid-mapping process is explained.

The antenna phases were calibrated to an initial model of the source structure, the same model (point or multiple point) as was used in the fringe fitting. The visibility data were then mapped, the antenna-beam being deconvolved from the data using the CLEAN algorithm. The resulting "clean components" were used to construct a new model of the source. The phases were calibrated using this new model, and the cycle repeated until satisfactory convergence was reached. Once one has a reasonably good model for the source, then further iterations of self-calibration and mapping iteratively improve the model, gradually converging to some final model, which should not depend too much on the sequence of mapping steps taken to reach it.

When phase-referencing, as at lower resolutions with the Very Large Array (VLA), the antenna phases can initially be calibrated by observing a nearby point source. In such a case, the initial maps of the program source show quite well the gross geometry of the source. Hybrid mapping then merely increases the dynamic range to closer to the thermal noise limit. A complication with Very Long Baseline Interferometry is that nearby point-like sources for phase referencing are generally not available. Therefore the initial model for self-calibration may be very dissimilar to the actual source — the model may only be a point or a few points, even for sources with complicated structure. The starting model may be so different from the source that, unless care is taken, the hybrid mapping sequence may never begin to converge. It was found necessary to require only a low signal-to-noise ratio (SNR = 1) on the

early phase calibration iterations, so a poor match between the visibility data and the as-yet-poor model would not cause data to be discarded.

The gravitational lenses, with multiple widely separated images, could not (except for B0218+357) each be mapped in a single field; the maximum pixel size was limited by the size of the beam, and due to computational resources the maximum practical field size was 1024×1024 pixels. However, mapping all images of a lens simultaneously was necessary when deconvolving the beam from the map field, so that the residual noise near one image would not be increased by the sidelobes of the other images. For all lenses except PKS B1830-211, the map pixel size used was 0.2 mas, the beam minor axis was at least 1.5 mas, and the beam major axis was at least 2.2 mas, so the beam was well sampled. For PKS B1830-211 the pixel size was 0.3 mas and the beam minor axis was 1.4 mas, so there were not quite 5 pixels across the beam. Fringe-rate mapping was used to verify the image locations and set the initial map field locations. Dirty maps verified the locations of the images, so that the map fields could be centered on them. The map extents in arcseconds were chosen large enough to contain the known flux-emitting regions, and was increased in size as hybrid mapping progressed if necessary to map flux that fell outside the originally chosen region.

When cleaning (section 2.7) the phase-calibrated data to deconvolve the beam pattern and construct a source model, the search region was restricted by setting "clean boxes", to prevent the creation of spurious "sources". To construct a model from the clean components for use in further phase-self-calibration, all clean components were merged (not merely those with amplitudes greater than that of the first negative clean component) and the negative merged clean components were discarded. (With standard Astronomical Image Processing System (AIPS) routines, it is difficult to, in an automated fashion, select clean components down to an intermediate amplitude level, say to half the amplitude of the first negative clean component. For the second pass of data reduction (section 3.4) an AIPS task was modified to do this.)

Several criteria are available to monitor the improvement from iteration to iteration in the hybrid mapping process. (a) In the very early stages of hybrid mapping,

Lens	Image	Expected thermal noise in map (mJy/Beam)	Actual rms map noise (mJy/Beam)
B0218+357	A	0.15	0.26
B0218+357	B	0.15	0.27
PKS B1830-211	A	0.15	0.81
PKS B1830-211	B	0.15	0.70
B1422+231	A	0.14	0.32
B1422+231	B	0.14	0.28
B1422+231	C	0.14	0.23
B1422+231	D	0.14	0.20
MG J0414+0534	A1	0.15	0.15
MG J0414+0534	A2	0.15	0.14
MG J0414+0534	B	0.15	0.14
MG J0414+0534	C	0.15	0.19
B0957+561	A	0.16	0.19
B0957+561	B	0.16	0.21

Table 3.2: Actual rms map noise in the gravitational lens maps compared with thermal noise expected in the maps.

the number of good phase solutions found during phase calibration is a useful criterion. This is low if the model poorly matches the data. If this number does not increase with subsequent hybrid mapping iterations, then the hybrid mapping process is not beginning to converge, and one may need to change one's strategy greatly. (b) The total flux of the clean components in the model that will be used for the phase calibration generally increases as hybrid mapping progresses, plateauing once further iterations cause little further change to the maps. (c) The dynamic range of the maps is another useful criterion, as is the closely related criterion of the rms noise in the regions of the map away from the sources. A higher dynamic range (lower map noise) generally indicates a better map. There is a lower limit on the attainable rms noise in an interferometer map, that due to the thermal noise in the visibility data.

Besides monitoring the criteria of number of phase solutions, total flux of clean components, dynamic range, and map noise, the maps themselves were looked at every iteration or every few iterations. This allowed observation of how, qualitatively, the map quality had improved, and allowed additional clean component search boxes

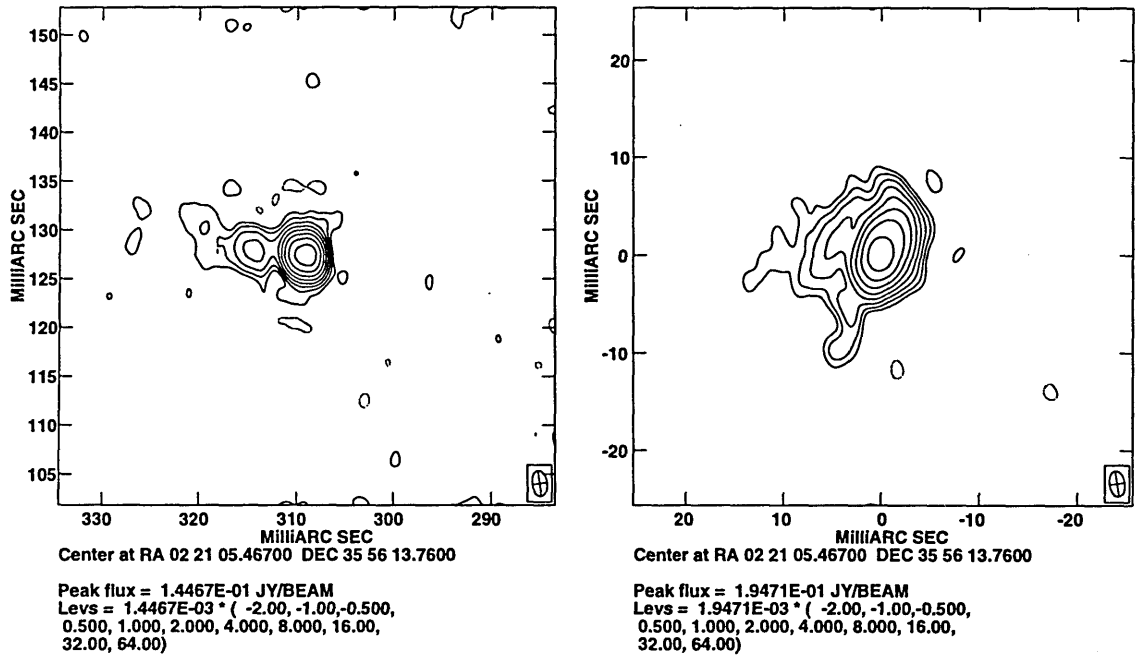

Figure 3-1: Gravitational Lens B0218+357, images B and A. VLBA, 5 GHz. The beam size is shown in the lower right corner of each plot.

to be set for secondary source structure as the map noise decreased sufficiently to uncover it. It was generally apparent when further iterations of hybrid mapping were making little change to the maps. The stopping criterion was not automated.

Fringe fitting to a clean component model and subsequent hybrid mapping

Hybrid mapping produced a clean component model that fairly well represented the source structure. With this source structure known, this model was used to redo the fringe fitting. The hybrid mapping process was repeated to produce a final map.

3.3 The gravitational lenses: particular considerations for each lens, and the completed maps.

3.3.1 Gravitational lens B0218+357

For the fringe fitting of B0218+357, a two-point model based on a VLBI map of

Patnaik *et al.* (1993) was used, with the model location of A at the origin, and B offset to the northeast. The data were fringe fitted to this two-point model, searching only a narrow delay and rate range, and ignoring the longer baselines. For the initial hybrid mapping of this source, all spectral channels were averaged together, and both A and B were mapped in a single field, which was possible for this lens since the image separation was small. After fringe fitting to a clean component model, subsequent mapping was done with the spectral averaging indicated in table 3.1, using two small fields, one around each lens image.

3.3.2 Gravitational lens PKS B1830-211

For PKS B1830-211, fringe fitting to a two-point model was tried, with model positions taken from a preprint of van Ommen *et al.* (1995), with the model flux used for A (the northeast image) somewhat higher than the flux used for B (the southwest image). However, fringe fitting to a point model, with limited u-v range (8×10^7 wavelengths) was more successful than fitting to the two-point model. In retrospect, the two-point model was a poor model: the B (southwest) image was actually the brighter image. Therefore a point model may be better for mapping a multi-component source than a bad multi-point model.

For PKS B1830-211, fringe fitting to a point model shifted the observation's center, since the fringe search found the brighter B image that was southwest of the correlation center. Fringe-rate maps of the pre-fringe-fit data showed a central component and a stronger second component to the southwest. Fringe-rate mapping after fringe fitting located the components center and northeast.

A complication arose in mapping this source, in that the central beam structure was very poorly represented by a Gaussian fit to the central regions. Such a Gaussian "beam" contained much less integrated flux than did the actual beam pattern. As a result, the clean-components restored using this Gaussian "beam" had too little flux compared with the background noise. As a consequence, cleaning a secondary feature made it appear to fade in significance relative to the background noise and the remaining uncleaned flux. This problem was treated by rescaling the residuals

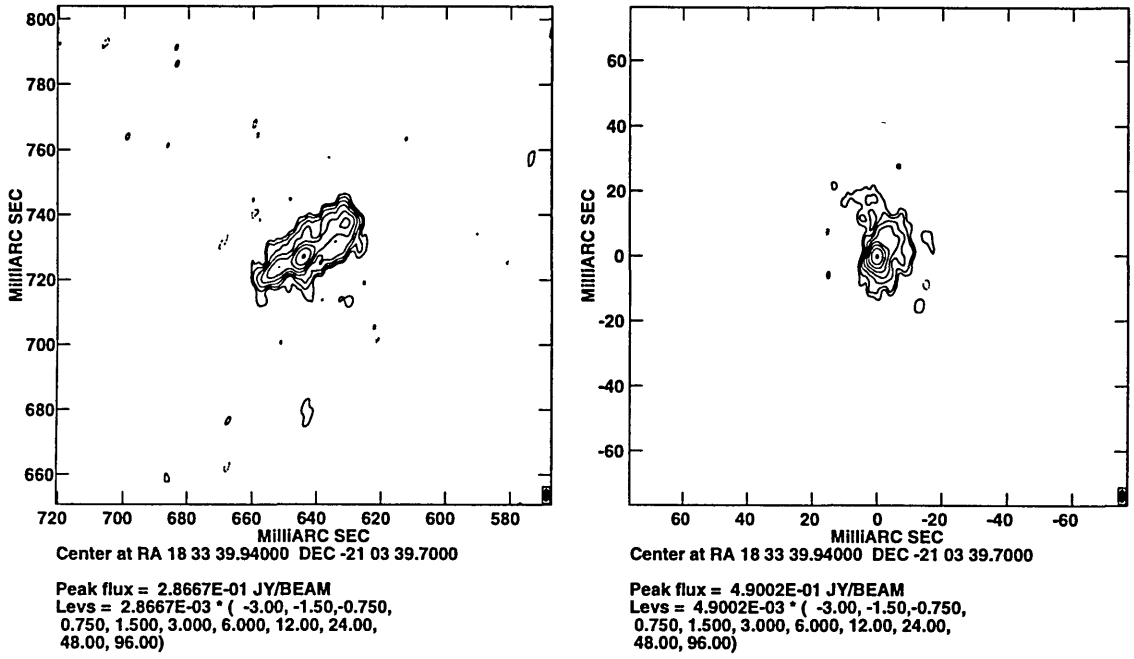

Figure 3-2: Gravitational Lens PKS B1830-211, images A and B. VLBA, 5 GHz. The beam size is shown in the lower right corner of each plot.

by a factor equal to the ratio of the restoring beam area to the actual beam area. In this case the scaling factor was 0.44. When residual rescaling was used, the flux of components no longer appeared to shrink relative to the background when they were cleaned, allowing one to better identify the location of the flux. Once all the source flux from an image is being cleaned, the choice of rescaling or not rescaling the residuals has no affect on the apparent source structure, it merely affects the residual noise level. The residuals were not rescaled in the final PKS B1830-211 maps.

3.3.3 Gravitational lens B1422+231

B1422+231 had a high signal-to-noise ratio, and images A, B, and C were all visible in the fringe-rate maps. Mapping began using three fields for A, B, and C. A fourth field was later added for D. The data reduction reduction process was straightforward for this lens.

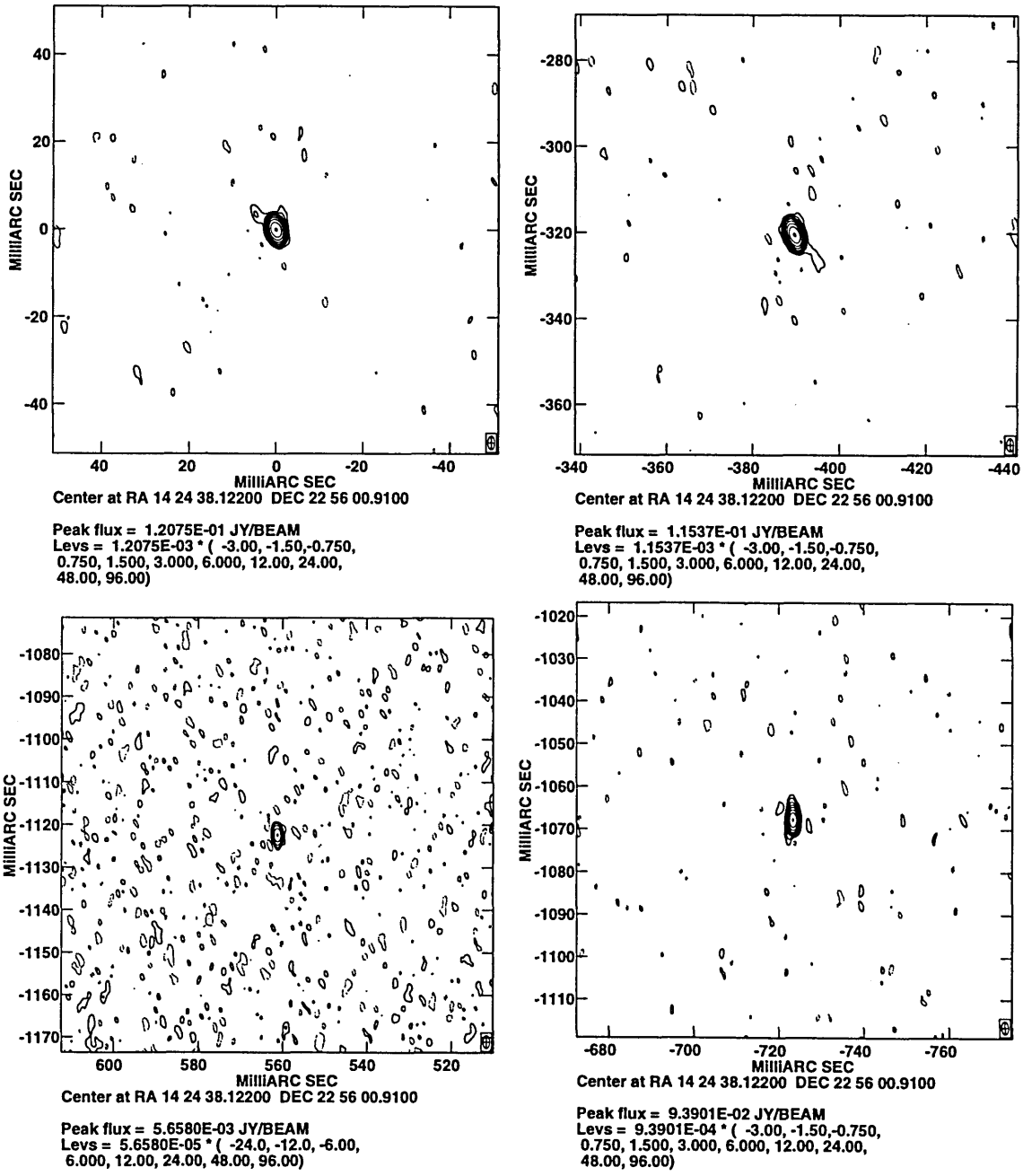

Figure 3-3: Gravitational Lens B1422+231. Clockwise from upper left, images A, B, C, D. VLBA, 5 GHz. The beam size is shown in the lower right corner of each plot.

3.3.4 Gravitational lens MG J0414+0534

For the first few hybrid mapping iterations, only fields for the brightest images, A1, A2, and B, were mapped. The side-lobes of A1 clearly showed up in the residuals of A2 and vice versa, illustrating the necessity of hybrid mapping all four fields simultaneously. After a few hybrid mapping iterations, C was located and a field was added for it. As the maps improved with more iterations of hybrid mapping, structure was revealed within the lens images. Clean boxes were added as new components appeared out of the noise. After some iterations, a wide-field map was made of the residuals, which indicated emission outside the fields originally mapped, so the fields were increased in size. The residuals were examined in the region of the deflector galaxy G, but no emission was seen.

3.3.5 Gravitational lens B0957+561

The signal-to-noise for B0957+561 was low enough that it could not be seen in the fringe-rate plots in all baselines to a central antenna. It was fringe fitted to a point model, and solutions were found for all antennas, except for St. Croix on one 6-minute scan. Despite doing no further time or frequency averaging after correlation, loss in peak flux is expected for image B, due to time average and bandwidth smearing — as B is 6 arcsec from image A. For detailed study of B0957+561 image B, reduction of data correlated at the location of B would be needed. Residual maps revealed no sign of emission near G or near the lobes. There is no sign of the jets that have been seen at other frequencies and higher sensitivities.

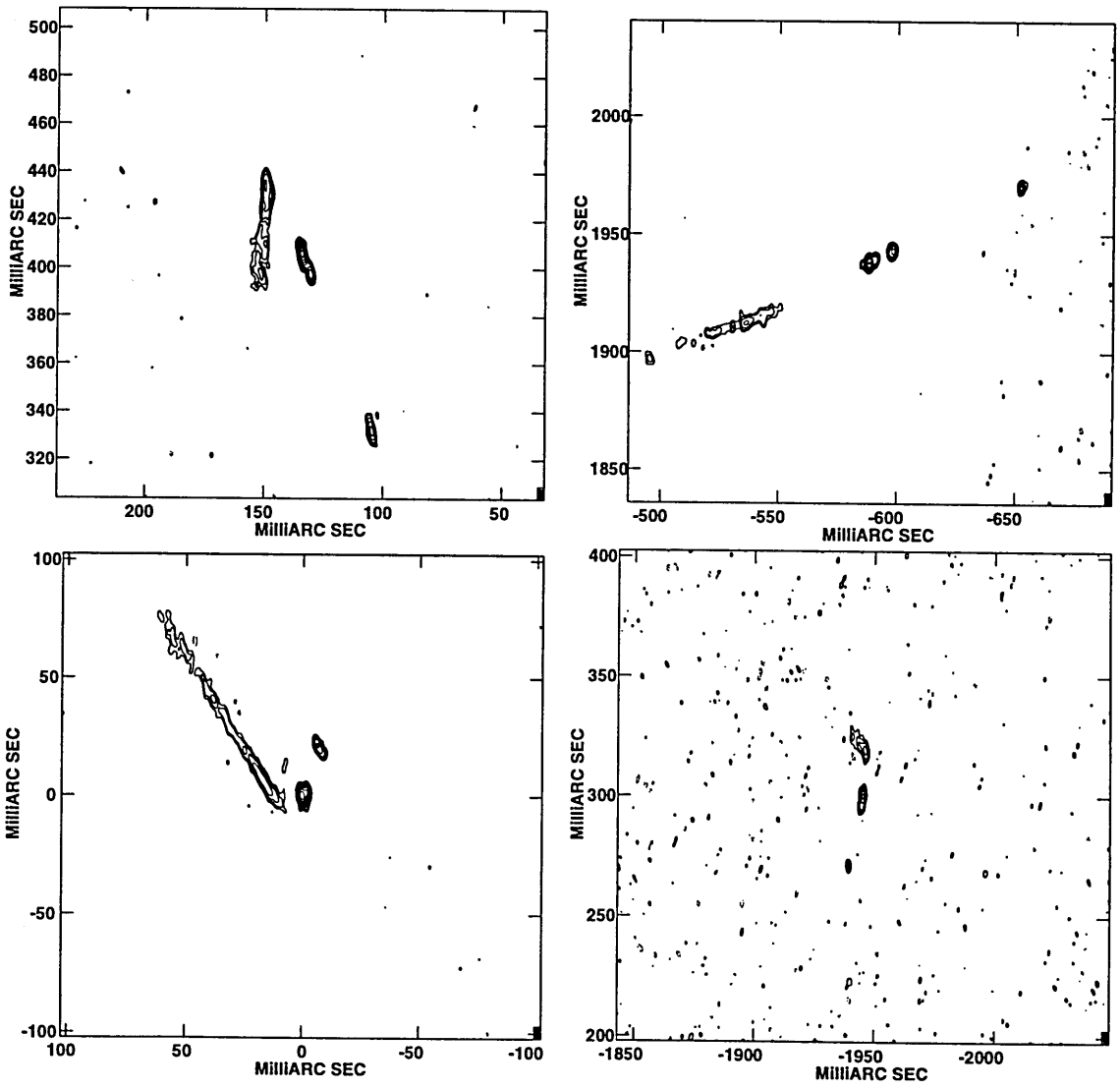

Figure 3-4: Clockwise from lower left, gravitational lens MG J0414+0534 images A1, A2, B, and C. Contour levels are at -2, -1, -0.5, 0.5, 1, 2, 4, 8, 16, 32, and 64 mJy/Beam. Peak fluxes in the fields are 109, 91, 34, and 21 mJy/Beam, respectively. VLBA, 5 GHz. The beam size is shown in the lower right corner of each plot. A map made with better amplitude calibration is in figure 3-6.

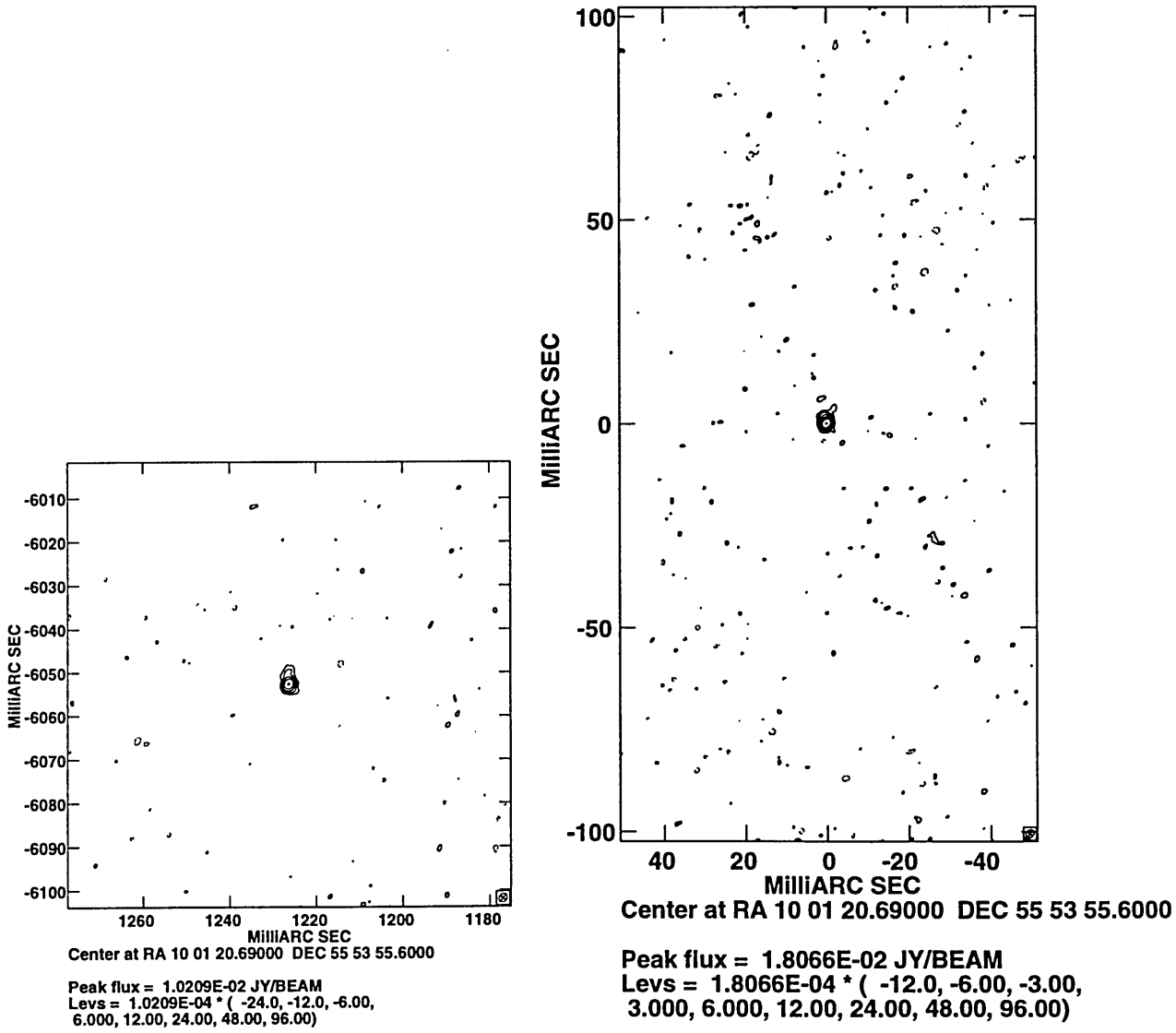

Figure 3-5: Gravitational Lens B0957+561, images B and A. VLBA, 5 GHz. The beam size is shown in the lower right corner of each plot.

3.4 Data reduction: a second pass with improved amplitude calibration and revised mapping technique

Only the simplest of amplitude calibration procedures was performed on the data before the first pass of mapping — amplitude calibration using the system temperatures only. It was desired to correct for certain other known effects, and redo the MG J0414+0534 map. It was also desired to see how sensitive the component locations, the component fluxes, and the lower level features of MG J0414+0534 were to the details of the clean component selection criteria in the hybrid-mapping process.

Improved amplitude calibration

The 15% amplitude difference between right and left circular polarizations was an obvious problem in the amplitude calibration. According to the AIPS Cookbook (1995), such an amplitude offset is characteristic of errors in the threshold levels of the digital samplers. The effect can be corrected from the autocorrelation data (Kogan 1995). The Astronomical Image Processing System (AIPS) task ACCOR was run with a short solution time interval to calculate gain solutions to correct for this effect. The obviously bad gain solutions at the scan ends and between the scans were edited out. Then these gain solutions were smoothed with a box width of 10 hours before being applied to the data. This removed the right-left amplitude differences due to digital sampler bias.

After correcting for the digital sampler bias, the system temperatures and the known gain curves were used for the standard amplitude calibration. This used more recent gain curve information from the VLBA than had been available before, and some of the nominal gain numbers had changed.

For an opacity correction, a simple secant of zenith-angle correction was done assuming a zenith opacity of $\tau = 0.01$ nepers.

The relative antenna amplitudes were corrected using the amplitude-check sources,

as follows. J1310+3220 and J1407+2827 were mapped to see the sources' structure. J1407+2827 clearly consists of two components; J1310+3220 has a main component with a hint of a weak secondary component. Simple models for the source structure were fit directly in the u-v plane to the amplitudes of the visibility data, ignoring the phases, and without allowing a variable factor for the antenna gains, in order to determine the integral flux of the primary source in each field. For J1407+2827 a pair of Gaussians were fit. For J1310+3220 the fitting was done twice: first with a single Gaussian and second with a pair of Gaussians. The model fitting was then done a second time in order to fit for a variable gain factor for each antenna. The position and integral flux of the chief peak were held fixed during the second fitting process, to prevent overall drift in the antenna gains. This gave three sets of antenna gain correction factors: from the 2-Gaussian-fit to J1407+2827, from the 1-Gaussian-fit to J1310+3220, and from the 2-Gaussian-fit to J1310+3220. For all the fits, the ratios were calculated of the fit RMS to the expected thermal noise in the visibilities; the ratios were 1.3, 1.5, and 1.3 respectively. The gain correction factors found from these fits were compared with each other. The standard deviations of the three estimates of the gain correction factors were under half a percent (except Mauna Kea and N. Liberty), even though the corrections were up to 1.8% (or 3.9% for Mauna Kea). It was concluded that the correction factors were generally consistent, and any of them was better than none. The gain corrections from the two-Gaussian fit to J1310-+3220 were applied to the data. The actual correction factors used were all under 1.7%

In summary, the correction of the right-left asymmetry caused by digital sampler bias was the largest amplitude correction, of roughly 15%. Next largest was, for J1310+3220, the antenna gain solutions found from fitting to the amplitude-check sources, with a 2.8% spread between the solutions for the different antennas. Finally, the smallest correction to the gains, for J1310+3220, was the opacity correction, with a 1.3% spread.

After the improved amplitude calibration had been done, the global delay solutions using J1310+3220 as a fringe-finder source were redone before the fringe fitting of any

program sources.

New techniques of cleaning and of choosing a clean component model for phase calibration

A couple of observations were made concerning the use of clean boxes and concerning the selection of clean components for the model to use in phase calibration: If the clean boxes extend well beyond the region of source flux, then obviously false structure appears, positive and negative holes in and at the edges of the box, which artificially lowers the map rms noise. Cleaning using these large clean boxes, but only selecting clean components from a smaller region to use as a model for the phase self-calibration, does not solve this false-structure problem. Therefore the clean boxes must not extend much beyond the extent of the source flux. More care was taken to keep the clean boxes small in the second pass of data reduction, using many small clean boxes to cover diagonal structures rather than few large clean boxes.

The method used in the first pass of data reduction, for creating a model for the phase-self-calibration from the clean components, in which all the cleaned flux is merged and the negative merged components are discarded, has a couple of problems if the source is cleaned very deeply. The model flux can be much higher that the total cleaned flux. As a consequence the cleaning process may not converge. The reason for this is that positive and negative clean components are found that lie on different but nearby pixels, and which fairly much cancel each other out in the map. Merging, then throwing out negatives, artificially raises the flux in the model. Therefore one should not clean too deeply in making a clean component model for self-calibration. This could be done by using only a small number of iterations in the CLEAN algorithm. Alternately, it can be done after the fact by cutting off the clean component list at some point and discarding the remaining, smaller amplitude, clean components. The selected portion of the clean component list may then be merged, and the negative merged components discarded, to construct a clean component model for further self-calibration.

One way of accomplishing this is to take all the clean components with flux of

absolute value larger than some multiple f of the (absolute value of the) flux of the first negative clean component. Standard AIPS routines allow one to do this easily only for $f \geq 1$ or for $f = 0$. Therefore the AIPS routine CCFND was modified to be able to do this for $0 < f < 1$. For subsequent mapping, clean components down to an amplitude of half that of the first negative clean component ($f = 0.5$) were used to construct the phase calibration model. This solved the problem of the model flux being significantly larger than the total cleaned flux, and the resultant non-convergence problem.

3.4.1 Gravitational lens MG J0414+0534: an improved map

With the improved amplitude calibration and the new clean-component selection technique, the MG J0414+0534 data were again reduced. The data were fringe fitted to the best clean-component model from the previous mapping. However, to avoid locking any wrong structure into the self-calibration process, the hybrid mapping iterations were begun using a four-point model for MG J0414+0534, consisting of a single point at the center of each field, with all flux for that field.

The hybrid mapping process started with one tiny clean box per field. During the hybrid mapping iterations, more clean boxes were added as the source structure appeared. The boxes were kept small so as not to include much non-flux region in the boxes. Therefore the long diagonal structures were, eventually, covered by many small clean boxes.

Clean component models for self-calibration were made by taking clean components down to half the level of the first negative, merging, and then discarding the negatives. This differs from what was done in the first pass of data reduction in which all clean components were merged and the negatives discarded.

It was found that the lowest level structures appeared to change depending on where clean boxes were set. Therefore the hybrid mapping process was stopped with a certain amount of structure left in the residuals of the A2 field.

The VLBA maps of MG J0414+0534 revealed structure that was not resolvable on VLA scales. Each image, A1, A2, B, and C, has at least four distinct features —

Lens	Image	Expected thermal noise in map (mJy/Beam)	Actual rms map noise (mJy/Beam)
MG J0414+0534	A1	0.15	0.22
MG J0414+0534	A2	0.15	0.26
MG J0414+0534	B	0.15	0.21
MG J0414+0534	C	0.15	0.19

Table 3.3: Actual rms map noise in the MG J0414+0534 maps compared with thermal noise expected in the maps.

corresponding to four distinct components in the actual source. Proceeding from the central peak in the direction of decreasing RA (increasing RA for C), these components are designated 'a', 'b', and 'c'. The extended component on the opposite side of the central peak is designated 'd'. There is an indication of a fifth component 'e' just east of 'a' in image B, which is not distinguishable from 'a' in the other images since they have less linear magnification in that region of the source.

The locations, fluxes, and extents of the MG J0414+0534 components were found by fitting an elliptical Gaussian to each component. The statistical errors on the component positions due to the thermal noise in the map were calculated from the map rms noise and the fitted Gaussian parameters. The formulas are given in appendix A of this thesis for such statistical errors on elliptical Gaussian fits to maps in which the noise is correlated from pixel to pixel. The component positions and position errors are listed in tables 3.4 and 3.5. The peak and integral fluxes are listed in table 3.7. The fitted component major axes, minor axes, and orientation angles and the deconvolved component sizes are listed in table 3.6.

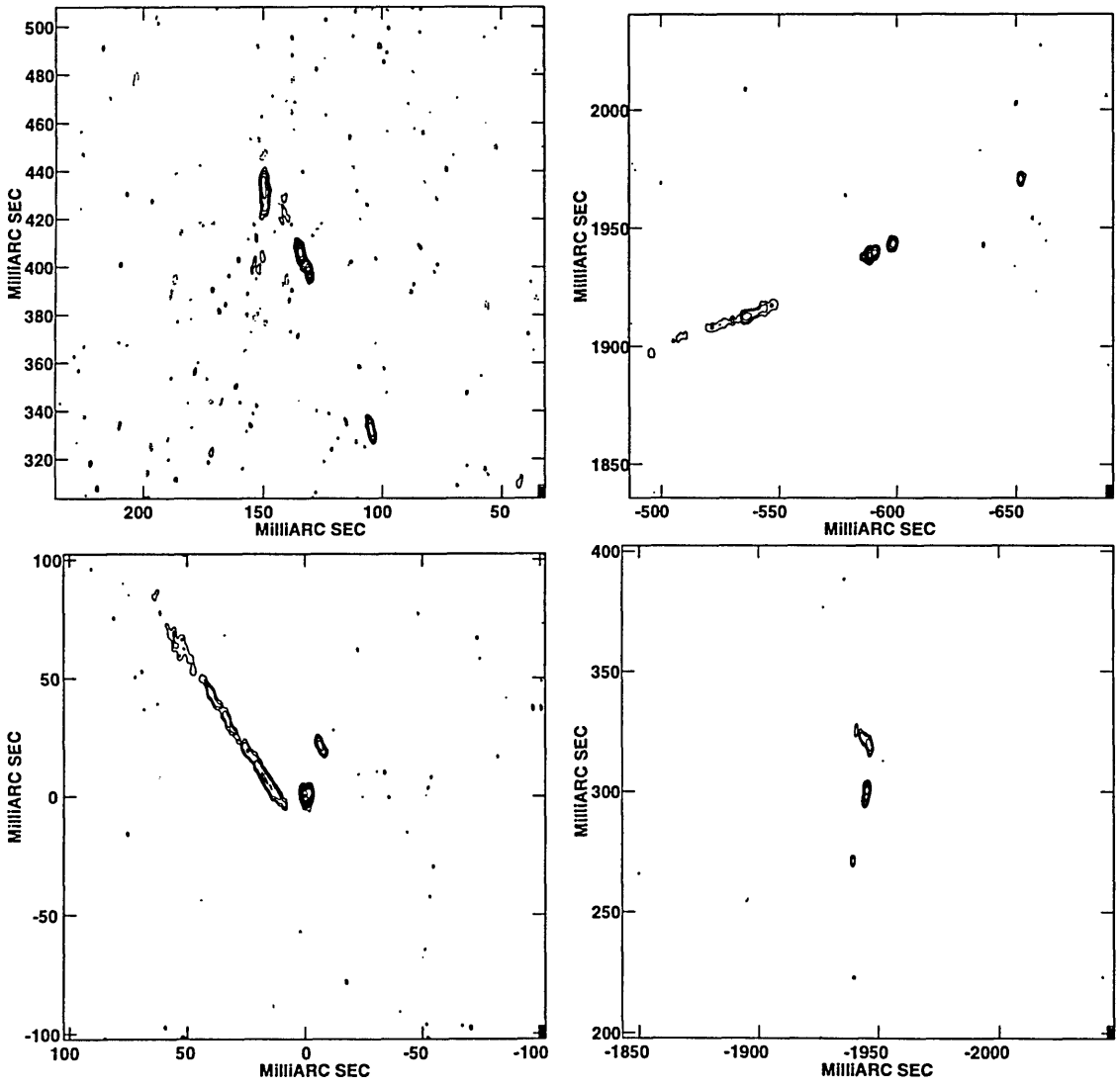

Figure 3-6: Clockwise from lower left, gravitational lens MG J0414+0534 images A1, A2, B, and C. Proceeding from the central peak of each field in the direction of decreasing RA (increasing RA for C) the subcomponents are designated 'a', 'b', and 'c'. The extended component on the opposite side of each central peak is designated 'd'. Contour levels are at -3, -1.5, -0.75, 0.75, 1.5, 3, 6, 12, 24, 48, and 96 mJy/Beam. Peak fluxes in the fields are 110, 91, 34, and 21 mJy/Beam, respectively. VLBA, 5 GHz. The beam size is shown in the lower right corner of each plot.

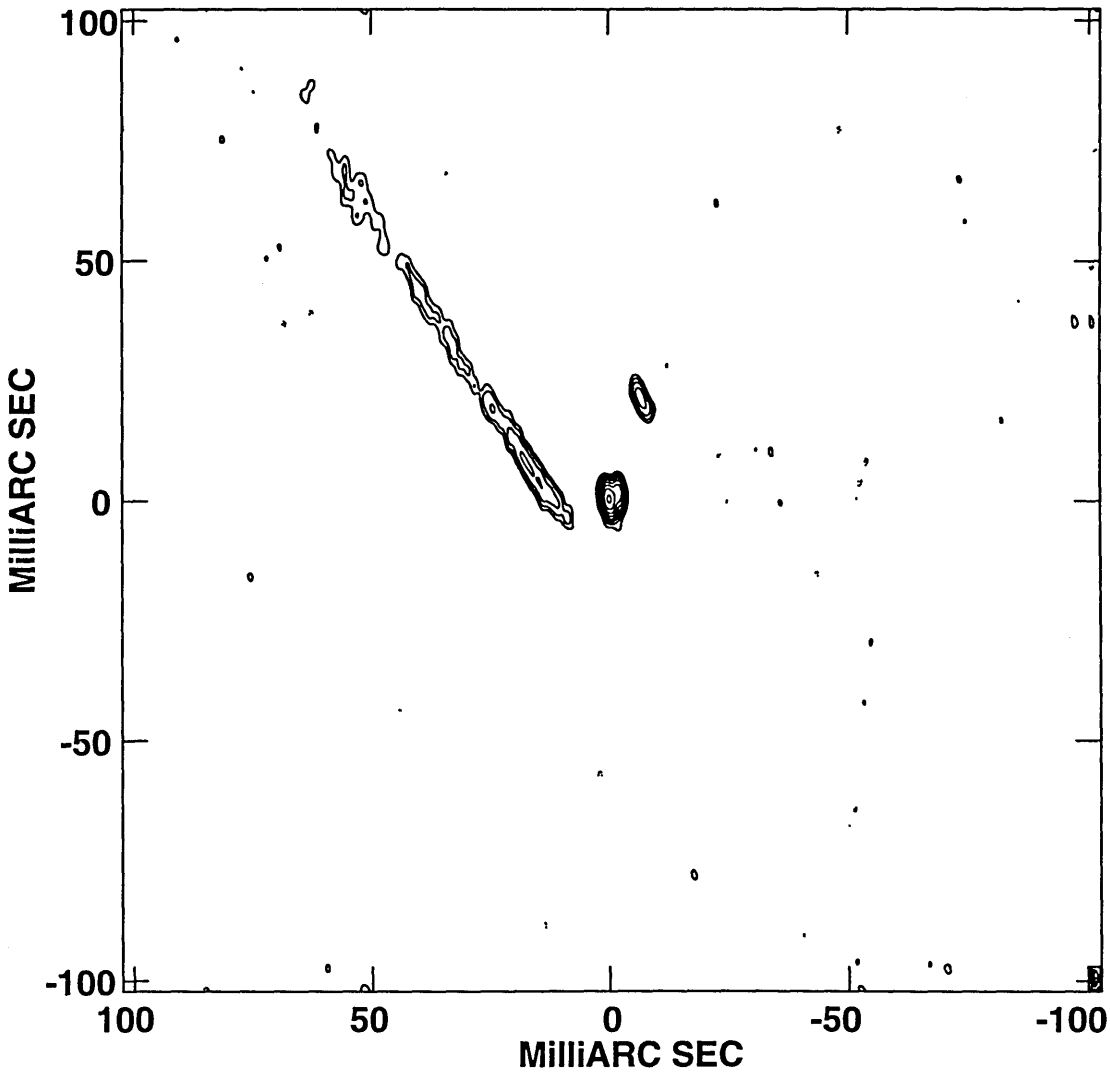

Figure 3-7: Gravitational lens MG J0414+0534 image A1. Contour levels are at -3, -1.5, -0.75, 0.75, 1.5, 3, 6, 12, 24, 48, and 96 mJy/Beam. Peak flux is 110 mJy/Beam. VLBA, 5 GHz. The beam size is shown in the lower right corner of the plot. This is the same plot as in figure 3-6, enlarged to show detail.

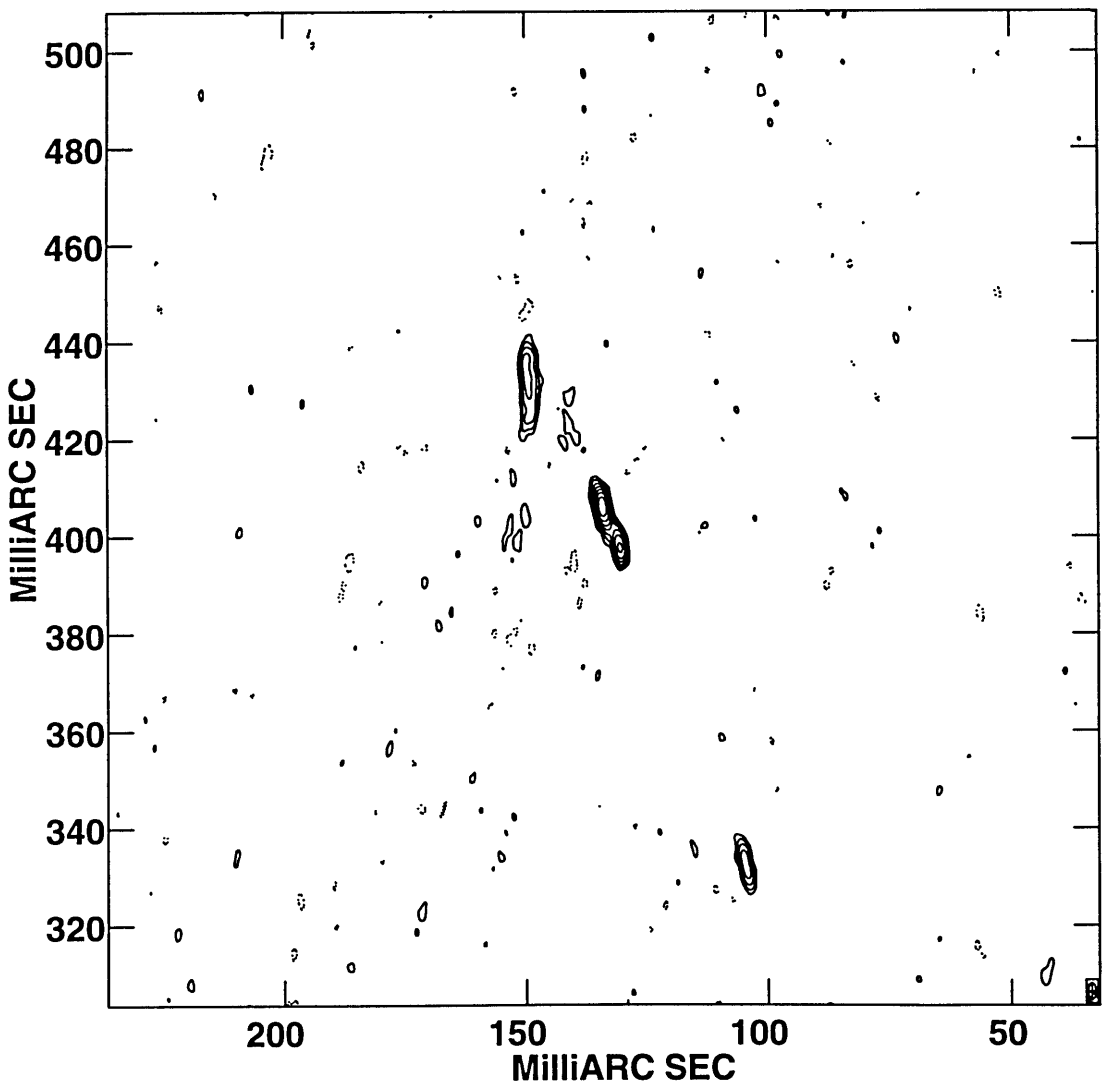

Figure 3-8: Gravitational lens MG J0414+0534 image A2. Contour levels are at -3, -1.5, -0.75, 0.75, 1.5, 3, 6, 12, 24, 48, and 96 mJy/Beam. Peak flux is 91 mJy/Beam. VLBA, 5 GHz. The beam size is shown in the lower right corner of the plot. This is the same plot as in figure 3-6, enlarged to show detail.

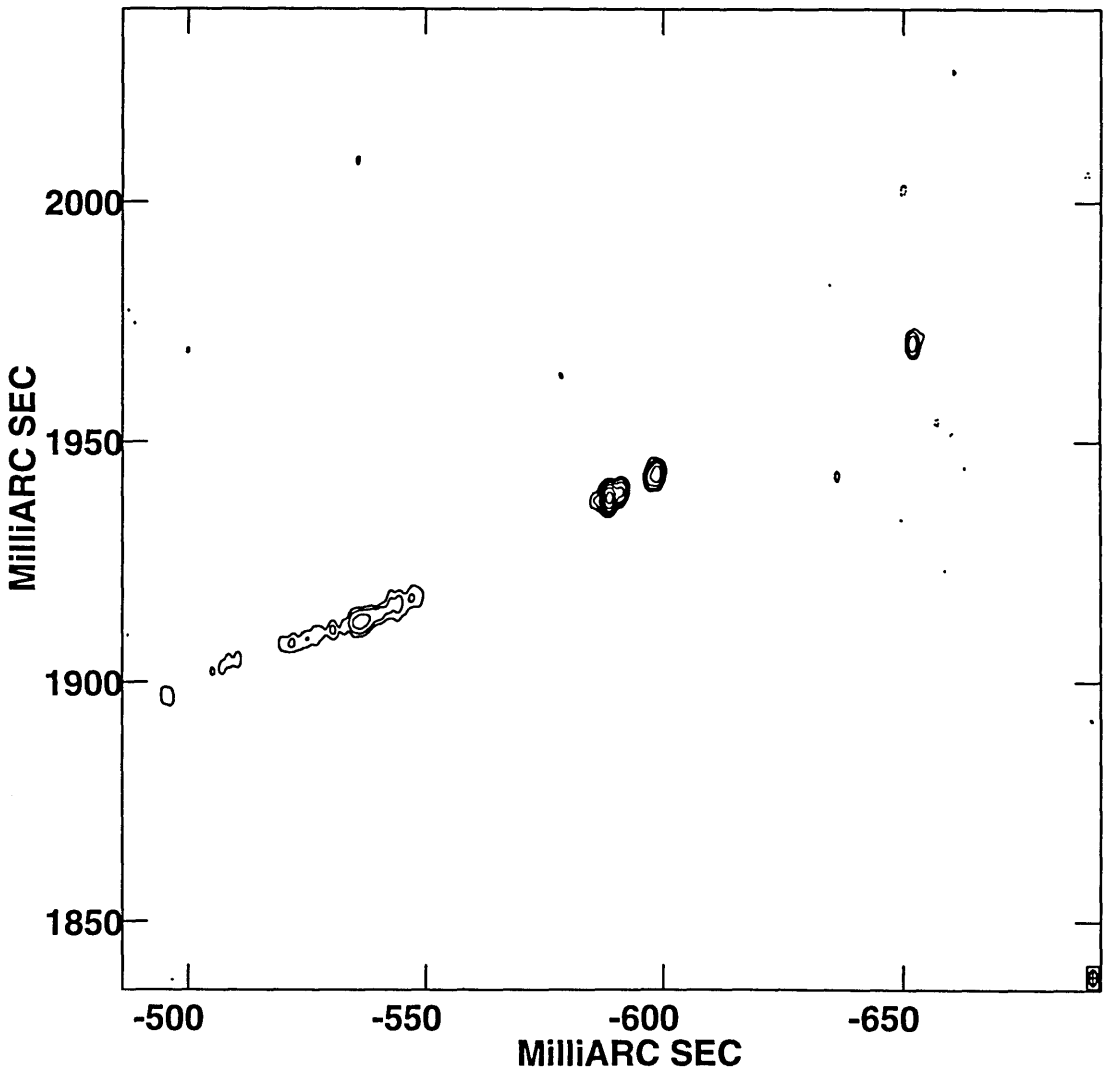

Figure 3-9: Gravitational lens MG J0414+0534 image B. Contour levels are at -3, -1.5, -0.75, 0.75, 1.5, 3, 6, 12, 24, 48, and 96 mJy/Beam. Peak flux is 34 mJy/Beam. VLBA, 5 GHz. The beam size is shown in the lower right corner of the plot. This is the same plot as in figure 3-6, enlarged to show detail.

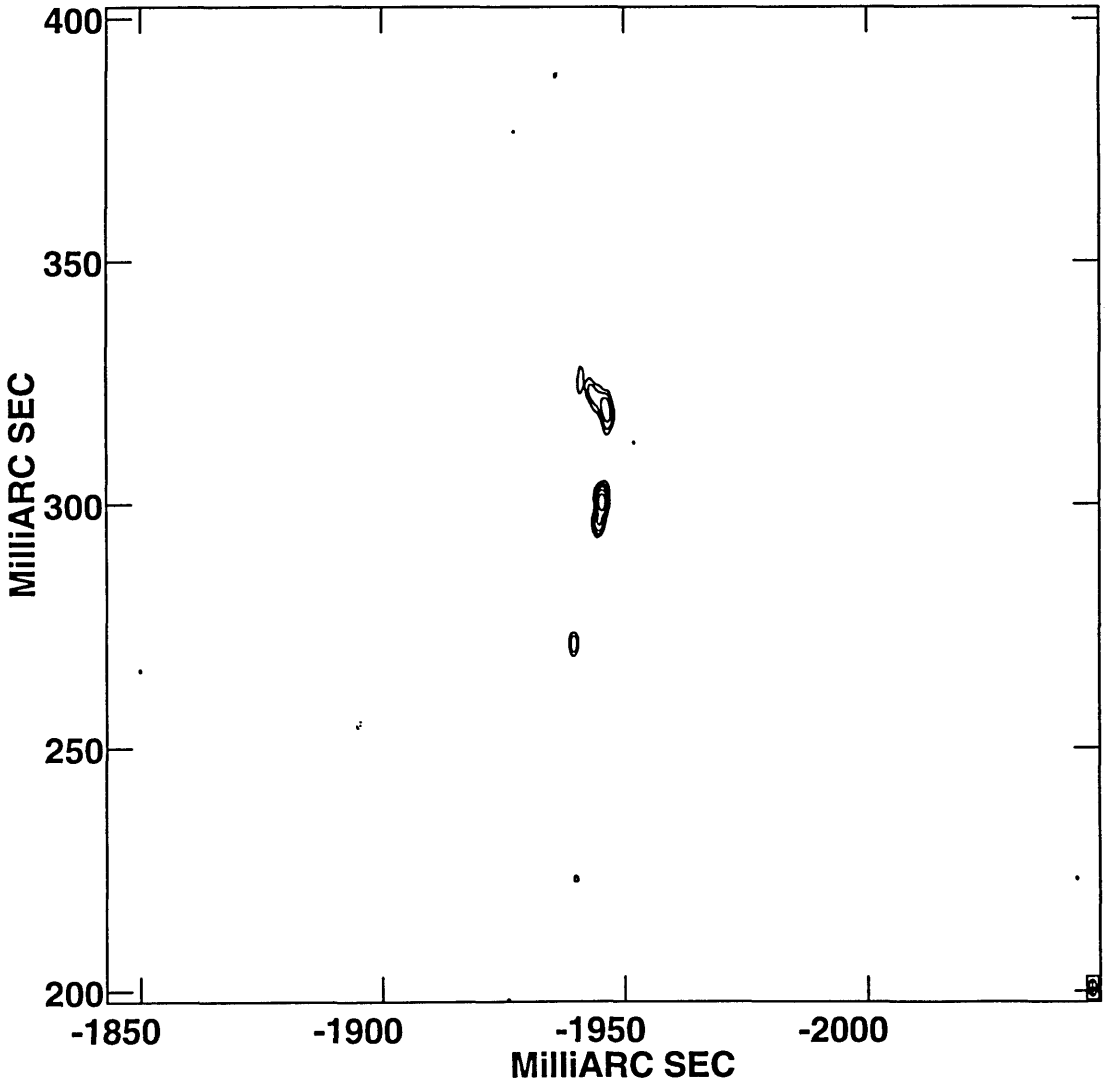

Figure 3-10: Gravitational lens MG J0414+0534 image C. Contour levels are at -3, -1.5, -0.75, 0.75, 1.5, 3, 6, 12, 24, 48, and 96 mJy/Beam. Peak flux is 21 mJy/Beam. VLBA, 5 GHz. The beam size is shown in the lower right corner of the plot. This is the same plot as in figure 3-6, enlarged to show detail.

Component	Component location		Centroid errors		
	$x = -$R.A. (mas)	$y =$ Dec. (mas)	major axis (mas)	minor axis (mas)	angle (degrees E of N)
A1a	0.0144	0.3818	0.0090	0.0040	11.8913629
A2a	-134.0714	405.9972	0.0149	0.0051	11.7622096
Ba	588.6037	1938.3514	0.0211	0.0130	-9.5048033
Ca	1945.3597	300.4118	0.0368	0.0144	-1.6814546
A1b	1.9438	1.6872	0.0245	0.0102	6.3055983
A2b	-130.5108	397.7602	0.0468	0.0152	11.8470413
Bb	598.4155	1943.2536	0.0653	0.0463	-26.6995069
Cb	1944.7253	296.2398	0.0956	0.0455	-10.2059743
A1c	6.7992	21.2226	0.1444	0.0405	24.4282210
A2c	-104.5880	332.3330	0.1860	0.0398	13.0979305
Bc	652.1233	1970.6440	0.1475	0.0835	-3.6386203
Cc	1939.5305	271.3034	0.2365	0.0859	-0.0131469
A1d	-16.1756	6.1174	0.3762	0.0424	31.7455935
A2d	-149.3272	432.3058	0.3147	0.0526	3.2990544
Bd	536.6837	1912.7346	0.2444	0.1041	-68.3828628
Cd	1945.9419	319.7664	0.2676	0.0953	26.0362592

Table 3.4: Fitted locations of the components of MG J0414+0534: the centroid positions for elliptical Gaussians fitted to the components of MG J0414+0534, and the errors on the centroid positions. The positions are relative to the correlation and fringe-fitting center at A1. The x-coordinate increases to the west and the y-coordinate increases to the north. The errors on the centroid positions are given as the major axis, minor axis, and orientation angle of the x-y error distribution, assumed to be Gaussian.

Component	Component location		Centroid errors		
	$x = -$R.A. (mas)	$y =$ Dec. (mas)	σ_x (mas)	σ_y (mas)	σ_{xy}^2 (mas^2)
A1a	0.0144	0.3818	0.0018	0.0037	-2.33e-6
A2a	-134.0714	405.9972	0.0025	0.0062	-7.09e-6
Ba	588.6037	1938.3514	0.0056	0.0089	8.09e-6
Ca	1945.3597	300.4118	0.0061	0.0156	6.06e-6
A1b	1.9438	1.6872	0.0044	0.0103	-9.76e-6
A2b	-130.5108	397.7602	0.0075	0.0195	-7.12e-5
Bb	598.4155	1943.2536	0.0215	0.0263	1.53e-4
Cb	1944.7253	296.2398	0.0203	0.0401	2.22e-4
A1c	6.7992	21.2226	0.0298	0.0563	-1.30e-3
A2c	-104.5880	332.3330	0.0243	0.0770	-1.31e-3
Bc	652.1233	1970.6440	0.0356	0.0626	1.69e-4
Cc	1939.5305	271.3034	0.0365	0.1004	2.01e-6
A1d	-16.1756	6.1174	0.0854	0.1362	-1.13e-2
A2d	-149.3272	432.3058	0.0236	0.1334	-9.98e-4
Bd	536.6837	1912.7346	0.0978	0.0561	3.02e-3
Cd	1945.9419	319.7664	0.0617	0.1036	-4.45e-3

Table 3.5: Fitted locations of the components of MG J0414+0534: the centroid positions for elliptical Gaussians fitted to the components of MG J0414+0534, and the errors on the centroid positions. The positions are relative to the correlation and fringe-fitting center at A1. The x-coordinate increases to the west and the y-coordinate increases to the north. The errors on the centroid positions are given as standard deviations on the x- and y-positions, and the correlation between the x- and y-position errors.

	Fitted size			Deconvolved size		
	angle	major axis	minor axis	angle	major axis	minor axis
	(° E of N)	(mas)	(mas)	(° E of N)	(mas)	(mas)
A1a	7.6 ± 0.1	4.017 ± 0.007	1.810 ± 0.003	26.4	2.10	0.68
A2a	8.6 ± 0.1	4.456 ± 0.012	1.692 ± 0.004	18.8	2.83	0.42
Ba	176.1 ± 0.3	3.568 ± 0.015	1.931 ± 0.011	-63.6	1.37	0.19
Ca	179.3 ± 0.3	3.765 ± 0.028	1.539 ± 0.010	-8.2	1.38	0.27
A1b	4.1 ± 0.2	3.939 ± 0.019	1.666 ± 0.008	15.7	1.84	0.60
A2b	8.7 ± 0.3	4.456 ± 0.039	1.647 ± 0.011	18.7	2.83	0.14
Bb	169.0 ± 1.4	3.660 ± 0.048	2.347 ± 0.042	-65.2	2.07	0.27
Cb	175.7 ± 1.0	3.410 ± 0.066	1.572 ± 0.034	-90.0	0.52	0.00
A1c	20.8 ± 0.6	5.770 ± 0.132	2.005 ± 0.031	29.2	4.76	0.32
A2c	11.7 ± 0.5	6.713 ± 0.173	1.768 ± 0.031	14.9	5.77	0.63
Bc	178.5 ± 1.8	3.753 ± 0.111	1.922 ± 0.070	-33.4	1.43	1.09
Cc	0.3 ± 1.7	3.470 ± 0.165	1.380 ± 0.055	0.0	0.00	0.00
A1d	31.4 ± 0.2	18.825 ± 0.374	2.625 ± 0.034	32.1	18.57	1.40
A2d	3.2 ± 0.3	13.057 ± 0.310	2.393 ± 0.047	3.4	12.58	1.86
Bd	115.2 ± 1.9	6.271 ± 0.238	3.492 ± 0.079	-71.2	5.98	1.10
Cd	23.2 ± 1.3	6.826 ± 0.252	2.681 ± 0.083	29.4	6.01	1.75

Table 3.6: Fitted sizes and deconvolved sizes of elliptical Gaussians fitted to the components of MG J0414+0534. The beam had major axis 3.51 mas, minor axis 1.50 mas, and orientation angle 0.7° E of N.

Component	Peak flux (mJy/Beam)	Integral flux (mJy)
A1a	108.1 ± 0.2	149.1 ± 0.4
A2a	88.0 ± 0.3	125.9 ± 0.5
Ba	33.2 ± 0.2	43.4 ± 0.4
Ca	20.5 ± 0.2	22.5 ± 0.3
A1b	39.2 ± 0.2	48.8 ± 0.4
A2b	28.1 ± 0.3	39.2 ± 0.5
Bb	10.7 ± 0.2	17.5 ± 0.4
Cb	6.5 ± 0.2	6.6 ± 0.3
A1c	9.2 ± 0.2	20.2 ± 0.6
A2c	10.1 ± 0.2	22.7 ± 0.7
Bc	5.2 ± 0.2	7.1 ± 0.4
Cc	2.7 ± 0.2	2.5 ± 0.2
A1d	6.6 ± 0.1	61.6 ± 1.2
A2d	8.4 ± 0.2	49.7 ± 1.2
Bd	4.2 ± 0.1	17.5 ± 0.7
Cd	4.4 ± 0.2	15.2 ± 0.6

Table 3.7: Peak and integral fluxes of elliptical Gaussians fitted to the components of MG J0414+0534.

3.5 Comparison of preliminary and revised data reduction of MG J0414+0534

The two maps of MG J0414+0534 differ somewhat from each other, in qualitative source shape, in component fluxes, and in component locations. In particular, the detailed shape of the low level extended structure in each component is affected by the details of the data reduction process. See figures 3-4 and 3-6.

It might be expected that absolute fluxes would be changed by the corrections to the amplitude calibration while the ratios of the component fluxes would be unaffected by the reduction process. However, the flux ratios changed by a few percent. For the integral fluxes of the elliptical Gaussians fitted to the brightest subcomponent in each field, the first pass at data reduction yielded a flux ratio of A1a/A2a = 1.165, whereas the second data reduction yielded a flux ratio of A1a/A2a = 1.184. This is a 1.6% difference in relative flux of the two brightest image components, just due

to the reduction technique. To check whether this is attributable to the difficulty in separating A1a and A2a from their close companions A1b and A2b, consider instead the summed flux in a box containing both the 'a' and 'b' components. The first pass at data reduction yielded a flux ratio of (A1a&b)/(A2a&b) = 1.1686, whereas the second yielded a flux ratio of (A1a&b)/(A2a&b) = 1.2005. This is a 2.7% difference in relative flux of the brightest image components, again due solely to the reduction technique. This is significantly larger than the 0.3% and 0.4% nominal statistical errors in the A1a and A2a integral fluxes.

The component locations are much less affected by the reduction process than are the component fluxes. Gaussians were fit to each component 'a', 'b', 'c', and 'd', of each lens image A1, A2, B, and C. The shifts in the image centroids (positions of the other 15 components relative to A1a) between the two data reduction passes were comparable to or less than the nominal statistical error on the first data reduction pass, except for A2a, A1c, and A1c, for which the shift was approximately twice or $2\frac{1}{2}$ times the nominal statistical errors. Note that the data, including the u-v coverage and the noise, were identical in both cases. What changed was the amplitude calibration, the clean component selection procedure, and the choice of clean boxes. Therefore it appears that the deconvolution error is at least comparable to the statistical error, and may be larger since the effects of changing u-v coverage and changing thermal noise on the deconvolution has not been examined.

The relative fluxes are much more sensitive to the details of the data reduction process than are the positions of sources. The cleaning and self-calibration may to some extent shift flux between components, but it does not tend to shift the component locations.

3.6 Significance of MG J0414+0534 maps for flux monitoring

The absolute gain calibration of the VLBA is difficult. However, it was anticipated that the relative fluxes of different components of a gravitational lens could be measured accurately so that the relative fluxes rather than the absolute fluxes could be used to monitor for time delays in lenses that are only slightly variable.

In section 3.5, the ratios of fluxes of the components of MG J0414+0534 were shown to change by at least a few percent with a change in the data reduction technique. This suggests that even using flux ratios rather than absolute fluxes, it will be difficult to use the VLBA to measure variability at a level of a few percent.

3.7 Significance of MG J0414+0534 maps for lens modeling.

The VLBA maps of MG J0414+0534 revealed structure that was not resolvable on VLA scales. Each image, A1, A2, B, and C, has at least four distinct components — corresponding to four distinct components in the actual source. The presence of so many components is good for modeling the potential of the deflector. If the magnification matrix is constant over the region of each image, then the positions of three non-collinear components are sufficient to constrain the magnification matrix (second derivative of the potential) at each image location A1, A2, B, and C. In contrast, the fluxes of single-component images constrain merely the total magnification (the determinant of the magnification matrix).

However the MG J0414+0534 images are not merely rotated, stretched, and sheared images of each other. That there is not a simple linear mapping from one image to another indicates that there is a non-zero gradient to the magnification over the extent of the source.

Chapter 4

An Introduction to Gravitational Lensing: Deflector Masses and Cosmological Applications

In this chapter the basic equations of gravitational lensing are introduced. It is shown how gravitational lensing gives information about the surface mass density of the deflector. It is also shown how cosmological parameters may be measured using observations of gravitational lenses.

4.1 Basic lensing equations

When a particle with nonzero mass passes near a massive object, it is deflected by the gravitational attraction. A photon, though massless, is also deflected; the path of light is bent by gravity. If the deflector mass is large enough, and if the background light source is well enough aligned with the deflector, the result is a distorted image of the background source. In some cases there are multiple images of the background source. This is the well-known phenomenon called "gravitational lensing".

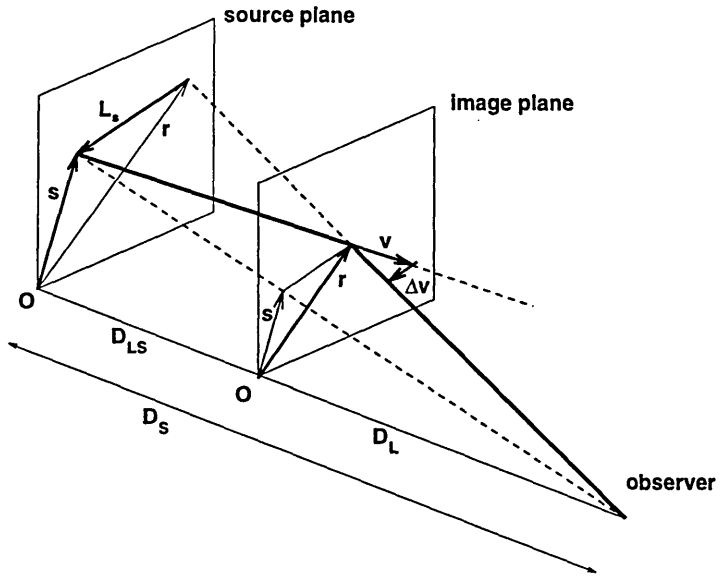

Figure 4-1: Gravitational lensing in the thin lens approximation. Light travels from the angular location $\vec{s}$ of the source to an angular location $\vec{r}$ in the plane of the deflector, where its path is bent abruptly. $\vec{v}$ and $\Delta\vec{v}$ are the photon's initial velocity and the change thereof. D_L is the angular diameter distance to the deflector, located in the image plane. D_S is the angular diameter distance to the source. The angular diameter distance from the deflector to the source is D_{LS}. $\vec{s}$ and $\vec{r}$ are angles subtended at the earth. The origin O may be an arbitrary angular location near the deflector and the images. See text for more explanation.

4.1.1 Deflections and image locations in the thin lens approximation

The gravitational lensing effect is readily calculable in the so-called "thin lens" approximation. The effect of the deflector on the photons is assumed to be significant only when the photons are very close to the deflector. The photons are assumed to travel in a straight line from the source to the plane of the deflector, at which point their path is bent. The photons then continue in a straight path to the observer. (See figure 4-1.)

The actual location of the source on the celestial sphere is at an angular position $\vec{s}$. Light travels from a location $\vec{s}$ on the source plane, to an location $\vec{r}$ on the image plane — the plane containing the deflector. At the image plane the trajectory is bent abruptly. The light proceeds thence in a straight line to the observer. Due to the

bending of the light, the source is not seen at $\vec{s}$; instead an image appears, displaced, at an angular location $\vec{r}$. The deflection angle of the light may be written in terms of the change of the photon's velocity, $\vec{\alpha} = \frac{\Delta\vec{v}}{|\vec{v}|}$. The physical displacement in the source plane, between the apparent location of the source and the actual location of the source, is $\vec{L}_{src} = D_S(\vec{s} - \vec{r})$, where D_S is the angular diameter distance to the source, since $\vec{r}$ and $\vec{s}$ are the angular locations of the source and image. The displacement may also be specified in terms of the bending angle at the deflector and the angular diameter distance from the deflector to the source $\vec{L}_{src} = D_{LS}\frac{\Delta\vec{v}}{|\vec{v}|}$, so that

$$\vec{s} = \vec{r} + \frac{D_{LS}}{D_S}\frac{\Delta\vec{v}}{|\vec{v}|}. \tag{4.1}$$

The bending of the photon trajectory by the gravitational field of the deflector is properly a general relativistic effect. However we assume that the gravitational field of the deflector is not strong, $\frac{1}{c^2}\Phi_{\text{Newtonian}} \ll 1$,[1] where $\Phi_{\text{Newtonian}}$ is the standard Newtonian-mechanics gravitational potential. The deflection of a photon properly calculated using general relativity, is twice the deflection calculated using Newtonian mechanics, for a massive particle with the same speed and impact parameter as the photon. Moreover, for a photon, the deflection is so small, and the path so nearly straight, that in calculating the deflection the photon's trajectory can be approximated as a straight line through the gravitational potential. It seems perverse to ignore the deflection in calculating the deflection, but it can be shown that this just ignores terms of higher order in $\Phi_{\text{Newtonian}}$. Thus the deflection, or more specifically the change in velocity, is the integral along a straight line trajectory from the source plane to the observer of twice the Newtonian force per unit mass, $\Delta\vec{v} = 2\int_{\text{src}}^{\text{obs}} dt\frac{\vec{F}}{m}$. Since the force per unit mass is the negative gradient of the gravitational potential, $\Delta\vec{v} = -2\int_{-\infty}^{\infty} dt\ \vec{\nabla}_{\vec{x}}\Phi_{\text{Newtonian}}(\vec{x})\Big|_{\vec{x}=\vec{B}-vt\hat{z}}$, where the photon trajectory has been written as $\vec{x} = \vec{B} - vt\hat{z}$, in terms of its velocity v and its impact parameter $\vec{B}$ which is a physical length. Changing variables from time, t, to physical distance along the trajectory, $z = -vt$, then $\Delta\vec{v} = -\frac{2}{v}\int_{-\infty}^{\infty} dz\vec{\nabla}_{\vec{x}}\Phi_{\text{Newtonian}}(\vec{B}, z)$, where it can be seen that

[1] Henceforth c will be set equal to 1 in all equations in this chapter.

the partial derivative with respect to z integrates to zero. It is convenient to write the impact parameter in terms of the angle $\vec{r}$ subtended at the earth, $\vec{r} = \vec{B}/D_L$:

$$\frac{\Delta\vec{v}(\vec{r})}{|\vec{v}|} = -\vec{\nabla}_{\vec{r}}\left\{\frac{2}{D_L}\int dz \frac{1}{v^2}\Phi_{\text{Newtonian}}(D_L\vec{r}, z)\right\}. \tag{4.2}$$

Combining this with equation 4.1, recalling that the speed of light is $v = c = 1$, gives the mapping from the image plane to the source plane:

$$\vec{s} = \vec{r} - \vec{\nabla}_{\vec{r}}\left\{\frac{2D_{LS}}{D_L D_S}\int dz \Phi_{\text{Newtonian}}(D_L\vec{r}, z)\right\}. \tag{4.3}$$

Note the following concerning the mapping. Any image plane location $\vec{r}$ necessarily maps back to a unique location $\vec{s}(\vec{r})$ in the source plane. However, depending on the mass distribution of the deflector, and on the distances to the source and the deflector, the mapping from $\vec{s}$ to $\vec{r}$ may not be unique. That is, there may be multiple images of the same source. Note also that the cosmological model used for spacetime enters the mapping only via the angular diameter distances.

4.1.2 The 2-D lens potential

For convenience, a 2-dimensional gravitational-lensing potential is introduced:

$$\Phi(\vec{r}) = \frac{2D_{LS}}{D_L D_S}\int dz \Phi_{\text{Newtonian}}(D_L\vec{r}, z). \tag{4.4}$$

The divergent part of the integral may be ignored, since it is just an infinite constant which has no dependence on the impact parameter $\vec{r}$. The mapping from image-plane to source-plane then takes the well-known form

$$\vec{s} = \vec{r} - \vec{\nabla}_{\vec{r}}\Phi(\vec{r}). \tag{4.5}$$

The Newtonian gravitational potential satisfies a 3-dimensional Poisson equation,

$$\nabla^2_{\vec{x}}\Phi_{\text{Newtonian}}(\vec{x}) = 4\pi G\rho(\vec{x}). \tag{4.6}$$

By integrating along the line of sight, it may readily be shown that the lensing potential satisfies a 2-dimensional Poisson equation. First the surface mass density is defined.

$$\sigma(\vec{r}) = \int_{-\infty}^{\infty} dz\, \rho(z, D_L\vec{r}). \tag{4.7}$$

Then integrating the 3-dimensional Poisson equation (4.6) along the line of sight,

$$\nabla^2_{(D_L\vec{r})}\left(\frac{D_L D_S}{2D_{LS}}\Phi(\vec{r})\right) = 4\pi G\sigma(\vec{r}). \tag{4.8}$$

This is more familiar in the form

$$\nabla^2_{\vec{r}}\Phi(\vec{r}) = \frac{2\sigma(\vec{r})}{\sigma_{\rm crit}}, \tag{4.9}$$

where $\sigma_{\rm crit} = \frac{1}{4\pi G}\frac{D_S}{D_L D_{LS}}$ is the critical mass density necessary to guarantee the ability to produce multiple images by gravitational lensing (Schneider *et al.* 1992, sections 5.4.3 and equations 5.4 and 5.5) Note that $\vec{r}$ is a dimensionless angular position on the sky (i.e. measured in radians).

4.1.3 The magnification matrix

It is often useful to know how a small displacement in the source plane affects the image location. The magnification matrix is defined so that $\Delta\vec{r} = M\Delta\vec{s}$. This is a 2×2 matrix which is a function of the image plane location, $M_{ij}(\vec{r}) = \frac{\partial r_i}{\partial s_j}$. The inverse of the magnification matrix is readily calculable from the lens potential, ${M^{-1}}_{ij}(\vec{r}) = \frac{\partial s_i(\vec{r})}{\partial r_j} = \delta_{ij} - \frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\Phi(\vec{r})$.

4.1.4 Specific intensity

It can be shown that the specific intensity of radiation is unchanged by the presence of a lens. The sky brightness (Jy/arcsec^2) of an object in the presence of a gravitational lens is the same as that in the absence of the lens; only its apparent size changes. (Schneider *et al.* 1992, sections 3.6 and 2.3) This does of course mean that the total

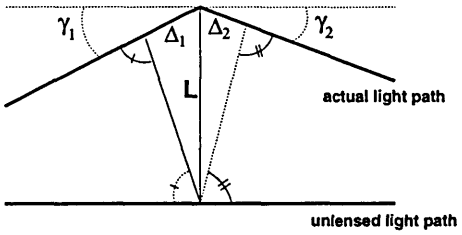

Figure 4-2: The geometrical path length increase, $\Delta = \Delta_1 + \Delta_2$, due to gravitational lensing. Thin lens approximation.

spectral flux density (Jy) from a source reaching the observer is changed by the lensing — it scales as the area of the image. For a nearly point-like source, the flux increases by a factor of $\det \boldsymbol{M}$.

4.1.5 Time delays in the thin lens approximation

In a multiple-image gravitational lens, there are multiple light paths from the source to the observer. In general the light travel time along one path is different from the travel time along another path. If a lensed source varies in flux, the observer will see the fluxes of the images change at different times. This lens "time delay" is another observable feature of gravitational lenses.

The travel-time increase due to lensing has two parts. One is due to the increase in the geometrical path length. The other is the retardation of the light in the gravitational field of the deflector.

That the physical path length is increased due to deflection may be seen in figure 4-1. The physical path length increase is illustrated in figure 4-2. The photon's actual trajectory is displaced from what it would be without the deflector by a physical distance $L_{img} = D_L|\vec{r} - \vec{s}|$. The path length increase in the frame of the deflector, due to the deflection, is the physical distance $\Delta = \Delta_1 + \Delta_2$ (see figure 4-2), where $\Delta_1 = \frac{1}{2}\gamma_1 L_{img}$ and $\Delta_2 = \frac{1}{2}\gamma_2 L_{img}$. The sum of the angles $\gamma_1 + \gamma_2$ is simply the total deflection angle α calculated above in section 4.1.1, $\gamma_1 + \gamma_2 = \alpha = \frac{|\Delta\vec{v}|}{|\vec{v}|} = \frac{D_S}{D_{LS}}|\vec{s} - \vec{r}|$. Therefore, combining these expressions, the geometrical path length increase Δ, as

measured in the frame of the deflector is

$$t_{\substack{\text{delay}\\\text{geom}}} = \Delta = \frac{1}{2}\frac{D_S D_L}{D_{LS}}|\vec{r} - \vec{s}|^2, \tag{4.10}$$

which is also the travel time increase due to the geometrical path length increase along this trajectory.

There is also a travel time increase due to the passage of the light through the potential of the deflector. The metric near the deflector may be approximated (Misner *et al.* 1973, equation 18.15c) as

$$ds^2 = -(1 + 2\Phi_{\text{Newtonian}})dt^2 + (1 - 2\Phi_{\text{Newtonian}})d\vec{x}^2. \tag{4.11}$$

A photon travels on a light-like path, a null-geodesic, with $ds = 0$, so that $dt = -(1 - 2\Phi_{\text{Newtonian}})dz$. Therefore the time taken for a photon to travel from the source location z_{src} to the observer at location z_{obs} is

$$t_{\text{obs}} - t_{\text{src}} = \int_{t_{\text{src}}}^{t_{\text{obs}}} dt \tag{4.12}$$

$$= z_{\text{src}} - z_{\text{obs}} - 2\int_{z_{\text{obs}}}^{z_{\text{src}}} \Phi_{\text{Newtonian}}(D_L\vec{r}, z)dz. \tag{4.13}$$

Comparing this with the travel time in the absence of a deflector, $t_{\text{obs}} - t_{\text{src}} = z_{\text{src}} - z_{\text{obs}}$, the delay in the arrival time of a photon due to passing through the deflector's potential well is

$$t_{\substack{\text{delay}\\\text{grav}}} = -2\int_{z_{\text{obs}}}^{z_{\text{src}}} \Phi_{\text{Newtonian}}(D_L\vec{r}, z)dz. \tag{4.14}$$

Only the integral in the region near the deflector contributes to travel time differences between the various images of various objects in the source plane, and only the travel time differences are of interest, so the limits of integration may be extended to $\pm\infty$. Therefore the relative arrival time delay of a photon due to passing through the deflector's potential well is

$$t_{\substack{\text{delay}\\\text{grav}}} = -2\int_{-\infty}^{+\infty} \Phi_{\text{Newtonian}}(D_L\vec{r}, z)dz \tag{4.15}$$

$$= -\frac{D_L D_S}{D_{LS}}\Phi(\vec{r}), \tag{4.16}$$

where equation 4.4 has been used in the last step.

The geometrical and the gravitational time delays are added and transformed from the frame of the deflector to the frame of the observer. The deflector-induced delay $t(\vec{r})$ in the arrival time of a photon at the observer is

$$t(\vec{r}) = t^{\text{delay}}_{\text{geom}} + t^{\text{delay}}_{\text{grav}} \tag{4.17}$$

$$t(\vec{r}) = (1+z_L)\frac{D_L D_S}{D_{LS}}\left(\frac{1}{2}|\vec{r}-\vec{s}|^2 - \Phi(\vec{r})\right), \tag{4.18}$$

$$t(\vec{r}) = (1+z_L)\frac{D_L D_S}{D_{LS}}\left(\frac{1}{2}|\vec{\nabla}_{\vec{r}}\Phi(\vec{r})|^2 - \Phi(\vec{r})\right), \tag{4.19}$$

(Schneider *et al.* 1992, equation 4.67) and the time delay between two images, at $\vec{r}_\alpha$ and $\vec{r}_\beta$, of the same source is

$$\Delta t_{\beta\alpha} = t(\vec{r}_\alpha) - t(\vec{r}_\beta) = (1+z_L)\frac{D_L D_S}{D_{LS}}\left(\frac{1}{2}|\vec{\nabla}_{\vec{r}_\alpha}\Phi(\vec{r}_\alpha)|^2 - \frac{1}{2}|\vec{\nabla}_{\vec{r}_\beta}\Phi(\vec{r}_\beta)|^2 - \Phi(\vec{r}_\alpha) + \Phi(\vec{r}_\beta)\right). \tag{4.20}$$

4.1.6 Degeneracies: transformations of the lens potential that do not affect the lens image sky brightness distribution

There are degeneracies in the lensing equation $\vec{s} = \vec{r} - \vec{\nabla}_{\vec{r}}\Phi(\vec{r})$ such that the mapping between the image plane and the source plane does not completely determine the potential $\Phi(\vec{r})$.

Suppose there are N_{src} sources, with $N_{\text{img},k}$ images of the k^{th} source. For each source k, the lens potential must satisfy

$$\text{for all } \alpha = 1 \ldots N_{\text{img},k}, \qquad \vec{s}_k = \vec{r}_{k\alpha} - \vec{\nabla}\Phi(\vec{r}_{k\alpha}). \tag{4.21}$$

If some potential $\Phi(\vec{r})$ satisfies this relation, can other potentials $\Phi'(\vec{r})$ be found that also satisfy this relation? Obviously one could simply add a constant to the potential,

and nothing would change — indeed the lens potential is ill defined up to an (infinite) additive constant. Falco, Gorenstein, and Shapiro (Falco *et al.* 1985; Gorenstein *et al.* 1988) showed that there are other transformations of the lens model parameters that leave the lens image surface brightness unchanged; that is, the image locations and fluxes are unchanged. Under this set of transformations, however, the time delays between all pairs of images are all scaled by the same factor, which can be confused with a rescaling of the Hubble parameter. The particular set of transformations that leaves the sky brightness unchanged is:

$$\begin{aligned}
\Phi(\vec{r}) &\longrightarrow \Phi'(\vec{r}) = (1-\kappa)\Phi(\vec{r}) + \Phi_o + \vec{\Phi}_1 \cdot \vec{r} + \tfrac{1}{2}\kappa|\vec{r}|^2 && \text{(2-D lens potential)} \\
\vec{s}_k &\longrightarrow \vec{s}'_k = (1-\kappa)\vec{s}_k - \vec{\Phi}_1 && \text{(source locations)} \\
S_k &\longrightarrow S'_k = (1-\kappa)^2 S_k && \text{(source fluxes}^2\text{)}
\end{aligned} \tag{4.22}$$

These transformations have four parameters: κ and Φ_o are scalar constants; $\vec{\Phi}_1$ is a vector constant. Gorenstein *et al.* (1988) called $\vec{\Phi}_1$ the prismatic transformation and κ the magnification transformation; the κ degeneracy is also referred to as the mass sheet degeneracy. These transformations change neither the image locations nor the images' brightness distributions. Their effect on the magnification matrix is $\boldsymbol{M}(\vec{r}) \longrightarrow \boldsymbol{M}'(\vec{r}) = \frac{1}{(1-\kappa)}\boldsymbol{M}(\vec{r})$, but this has no visible effects, since it merely balances the shifts in the source positions and the rescaling of the source fluxes. The effect of the transformations on the deflector surface mass density is

$$\sigma(\vec{r}) \longrightarrow \sigma'(\vec{r}) = (1-\kappa)\,\sigma(\vec{r}) + \kappa\sigma_{\rm crit}. \tag{4.23}$$

The mass of the principal deflector is reduced by a factor of $(1-\kappa)$. A uniform mass sheet, of surface density κ times the critical density, has been added to the system. This reduces the convergence attributable to the principal deflector, while adding to the lens system a constant convergence κ associated with the mass sheet.

[2]The source surface brightness (Jy/arcsec^2) is unchanged. But since the source area changes, the total flux (Jy) for each source is transformed.

For the α^{th} image at location $\vec{r}_{k\alpha}$ of the k^{th} source at location $\vec{s}_k$, the arrival time delay $t(\vec{r}_{k\alpha})$ (equation 4.19) transforms as

$$t(\vec{r}_{k\alpha}) \longrightarrow t'(\vec{r}_{k\alpha}) = (1-\kappa)\, t(\vec{r}_{k\alpha}) - (1-\kappa)\,(1+z_L)\frac{D_L D_S}{D_{LS}}\left\{\vec{s}_k \cdot \vec{\Phi}_1 + \frac{\kappa}{2}|\vec{s}_k|^2\right\}. \quad (4.24)$$

The second term does not contribute to a time delay between two images of the same source, though it does contribute to the time delays between images of different sources. In particular, for events simultaneous in the source plane but at, for example, source plane positions spaced along a line, the prismatic degeneracy $\vec{\Phi}_1$ adds an arrival time delay which increases linearly with source position, and the mass sheet degeneracy κ adds a delay which goes quadratically with source position. For two images, α and β, of the same source, the change of the time delay under this transformation is

$$\Delta t_{\beta\alpha} \longrightarrow \Delta t'_{\beta\alpha} = (1-\kappa)\,\Delta t_{\beta\alpha} \quad (4.25)$$

The time delay (equation 4.20) is proportional to the inverse of the Hubble parameter through its proportionality to a distance factor $\frac{D_L D_S}{D_{LS}}$. Therefore the presence of a mass sheet as part of the deflector can be confused with a rescaling of the Hubble parameter. The presence of such a mass sheet cannot be detected by the image geometry nor by the time delay. Actually, the deflector need not have a mass sheet extending indefinitely in all directions for this degeneracy term to cause a rescaling of the Hubble parameter estimated from a gravitational lens. It will be discussed in chapter 6 how a mismatch between the actual deflector radial profile and that used to model the deflector can cause an error in the deduced Hubble parameter.

4.2 Deriving information on the deflector mass distribution and on the deflector and source distances from observations

From observations of the lens system it is possible to deduce information about the deflector mass distribution, and about the angular diameter distances to the deflector and the source. Section 4.2.1 lists the available observables. Sections 4.2.2 and 4.2.3 explain how these put constraints on the deflector potential $\Phi(\vec{r})$. Section 4.2.4 explains how the deflector potential is used to find the surface mass density and, combined with a time delay measurements, an angular diameter distance to the lens system.

4.2.1 Observable quantities

For every gravitational lens system there are certain things that are potentially observable.

Lens image geometry

Firstly, one can observe the brightness and shape of the lens images. There may be images of more than one background object. The correspondence between observed components may be apparent; it may be obvious which observed components are images of the same background source; see, for example, figure 3-6. One can determine the images' angular locations, angular extents, and fluxes, and perhaps more detailed geometry information as well.

Time delays

A series of observations may be made to monitor the lens images for variability. If the source is variable, the time delay may be determined for one or more pairs of images of the source.

Redshifts

Source and deflector redshifts may be measured from spectral lines.

The deflector

If the deflector itself is luminous, then it may be observed. The location, radial profile of the light, ellipticity of the luminous matter, velocity dispersion, etc., may be determined. Gravitational lensing is sensitive to all the mass of the deflector — including any dark matter. Observations of the deflector itself just detect the luminous mass; nevertheless these may be useful in modeling.

4.2.2 Lens potential — constraints from the lens images

The mapping from the image plane to the source plane is $\vec{s} = \vec{r} - \vec{\nabla}_{\vec{r}}\Phi(\vec{r})$. The lens potential may be modeled based on the geometry of the images. The general concept is as follows. Some parameterized form is assumed for the lens deflector potential $\Phi(\vec{r})$. Suitable model forms are the subject of chapter 6 of this thesis. The source brightness distribution is also parameterized, for example as a set of objects with specified fluxes, shapes, and locations. The deflector parameters and source parameters are then varied to find the potential $\Phi(\vec{r})$ that best matches the observed flux. The manner in which to do this is discussed in chapter 5 of this thesis. Note that the only constraints on $\Phi(\vec{r})$ from the image brightness distribution are on $\vec{\nabla}\Phi(\vec{r}_{\rm img})$ and $\frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\Phi(\vec{r}_{\rm img})$ at the locations $\vec{r}_{\rm img}$ where there is image flux, since the deflector gravitational potential is only sampled at the angular locations where there is image flux. The only direct constraint on $\Phi(\vec{r})$ in the regions without image flux is that $\Phi(\vec{r})$ must not produce images that would be seen but are not seen. In general, more images of more background sources better constrain $\Phi(\vec{r})$.

4.2.3 Lens potential — constraints from direct observation of the deflector

Constraints on the deflector mass distribution and its potential may also be found by direct observation of the deflector itself.

In particular, the velocity dispersion of the deflector galaxy may be used as a measure of the depth of the deflector's gravitational potential. This is particularly useful for lenses such as B0957+561 where the principal lensing galaxy is in a cluster of galaxies. The cluster, modeled using a uniform mass sheet, may add a significant amount of convergence κ to the lens. This, because of the mass sheet degeneracy, causes trouble in making a model prediction of the time delay, unless it can be determined how much of the lens's convergence is due to the mass sheet and how much is due to the primary deflector. Measuring the strength of the galaxy potential itself, via measurement of the galaxy's 1-dimensional velocity dispersion, $\sigma^2_{v_{1D}}$, allows the mass sheet degeneracy to be broken (Falco *et al.* 1985; Gorenstein *et al.* 1988; Narayan 1991). In fact,

$$(1-\kappa) = \sigma^2_{v_{1D}} \frac{D_{LS}}{D_S} \times C_{\substack{\text{model} \\ \text{dependent}}}. \tag{4.26}$$

The constant depends on the parameters of the lens gravitational potential — as fitted to the lens with the assumption of no mass sheet. It also depends on the shape of the distribution of luminous matter. Though it depends on the models used for the total matter distribution and the luminous matter distribution, it does not depend on cosmological parameters.

This may be readily seen for the case of the singular isothermal sphere,

$$\Phi_{\text{SIS}}(\vec{r}) = b|\vec{r}| = 4\pi\sigma^2_{v_{1D}} \frac{D_{LS}}{D_S}|\vec{r}|. \tag{4.27}$$

The apparent potential is determined by modeling the lens images as a singular

isothermal sphere with no mass sheet present,

$$\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}) = b_{\text{model}}|\vec{r}|, \tag{4.28}$$

where b_{model} is determined by modeling the lens images. The true potential of the galaxy is also assumed to be a self-gravitating singular isothermal sphere. Its depth is determined by observations of the velocity dispersion of the galaxy

$$\Phi_{\substack{\text{galaxy}\\ \text{true}}} = 4\pi\sigma_{\substack{v\\ 1D}}^2 \frac{D_{LS}}{D_S}|\vec{r}|. \tag{4.29}$$

From equation 4.22 the true galaxy potential $\Phi_{\substack{\text{galaxy}\\ \text{true}}}$ and the apparent potential $\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}$ are related by

$$\Phi_{\substack{\text{galaxy}\\ \text{true}}} = (1-\kappa)\Phi_{\substack{\text{model},\\ \text{no}\,\kappa}} \tag{4.30}$$

in the presence of a mass sheet $\kappa\sigma_{\text{crit}}$ which is not taken into account in modeling $\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}$. Therefore for this case, comparing equations 4.28 and 4.29 it can be seen that

$$(1-\kappa) = \sigma_{\substack{v\\ 1D}}^2 \frac{D_{LS}}{D_S}\frac{4\pi}{b_{\text{model}}}. \tag{4.31}$$

The constant of proportionality between $(1-\kappa)$ and $\sigma_{\substack{v\\ 1D}}^2 \frac{D_{LS}}{D_S}$ can also be worked out for other models for the gravitational potential. For example, if the lens is modeled by a singular isothermal sphere with a lens Einstein radius of b_{model}, and if the luminous matter is assumed to be an isothermal tracer population sitting in the potential and having a mass distribution $\rho_{\text{lum}} \propto \frac{1}{r^n}$, then

$$(1-\kappa) = \sigma_{\substack{v\\ 1D}}^2 \frac{D_{LS}}{D_S}\frac{2\pi n}{b_{\text{model}}}. \tag{4.32}$$

Therefore the breaking of the degeneracy by use of the velocity dispersion is model dependent, and depends on assumptions of how the luminous and dark matter in the galaxy are distributed.

The lensing galaxy's radial profile, ellipticity, and positions may also be used to select a suitable potential model and constrain the model parameters. However,

relating properties of the inner regions of the galaxy, as sampled by its light, to properties of the regions sampled by lensing well beyond the bright central regions of the galaxy, is model dependent.

4.2.4 Using the lens potential to find the surface mass density $\sigma(\vec{r})$, the effective lens distance $D_{\rm eff}$, and the angular diameter distance to the deflector D_L

The surface mass density of the lens deflector may be found from the divergence of the 2-D lens potential if it is well constrained. If observations of the luminous matter in the deflector are also available, then the surface mass density deduced from the lensing can be compared with them. A discrepancy would indicate the location of dark matter. In particular, the surface mass density is (equation 4.9)

$$\sigma(\vec{r}) = \frac{1}{2}\sigma_{\rm crit}\nabla^2_{\vec{r}}\Phi(\vec{r}) = \frac{1}{2}\frac{1}{4\pi G}\frac{D_S}{D_L D_{LS}}\nabla^2_{\vec{r}}\Phi(\vec{r}). \tag{4.33}$$

Even if the source or lens redshifts are unknown, or if the Hubble parameter is not well known, or if there is some unknown uniform mass sheet in the lens system, the angular dependence of $\sigma(\vec{r})$ is given by the divergence of the modeled $\Phi(\vec{r})$.

Measurements of a time delay and the deflector redshift, combined with the modeled potential, give a measure of the distance scales in the system. The time delay between two images α and β of the same source is (from equation 4.20)

$$t(\vec{r}_\alpha) - t(\vec{r}_\beta) = \\ (1+z_L)\frac{D_L D_S}{D_{LS}}\left(\frac{1}{2}|\vec{\nabla}_{\vec{r}_\alpha}\Phi(\vec{r}_\alpha)|^2 - \frac{1}{2}|\vec{\nabla}_{\vec{r}_\beta}\Phi(\vec{r}_\beta)|^2 - \Phi(\vec{r}_\alpha) + \Phi(\vec{r}_\beta)\right). \tag{4.34}$$

If the model deflector mass distribution contains a sheet of constant mass density as may be used for a galaxy cluster, $\sigma_{sheet} = \kappa\sigma_{\rm crit}$, then this is undetectable via the image geometry. Let $\Phi_{\substack{\rm model \\ \rm no\,\kappa}}$ be the potential as modeled from the image geometry under the

assumption that there is no such sheet. Then the time delay is (equation 4.25)

$$t(\vec{r}_\alpha) - t(\vec{r}_\beta) = \\ (1-\kappa)(1+z_L)\frac{D_L D_S}{D_{LS}} \\ \times \left(\frac{1}{2}|\vec{\nabla}_{\vec{r}_\alpha}\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\alpha)|^2 - \frac{1}{2}|\vec{\nabla}_{\vec{r}_\beta}\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\beta)|^2 - \Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\alpha) + \Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\beta)\right). \quad (4.35)$$

The deflector's luminous matter velocity dispersion and radial profile can be used to break the degeneracy between the mass sheet and the primary deflector, so that the time delay is

$$t(\vec{r}_\alpha) - t(\vec{r}_\beta) = \\ (1+z_L)C_{\substack{\text{model}\\ \text{dependent}}}\sigma_{v\,1D}^2 D_L \\ \times \left(\frac{1}{2}|\vec{\nabla}_{\vec{r}_\alpha}\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\alpha)|^2 - \frac{1}{2}|\vec{\nabla}_{\vec{r}_\beta}\Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\beta)|^2 - \Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\alpha) + \Phi_{\substack{\text{model}\\ \text{no}\,\kappa}}(\vec{r}_\beta)\right), \quad (4.36)$$

where $C_{\substack{\text{model}\\ \text{dependent}}}$ is a model dependent constant that has no dependence on the cosmological model used for the universe.

Consider a lens system for which a time delay and the deflector redshift are known and for which Φ has been modeled based on the image brightness distribution. If the lens system is thought to contain a component that approximates a mass sheet, such as a galaxy cluster, then the degeneracy between the mass sheet and the primary (galaxy) deflector may be broken by measurement of the primary deflector's velocity dispersion and luminous matter radial profile, as discussed in section 4.2.3. Equation 4.36 then gives a measure of the angular diameter distance to the deflector D_L, in one step without the cosmic distance ladder, as pointed out by Narayan (1991). If no mass sheet component to the deflector is observed, so that it is not appropriate to include a uniform mass sheet as part of the deflector model, then $\kappa = 0$. Equation 4.35 then gives a measure of $\frac{D_L D_S}{D_{LS}}$ in one step, with no dependence on the cosmic distance ladder. This ratio $\frac{D_L D_S}{D_{LS}}$ is sometimes called the "effective lens distance" $D_{\text{eff}} \equiv \frac{D_L D_S}{D_{LS}}$, but this is a misnomer, since the ratio $\frac{D_L D_S}{D_{LS}}$ does not correspond to any particular distance in the lens system.

4.3 Applications to cosmology

4.3.1 The standard cosmology

The Robertson-Walker line element for a homogeneous and isotropic spacetime is (Misner *et al.* 1973, equation 27.22)

$$ds^2 = -dt^2 + a^2(t)\left\{d\chi^2 + S^2(\chi)d\Omega^2\right\}, \tag{4.37}$$

where the function S depends on the curvature $\kappa_{\rm curv}$ of the universe:

$$S(\chi) = \begin{cases} \frac{1}{\sqrt{\kappa_{\rm curv}}}\sin(\chi\sqrt{\kappa_{\rm curv}}) & \text{for } \kappa_{\rm curv} > 0 \\ \chi & \text{for } \kappa_{\rm curv} = 0 \\ \frac{1}{\sqrt{-\kappa_{\rm curv}}}\sinh(\chi\sqrt{-\kappa_{\rm curv}}) & \text{for } \kappa_{\rm curv} < 0 \end{cases} \tag{4.38}$$

The change of the scale factor $a(t)$ with time is described by the Friedmann equation which, for a matter-dominated universe, is (Misner *et al.* 1973, equations 27.74 and 27.75)

$$\frac{1}{a}\frac{da}{dt} = H_o\left\{\Omega_o\left(\frac{a_o}{a}\right)^3 + \Omega_{\kappa o}\left(\frac{a_o}{a}\right)^2 + \Omega_{\Lambda o}\right\}^{\frac{1}{2}}. \tag{4.39}$$

This has three independent parameters and one dependent parameter:

- Hubble constant H_o
- Average mass density of the universe $\Omega_o \equiv \frac{8\pi G\rho_o}{3H_o^2}$
- Curvature of the universe $\Omega_{\kappa o} \equiv \frac{-\kappa_{\rm curv}}{H_o^2 a_o^2}$
- Cosmological constant $\Omega_{\Lambda o} \equiv \frac{\Lambda}{3H_o^2} = 1 - \Omega_o - \Omega_{\kappa o}$

The angular diameter distance from A to B, D_{AB}, is the physical length of an object located at B divided by the angle subtended at A. In a Friedmann-Robertson-Walker matter-dominated universe, the angular diameter distance from co-moving source A to co-moving source B, derived from the Friedmann equation and the Robertson-

Walker metric, is

$$D_{AB} = \frac{a_o}{1+z_B} S\left(\frac{1}{a_o H_o}\int_{1+z_A}^{1+z_B} \frac{du}{\sqrt{\Omega_o u^3 + \Omega_{\kappa o} u^2 + \Omega_{\Lambda o}}}\right). \tag{4.40}$$

4.3.2 Dependence of $D_{\rm eff}$, D_L, and $\sigma_{\rm crit}$ on the standard cosmology

It was shown above that the quantities D_L and $D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$ may be accessible from measurements of individual gravitational lenses. $D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$ is a function of the source and deflector redshifts, as illustrated in figure 4-3. On the other hand, D_L, as shown in figure 4-4, is a function of the deflector, but not the source, redshift. In both cases the functional form depends on the cosmological model. Both are proportional to ${H_o}^{-1}$, the inverse of the Hubble parameter. For low source and lens redshifts, both $D_{\rm eff}$ and D_L have very little sensitivity to the other cosmological parameters Ω_o and $\Omega_{\Lambda o}$.

The measurement of a single gravitational lens, yielding $D_{\rm eff}$ or D_L, is sufficient to give a measure of the Hubble parameter. Of course there will be some contribution to the uncertainty in H_o from the lack of knowledge of Ω_o and $\Omega_{\Lambda o}$. Note that $D_{\rm eff}$ has less sensitivity to Ω_o and $\Omega_{\Lambda o}$ than does D_L, so that $D_{\rm eff}$ would introduce less error into H_o from the other cosmological parameters than would D_L. However, there is always the possibility of the mass sheet degeneracy introducing systematic error into the measurement of H_o from $D_{\rm eff}$. In comparison, D_L requires the knowledge of the velocity dispersion and radial profile of the luminous matter as well as the shape of the potential well, so it too is subject to systematic problems. The possibility of errors in the determination of H_o due to some peculiarity of a lens makes it desirable to measure many lenses.

The goal of measuring Ω_o and $\Omega_{\Lambda o}$ also makes it desirable to measure many lenses. The dependence of D_L and $D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$ on the cosmological parameters Ω_o and $\Omega_{\Lambda o}$ becomes more pronounced for large redshifts. Indeed the shape of the curve D_L vs. z_L depends on Ω_o and $\Omega_{\Lambda o}$, whereas the normalization depends on H_o. Similarly,

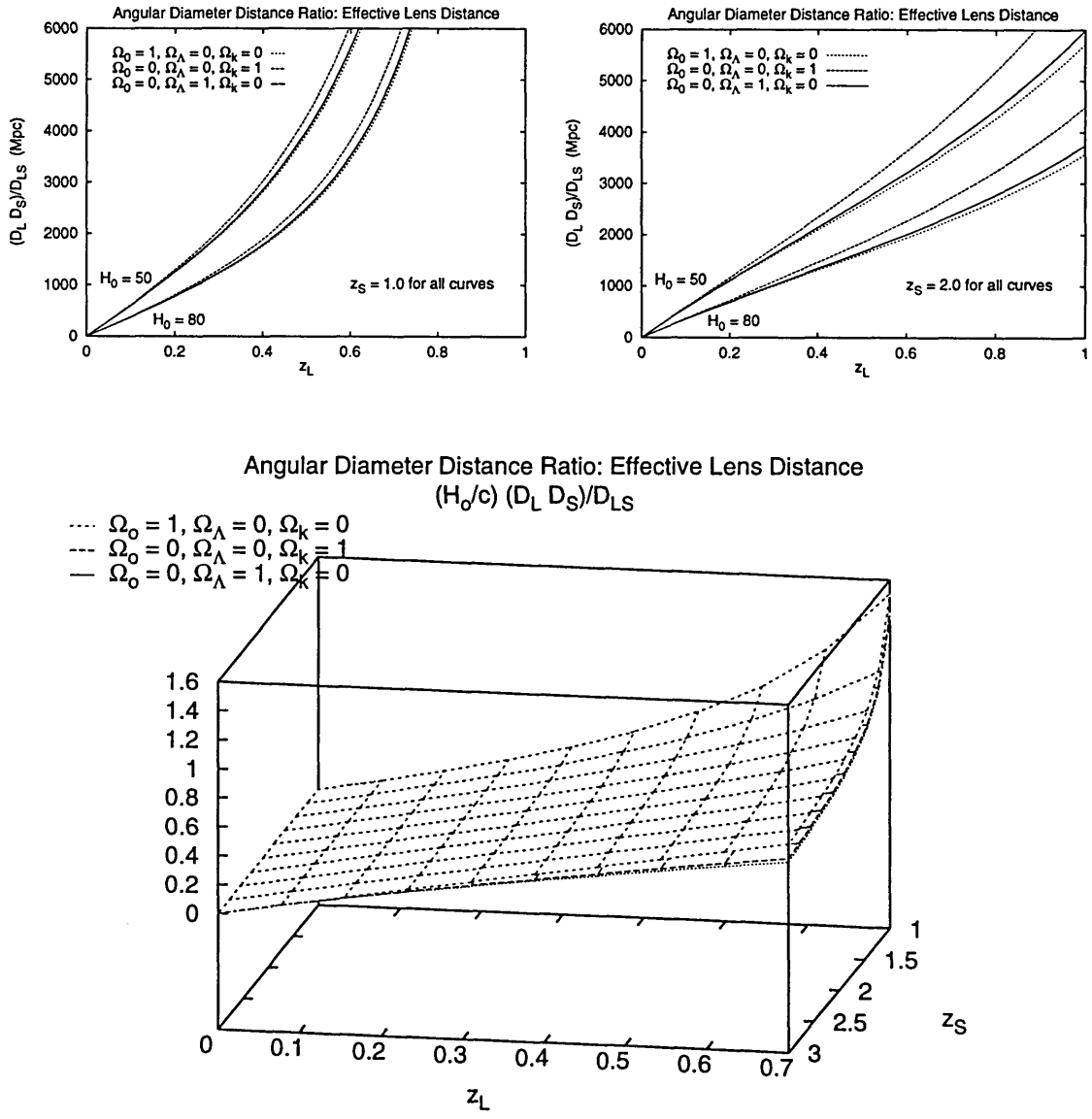

Figure 4-3: The dependence of the effective lens distance $D_{\text{eff}} = \frac{D_L D_S}{D_{LS}}$ on redshift. The curves depend on the Hubble parameter and the other cosmological parameters. The ordinate in the lower plot has been made dimensionless by multiplying $\frac{D_L D_S}{D_{LS}}$ by $\frac{1}{c} H_o$; therefore the lower plot illustrates only the dependence on Ω_o and $\Omega_{\Lambda o}$.

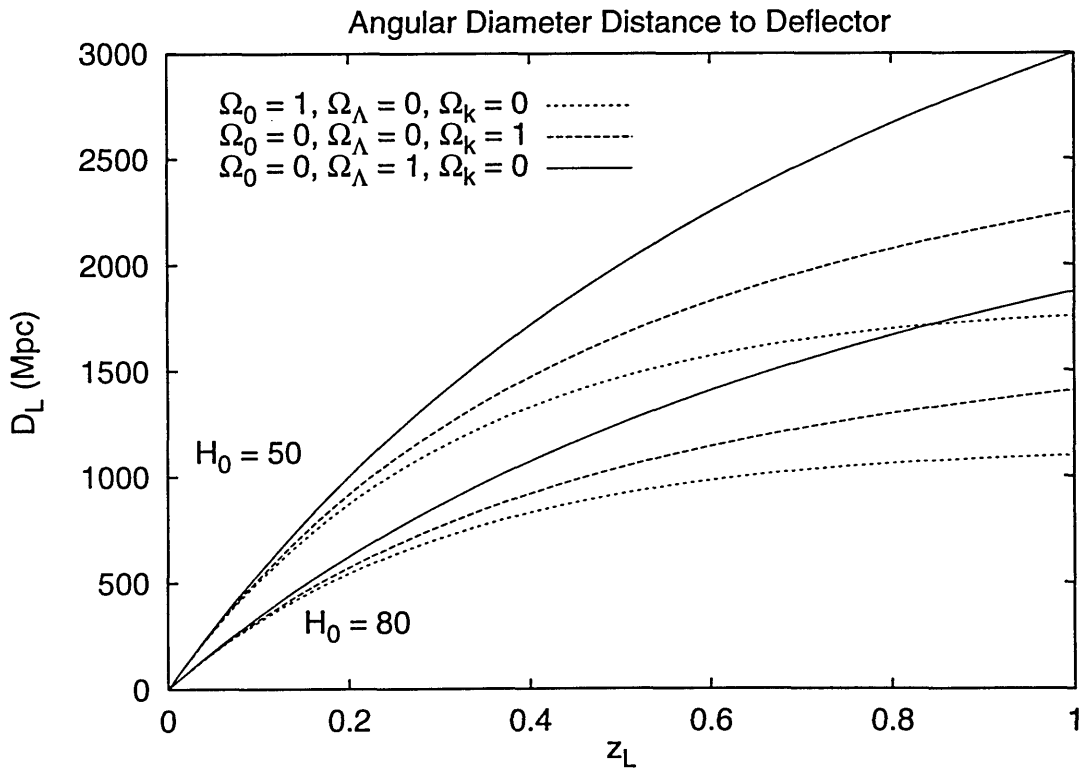

Figure 4-4: The angular diameter distance to the deflector, D_L is a function of the lens redshift. The curves show how this function depends on the Hubble parameter and on the other cosmological parameters.

the shape of the surface $D_{\rm eff}$ vs. z_L and z_S depends on Ω_o and $\Omega_{\Lambda o}$, whereas the normalization depends on H_o. If enough lenses at a variety of redshifts are measured, then it is possible in principle to trace out the shape of the surface or curve, thus determining Ω_o and $\Omega_{\Lambda o}$ as well as the normalization H_o.

$D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$ has some dependence on the curvature $\Omega_{\kappa o}$, becoming more pronounced with higher source and lens redshifts, but has very little sensitivity to the balance between the mass density Ω_o and the cosmological constant $\Omega_{\Lambda o}$ If the universe is assumed to be flat, $\Omega_{\kappa o} = 0$, as is popular for theoretical reasons, then the balance between Ω_o and $\Omega_{\Lambda o}$ will have little effect on the estimates of the Hubble parameter.

In contrast to the effective lens distance $D_{\rm eff}$, the distance D_L is more sensitive to all the cosmological parameters Ω_o, $\Omega_{\Lambda o}$, and $\Omega_{\kappa o}$. In particular it is more sensitive to the balance between the mass density Ω_o and the cosmological constant $\Omega_{\Lambda o}$. Were the Hubble parameter well known, this would be more sensitive for constraining the other cosmological parameters.

Another way of using lensing to constrain cosmological parameters is through statistics on the probability of lensing, comparing the number of lenses seen with the probability of lensing by some set of deflector objects. This probability of lensing depends on the relation of the deflector masses to the critical surface mass density $\sigma_{\rm crit} = \frac{1}{4\pi G}\frac{D_S}{D_L D_{LS}}$. Since the dependence of $\sigma_{\rm crit}$ on the deflector and source redshifts z_L and z_S (figure 4-5) itself depends on the cosmological parameters H_o, Ω_o and $\Omega_{\Lambda o}$, such probability-of-lensing methods may be used to constrain these parameters.

There is another way of using lensing to constrain the cosmological parameters Ω_o and $\Omega_{\Lambda o}$ which in contrast to the methods above does not require knowledge of the Hubble parameter (Paczynski & Gorski 1981; Breimer & Sanders 1992). Comparing the strength of the deflector potential as determined by lens modeling with the strength of the deflector potential as measured by its velocity dispersion under the assumption that there is no mass sheet component to the deflector, gives (from equation 4.26)

$$\frac{D_S}{D_{LS}} = \sigma_{v\ 1D}^2 \times C_{\substack{\rm model \\ \rm dependent}}. \tag{4.41}$$

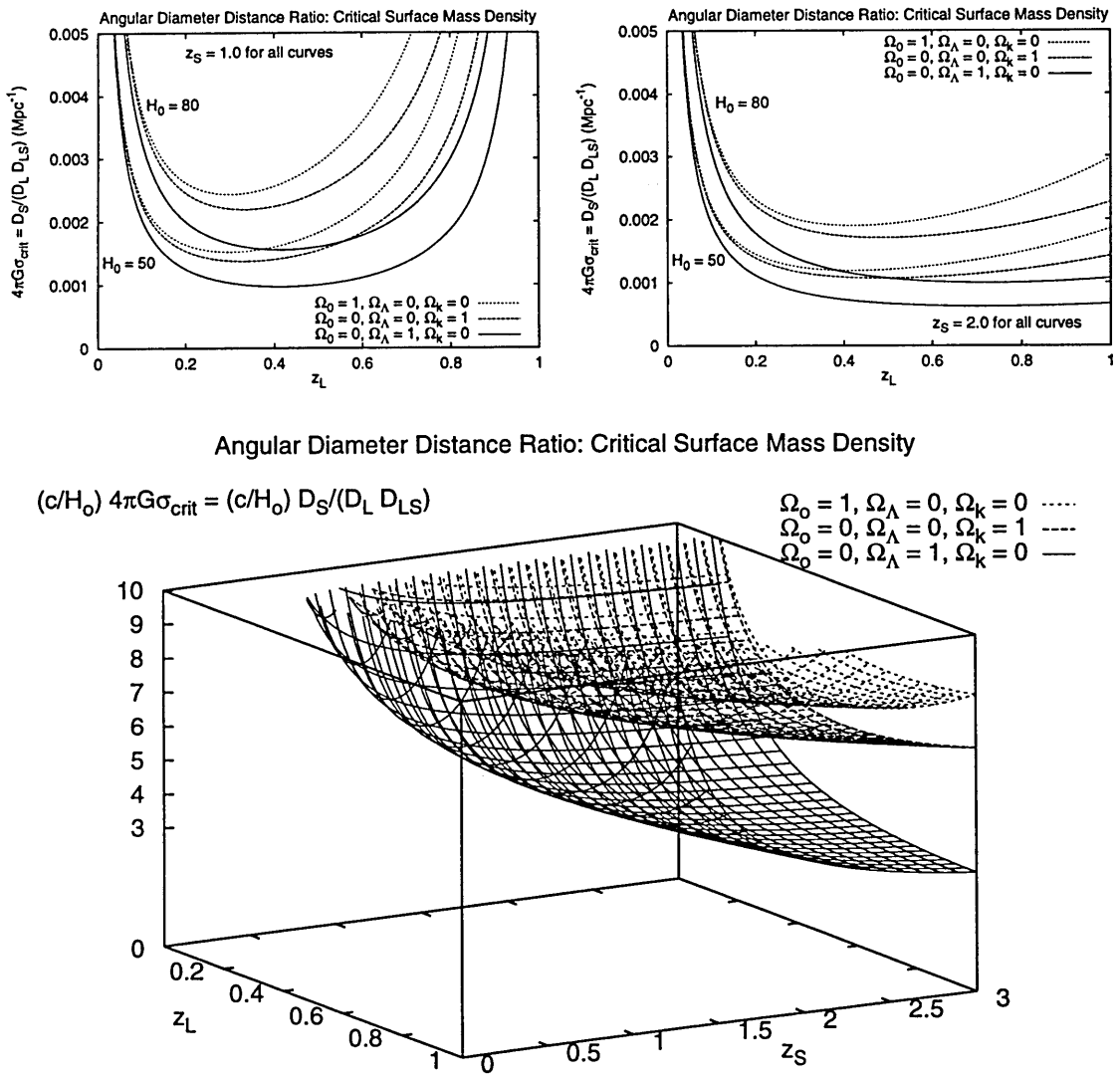

Figure 4-5: The dependence of the critical surface mass density for lensing $\sigma_{\rm crit}$ on the source and lens redshifts. The different surfaces illustrate the dependence on the cosmological parameters Ω_o and $\Omega_{\Lambda o}$.

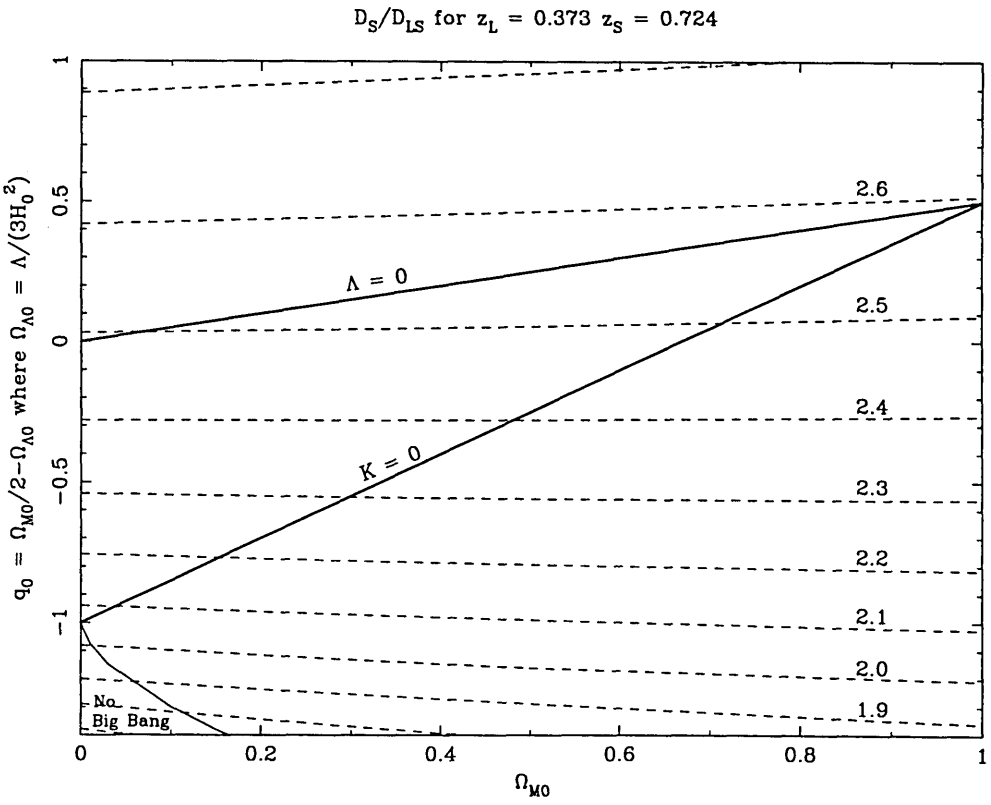

Figure 4-6: This contour plot of $\frac{D_S}{D_{LS}}$ shows the sensitivity of the angular diameter distance ratio to the cosmological parameters $\Omega_{Mo} \equiv \Omega_o$ and $\Omega_{\Lambda o}$, for fixed source and lens redshifts. (These redshifts are for the giant lensed arc, $z_S = 0.724$, (Soucail *et al.* 1988), in the cluster Abell 370, $z_L = 0.373$.) Note that the ratio is particularly sensitive to the deceleration parameter $q_0 = \frac{1}{2}\Omega_o - \Omega_{\Lambda o}$. The region below the curve at the lower corner of the plot parameterizes universes which did not evolve from an initial singularity, that is, ones which had $\frac{da}{d\tau} = 0$ for a finite scale factor a smaller than the present value (Carroll *et al.* 1992).

Equivalently, for a deflector which is itself a cluster of galaxies, its x-ray gas temperature may be used in place of the velocity dispersion

$$\frac{D_S}{D_{LS}} = T_{\rm gas} \times C'_{\substack{\rm model \\ \rm dependent}}. \tag{4.42}$$

$C_{\substack{\rm model \\ \rm dependent}}$ is a model dependent constant depending on the radial profile of the luminous matter and the shape of the potential well but having no dependence on the cosmological model used for the universe. Note that the ratio $\frac{D_S}{D_{LS}}$ is particularly sensitive to the deceleration parameter $q_o = \frac{1}{2}\Omega_o - \Omega_{\Lambda o}$. This is shown in figure 4-6.

4.4 Summary: requirements for determining the deflector mass distribution and the Hubble parameter and other cosmological parameters

Deflector mass distribution

The study of the image geometry of a single lens may be enough to give the angular dependence of the mass of that lens, $\sigma(\vec{r}) \propto \nabla^2\Phi(\vec{r})$. Combined with knowledge of the source and deflector redshifts, an assumed cosmology, and an assumption concerning the mass sheet degeneracy, this gives as absolute distribution of the mass. This distribution of all the matter, dark plus luminous, is of interest for comparing the distribution of dark matter in the deflector with the distribution of luminous matter.

Hubble parameter

Measurement of a single gravitational lens may give a measurement of the angular diameter distances in the lens system independent of the cosmic distance ladder. This gives a point-estimate of the Hubble parameter, independent of the cosmic distance ladder, but with some uncertainty due to Ω_o and $\Omega_{\Lambda o}$. To estimate the Hubble parameter, a number of things must be known about the lens system:

- The time delay between two images must have been measured.

- The deflector redshift must be known.
- There must be a good model for the 2-D lensing potential, from maps, and/or from observations of the deflector itself. One must
  - assume that there is no component to the deflector which approximates a mass sheet. (In this case one has determined $D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$.)
  - or measure the velocity dispersion of the lens galaxy to distinguish between the convergence due to the lens galaxy and convergence due to the mass sheet. (In this case one has determined D_L.)
- The source redshift as well as the deflector redshift must be known to deduce the Hubble parameter from angular diameter distances. (Only the deflector redshift need be known if the velocity dispersion of the principal deflector is being used to break the degeneracy.)

So far such measurements have been made for B0957+561 (Grogin & Narayan 1996a; Grogin & Narayan 1996b; Haarsma *et al.* 1997; Haarsma 1997; Kundic *et al.* 1997) and for PG B1115+080 (Schechter *et al.* 1997).

Cosmological parameters Ω_o and $\Omega_{\Lambda o}$

Measurements of many multiple-image lenses at various redshifts may, when they have been made, be used to trace out the dependence of $D_{\rm eff} = \frac{D_L D_S}{D_{LS}}$ on z_L and z_S or of D_L on z_L. The normalization of the curves will give the Hubble parameter H_o, and the shape of the curve will gives Ω_o and $\Omega_{\Lambda o}$.

Chapter 5

Lens Modeling: Using Observations to Constrain Models of the Deflector

5.1 Introduction

To model a gravitational lens system one needs two pieces. First is needed a model of the deflector, either its mass distributions or else its gravitational potential. One also needs a model for the source: its sky-brightness as it would be in the absence of the intervening deflector.

For purposes of studying the cosmological parameters, including H_o, it is the model of the deflector that is of particular importance. The deflector potential, as well as the deflector and source redshifts and the time delay between two images, must be known in order to constrain H_o. The model of the deflector potential also provides information on the mass distribution of the deflector. This is direct information on the matter's mass from its gravitational effects, rather than indirect information from the matter's luminosity.

This chapter discusses how one may use the observed sky brightness of a lens system to provide constraints on a model of the deflector.

The model sky brightness, mapped through the model deflector and then convolved with the detector response, must match the actual observed map of the lens — at least to within the noise of the observation. The flux in every pixel of the map provides constraints on the system, and all must be in agreement with the model for the model not to be rejected.

In practise, one would like to explore a wide range of lens models. For this purpose one would like a computationally fast way of rejecting the models that cannot come close to producing the images seen. This can be done if one can capture the most-important information from the observed images into just a few constraints. For example, the VLBA 6 cm maps of MG J0414+0534 (figure 3-6) show that each image (A1, A2, B, and C) of the background object has four resolved components. From the shapes of the components, it is clear which component of A1 corresponds to which components of A2, of B, and of C. A deflector model for MG J0414+0534 must be able to correctly account for the positions of all these components. A deflector model that correctly accounts for the positions of the components may or may not also correctly account for the component fluxes, the components' shapes, and the extended flux. But any deflector model that is rejected by the component positions alone need not be considered further when all the constraints, the flux in each pixel, are used.

5.2 "Point" modeling — using positions but not fluxes

The simplest approach is "point" modeling. One identifies some number N_l of features, or knots, or components in the background source, and the images of each. The positions of the images are used as constraints. The source model consists merely of N_l point sources, of unspecified flux. The requirement for the deflector model is that it correctly map the sources' positions onto the image positions.

More precisely, the modeling is done as follows:

From the map of the gravitational lens system, one identifies N_l features, each of

which has N_α images. The $N_l \times N_\alpha$ measured right-ascensions and $N_l \times N_\alpha$ measured declinations provide $2 \times N_l \times N_\alpha$ model constraints,

$$\vec{r}_{l\alpha} \text{ for } \begin{cases} l = 1, \ldots, N_l \text{ independent sources} \\ \alpha = 1, \ldots, N_\alpha \text{ images of each source.} \end{cases} \tag{5.1}$$

Each measured image location has its own estimate of the measurement error,

$$S_{l\alpha} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix}. \tag{5.2}$$

The source model consists of the locations of N_l point sources:

$$\vec{s}_l \text{ for } l = 1, \ldots, N_l \text{ independent sources.} \tag{5.3}$$

The deflector model is assumed to take the form of a 2-D lens potential $\Phi_{\{a\}}(\vec{r})$ that depends on some set of N_a parameters $\{a\}$.

The deflector model may be used to map positions from the source-plane to the image-plane

$$\vec{s}(\vec{r}) = \vec{r} - \vec{\nabla}_{\vec{r}} \Phi_{\{a\}}(\vec{r}). \tag{5.4}$$

The inverse mapping is not unique, since each source-plane position can map to multiple image-plane positions. We denote the position of the α^{th} image of a particular source location $\vec{s}$ by

$$\vec{r}_\alpha(\vec{s}). \tag{5.5}$$

Assuming that the errors are Gaussian, the maximum-likelihood statistic is the familiar chi-squared

$$\chi^2_{\text{positions}} = \sum_{\substack{\text{sources} \\ l}} \sum_{\substack{\text{images} \\ \alpha}} (\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}) \cdot \mathbf{S}_{l\alpha}^{-1} \cdot (\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}). \tag{5.6}$$

The chi-squared is minimized with respect to the model source locations $\vec{s}_l$ and the model deflector parameters $\{a\}$. (These model parameters enter the chi-squared

expression only in the mapping $\vec{r}_\alpha(\vec{s}_l)$.) The minimum of the chi-squared gives the maximum-likelihood estimate for the best-fit model parameters $\vec{s}_l$ and $\{a\}$.

5.2.1 Mathematical simplification: the source-plane approximation

The constraints on the model are the measured image positions, which are image-plane positions. The model source parameters are source-plane positions. To compute the chi-squared (equation 5.6), the model source-plane positions must be mapped to the image plane.

Source-plane positions can easily be computed from image-plane positions, for a given potential: $\vec{s}(\vec{r}) = \vec{r} - \vec{\nabla}_{\vec{r}}\Phi(\vec{r})$. But image-plane positions cannot easily be computed from source-plane positions. The equation cannot be analytically inverted for a general potential. Indeed in the interesting (strong lensing) case the inversion is non-unique. Each source-plane positions maps to multiple image-plane positions, and these solutions must be found numerically.

If the form of the model is such that a good fit can be obtained, then $(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha})$ is small near the minimum of the chi-squared. A Taylor expansion[1] may be made for $\vec{s}(\vec{r})$ near an image positions $\vec{r}_{l\alpha}$:

$$s_i(\vec{r}) = s_i(\vec{r}_{l\alpha}) + (r_j - r_{l\alpha,j}) \left.\frac{\partial s_i}{\partial r_j}\right|_{\vec{r}_{l\alpha}} + (r_j - r_{l\alpha,j})(r_k - r_{l\alpha,k}) \left.\frac{\partial^2 s_i}{\partial r_j \partial r_k}\right|_{\vec{r}_{l\alpha}} + \ldots . \quad (5.7)$$

The partial derivative is simply the inverse of the magnification matrix $\frac{\partial s_i}{\partial r_j} = {M^{-1}}_{ij}$. This Taylor expansion can be rewritten in term of the magnification matrix:

$$s_i(\vec{r}) = s_i(\vec{r}_{l\alpha}) + (r_j - r_{l\alpha,j}) \left.{M^{-1}}_{ij}\right|_{\vec{r}_{l\alpha}} + (r_j - r_{l\alpha,j})(r_k - r_{l\alpha,k}) \left.\frac{\partial}{\partial r_k}{M^{-1}}_{ij}\right|_{\vec{r}_{l\alpha}} + \ldots . \quad (5.8)$$

[1]Throughout this chapter, the coordinate indices i, j, k, and p appearing twice in a term indicate implied summation, in the usual manner. Note that all sums over sources l and images α are shown explicitly.

Evaluating this at $\vec{r} = \vec{r}_\alpha(\vec{s}_l)$, and using vector notation,

$$\vec{s}_l - \vec{s}(\vec{r}_{l\alpha}) = \tag{5.9}$$
$$M^{-1}(\vec{r}_{l\alpha})(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}) + \left((\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}) \cdot \vec{\nabla}_{\vec{r}} M^{-1}(\vec{r}_{l\alpha})\right)(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}) + \ldots.$$

Near the chi-squared minimum, for a well-fitting model, the model image position $\vec{r}_\alpha(\vec{s}_l)$ is near the observed image positions $\vec{r}_{l\alpha}$, so the expansion can be truncated with little error. As one can see, this amounts to assuming that the change in magnification is negligible between the observed image location $\vec{r}_{l\alpha}$ and the model image location $\vec{r}_\alpha(\vec{s}_l)$:

$$(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}) \cdot \vec{\nabla}_{\vec{r}} M^{-1}(\vec{r}_{l\alpha}) = \mathcal{O}(\epsilon') \tag{5.10}$$

An approximation is adequate as long as it is good near the chi-squared minimum, because we are unconcerned about the behavior of the function far from the minimum, unless new lower minima are introduced. Truncating the approximation to first order,

$$\vec{s}_l - \vec{s}(\vec{r}_{l\alpha}) \approx M^{-1}(\vec{r}_{l\alpha})(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}), \tag{5.11}$$
$$M(\vec{r}_{l\alpha})(\vec{s}_l - \vec{s}(\vec{r}_{l\alpha})) \approx \vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}. \tag{5.12}$$

This yields an approximation to $\chi^2_{\text{positions}}$ in the source plane

$$\chi^2_{\substack{\text{positions} \\ \text{srcplane}}} = \sum_{\substack{\text{sources} \\ l}} \sum_{\substack{\text{images} \\ \alpha}} (\vec{s}_l - \vec{s}(\vec{r}_{l\alpha}))\, M(\vec{r}_{l\alpha}) \cdot \mathbf{S}^{-1}_{l\alpha} \cdot M(\vec{r}_{l\alpha})\, (\vec{s}_l - \vec{s}(\vec{r}_{l\alpha}))$$
$$\text{where } \vec{s}(\vec{r}_{l\alpha}) = \vec{r}_{l\alpha} - \vec{\nabla}_{\vec{r}_{l\alpha}} \Phi_{\{a\}}(\vec{r}_{l\alpha}), \tag{5.13}$$

where the components of the inverse magnification matrix are $M^{-1}{}_{ij}(\vec{r}) = \delta_{ij} - \frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\Phi_{\{a\}}(\vec{r})$. This expression for the chi-squared is simple to compute, provided that the first and second derivatives of the model for the deflector potential are easy to calculate. A quickly computable chi-squared is desirable regardless of the minimization algorithm used to minimize the chi-squared. Note that the global minimum of each function, $\chi^2_{\text{positions}}$ and $\chi^2_{\substack{\text{positions}, \\ \text{srcplane}}}$, is zero in the case that the model exactly reproduces the observation. We expect that even with noise and measurement error,

the source-plane approximation to the chi-squared, $\chi^2_{\substack{\text{positions}\\\text{srcplane}}}$, has a global minimum which corresponds to the minimum of the true image-plane chi-squared, $\chi^2_{\text{positions}}$, and that no lower minimum is introduced by this approximation.

The "source-plane" approximation assumes that the change in magnification is small between the observed image location and the best-fit model's image location. Therefore the validity of the "source-plane" approximation depends on the model being a sufficiently good fit; if the deflector model perfectly reproduced the observed image locations, this approximation would be fine. The source-plane approximation can fail in two cases. The first case is when the deflector model is so poor that the model cannot reproduce the observed positions, but in that situation the model would be rejected anyway. The "source-plane" approximation can fail in a more subtle fashion if the error on the observed image locations is large enough that the magnification changes significantly within the error region. In such a case, the model's χ^2_{srcplane} could be low enough that the model would not be rejected, yet the magnification at the model image locations could be significantly different from the magnification at the true image locations.

5.2.2 An analytic expression for the best-fit source positions in the source-plane approximation

The source-plane approximation to the chi-squared (equation 5.13) is quadratic in the $2 \times N_k$ model source positions, $\vec{s}_l$. Therefore the minimization over these parameters may be done analytically. The value of $\vec{s}_l$ at the minimum may be found by setting $\vec{\nabla}_{\vec{s}_l} \chi^2_{\substack{\text{positions}\\\text{srcplane}}}$ to zero and solving for $\vec{s}_l$:

$$0 = \vec{\nabla}_{\vec{s}_l} \chi^2_{\substack{\text{positions}\\\text{srcplane}}} = 2 \sum_{\substack{\text{images}\\\alpha}} \boldsymbol{M}(\vec{r}_{l\alpha}) \cdot \mathbf{S}^{-1}_{l\alpha} \cdot \boldsymbol{M}(\vec{r}_{l\alpha}) \left(\vec{s}_l - \vec{s}(\vec{r}_{l\alpha})\right), \tag{5.14}$$

so that the model positions that minimize the chi-squared are

$$\vec{s}_l = \left[\sum_{\substack{\text{images} \\ \alpha}} M(\vec{r}_{l\alpha}) \cdot \mathbf{S}_{l\alpha}^{-1} \cdot M(\vec{r}_{l\alpha}) \right]^{-1} \left[\sum_{\substack{\text{images} \\ \alpha}} M(\vec{r}_{l\alpha}) \cdot \mathbf{S}_{l\alpha}^{-1} \cdot M(\vec{r}_{l\alpha}) \vec{s}(\vec{r}_{l\alpha}) \right] \tag{5.15}$$

where $\vec{s}(\vec{r}_{l\alpha}) = \vec{r}_{l\alpha} - \vec{\nabla}_{\vec{r}_{l\alpha}} \Phi_{\{a\}}(\vec{r}_{l\alpha})$.

This analytical minimization reduces the dimensionality of the parameter space from $(2 \times N_l) + N_a$ to merely N_a, where N_a is the number of parameters of the deflector model. For MG J0414+0534, where $N_l = 4$, this reduces the dimensionality of the parameter space by 8 dimensions. Therefore this technique is a significant aid to any minimization algorithm used to minimize the chi-squared.

5.2.3 Other simplifications to "point" modeling

Other simplifications to the "point" modeling chi-squared (equation 5.6) are discussed, for example, by Kochanek (1991). His aim was to simplify the chi-squared in such a way as to make the chi-squared quadratic in all model parameters in which the deflector's potential Φ was linear. This method of course made finding the parameters that minimize the chi-squared quite straightforward. However, in order to make the chi-squared quadratic, one must neglect the dependence of the magnification matrix $M(\vec{r}_{l\alpha})$ on the deflector model parameters $\{a\}$, (see equation 5.13), thereby distorting the weighting appropriate to the measured errors. This thesis does not use these sorts of approximations for the following reasons: (1) The approximations introduce some error into the weightings of the measured constraints. (2) This thesis uses more general deflector models which are nonlinear in some of their parameters, and a method was wanted that would work for any model for which $\vec{\nabla}\Phi$ and $\frac{\partial^2}{\partial r_i \partial r_j}\Phi$ could be calculated. (3) Ample computer power was available to do without the approximation.

5.3 "Point" modeling — using positions and fluxes

The point modeling technique may be extended to use image fluxes as well as image positions.

Again one identifies some number N_l of features, knots, or components in the background sources, and the images of each. The positions and the integral fluxes of the images are used as constraints. As before, the source model consists of N_l point sources, but in this case flux as well as position must be specified for each. The requirement for the deflector model is that it transform the model source flux into the observed image flux as well as correctly mapping sources positions onto image positions.

More precisely, the modeling is done as follows:

The N_l source features, each of which has N_α images, provide $3 \times N_l \times N_\alpha$ model constraints: x-position, y-position, and flux for each image of each source. The position constraints and error matrix are as in equations 5.1 and 5.2. The additional flux constraints are the fluxes of the lensed images, the "image-plane" fluxes

$$S^{\mathrm{I}}_{l\alpha} \text{ for } \begin{cases} l = 1, \ldots, N_l \text{ independent sources} \\ \alpha = 1, \ldots, N_\alpha \text{ images of each source.} \end{cases} \tag{5.16}$$

Each measured image flux has an error estimate $\sigma_{l\alpha}$.

The source model consists of locations and fluxes of the N_l point sources. The locations are as in equation 5.3. The fluxes are the unlensed source fluxes, the "source-plane" fluxes,

$$S^{\mathrm{S}}_{l} \text{ for } l = 1, \ldots, N_l \text{ independent sources.} \tag{5.17}$$

The type of deflector model used is the same as for point-modeling when flux constraints were not used, a 2-D lens potential $\Phi_{\{a\}}(\vec{r})$ parameterized by a set of N_a parameters $\{a\}$.

For point sources, the flux magnification is given by the determinant of the magnification at the model image location. Therefore the model image-plane flux to be

compared with the measured image-plane flux is:

$$S_l^{\rm S} \det \boldsymbol{M}(\vec{r}_\alpha(\vec{s}_l)). \tag{5.18}$$

Assuming that the errors are Gaussian, the maximum-likelihood statistic is the familiar chi-squared with the contribution of the fluxes added to the position-only point-modeling chi-squared

$$\chi^2 = \chi^2_{\rm positions} + \chi^2_{\rm flux}, \tag{5.19}$$

where $\chi^2_{\rm positions}$ is as given in equation 5.6 and where

$$\chi^2_{\rm flux} = \sum_{\substack{\rm sources \\ l}} \sum_{\substack{\rm images \\ \alpha}} \left\{ \frac{S_l^{\rm S} \det \boldsymbol{M}(\vec{r}_\alpha(\vec{s}_l)) - S_{l\alpha}^{\rm I}}{\sigma_{l\alpha}} \right\}^2. \tag{5.20}$$

As before, the chi-squared is minimized with respect to the model deflector parameters $\{a\}$ and the model source parameters. The minimum of the chi-squared gives the maximum-likelihood estimate for the best-fit model parameters, $\{a\}$, $\vec{s}_l$, and $S_l^{\rm S}$.

5.3.1 Mathematical simplification: the source-plane approximation

Reasonable approximations are again used to simplify the chi-squared computation. Taylor expanding around an observed image plane position $\vec{r}_{l\alpha}$, one finds

$$\det \boldsymbol{M}(\vec{r}') = \det \boldsymbol{M}(\vec{r}_{l\alpha}) + (\vec{r}' - \vec{r}_{l\alpha}) \cdot \vec{\nabla}_{\vec{r}} \det \boldsymbol{M}(\vec{r})\Big|_{\vec{r}_{l\alpha}} + \ldots. \tag{5.21}$$

As in section 5.2.1, this is evaluated at $\vec{r}' = \vec{r}_\alpha(\vec{s}_l)$ and terms of order $(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha}) \cdot \vec{\nabla} \det \boldsymbol{M}(\vec{r}_{l\alpha}) = \mathcal{O}(\epsilon')$ are neglected. This drops the higher order terms that involve the gradient of the magnification matrix, so that

$$\det \boldsymbol{M}(\vec{r}_\alpha(\vec{s}_l)) \approx \det \boldsymbol{M}(\vec{r}_{l\alpha}). \tag{5.22}$$

Then

$$\chi^2_{\rm flux} \approx \sum_{\substack{\rm sources \\ l}} \sum_{\substack{\rm images \\ \alpha}} \left\{ \frac{S^{\rm S}_l \det \boldsymbol{M}(\vec{r}_{l\alpha}) - S^{\rm I}_{l\alpha}}{\sigma_{l\alpha}} \right\}^2 . \tag{5.23}$$

This is plainly the same result as if we had mapped the observed fluxes and errors back to the source plane: $\chi^2_{\rm flux} \approx \chi^2_{\substack{\rm flux \\ \rm srcplane}}$ which is given by

$$\chi^2_{\substack{\rm flux \\ \rm srcplane}} = \sum_{\substack{\rm sources \\ l}} \sum_{\substack{\rm images \\ \alpha}} \left\{ \frac{S^{\rm S}_l - S^{\rm I}_{l\alpha} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha})}{\sigma_{l\alpha} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha})} \right\}^2 , \tag{5.24}$$

where ${\boldsymbol{M}^{-1}}_{ij}(\vec{r}) = \delta_{ij} - \frac{\partial}{\partial r_i}\frac{\partial}{\partial r_j}\Phi_{\{a\}}(\vec{r})$ is a 2×2 matrix that is already calculated in the course of calculating $\chi^2_{\substack{\rm positions \\ \rm srcplane}}$. (See equation 5.13.)

5.3.2 An analytic expression for the best-fit source fluxes in the source-plane approximation

When fluxes are used as constraints, the "point" modeling chi-squared is the sum of two pieces, one due the positions, and one due to the fluxes,

$$\chi^2_{\rm srcplane} = \chi^2_{\substack{\rm positions \\ \rm srcplane}} + \chi^2_{\substack{\rm flux \\ \rm srcplane}} , \tag{5.25}$$

where the expressions for $\chi^2_{\substack{\rm positions \\ \rm srcplane}}$ and $\chi^2_{\substack{\rm flux \\ \rm srcplane}}$ are given in equations 5.13 and 5.24. Only the second term of equation 5.25 depends on the model source fluxes. This second term is quadratic in the model fluxes. Therefore the minimization over the model fluxes may be done analytically. The value of $S^{\rm S}_l$ at the minimum may be found by setting $\vec{\nabla}_{S^{\rm S}_l}\chi^2_{\rm srcplane}$ to zero and solving for $S^{\rm S}_l$,

$$0 = \vec{\nabla}_{S^{\rm S}_l}\chi^2_{\rm srcplane} = \vec{\nabla}_{S^{\rm S}_l}\chi^2_{\substack{\rm flux \\ \rm srcplane}} = 2 \sum_{\substack{\rm images \\ \alpha}} \left\{ \frac{S^{\rm S}_l - S^{\rm I}_{l\alpha} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha})}{\left(\sigma_{l\alpha} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha})\right)^2} \right\} . \tag{5.26}$$

Thus the model fluxes that minimize the chi-squared are

$$S_l^{\mathrm{S}} = \frac{\sum_{\substack{\mathrm{images} \\ \alpha}} \left\{ \frac{S_{l\alpha}^{\mathrm{I}} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha})}{\left(\sigma_{l\alpha} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha})\right)^2} \right\}}{\sum_{\substack{\mathrm{images} \\ \alpha}} \left\{ \sigma_{l\alpha} \det \boldsymbol{M}^{-1}(\vec{r}_{l\alpha}) \right\}^{-2}} \ . \tag{5.27}$$

Since the second term in equation 5.25 does not depend of the model source positions $\vec{s}_l$, the $\chi^2_{\mathrm{srcplane}}$ is, like $\chi^2_{\substack{\mathrm{positions} \\ \mathrm{srcplane}}}$, quadratic in the model source positions. The minimization over the model source positions is done analytically, just as in section 5.2.2. The positions $\vec{s}_l$ that minimize $\chi^2_{\mathrm{srcplane}}$ are still the same as in equation 5.15.

Adding the fluxes as constraints increases the number of modeling constraints by $N_\alpha \times N_l$. It increases the number of model source parameters by N_l — one flux for each source object. However, these additional source parameters may all be found by analytical minimization of the chi-squared. Adding the fluxes as constraints does not increase the dimensionality of the parameter space to be searched by the algorithm used to minimize the chi-squared.

5.4 Corrections to "point" modeling for slightly extended sources in the presence of magnification gradients

"Point" modeling uses only the locations of each image, even for resolved objects, assuming that the centroids of the images of a source object correspond to each other. In other words, the centroid of the source flux is assumed to map to the centroid of the image flux under lensing. If this assumption is false, then "point" modeling introduces some error.

It will be shown below that if the magnification is constant across an image, the centroid of the image is the same as the image-plane position corresponding to the source centroid. For sources which are only somewhat extended, an expression for the displacement caused by a magnification gradient will be found,

5.4.1 Notation

Let the surface brightness distribution of a lensed image be denoted by $I_{\rm I}(\vec{r})$. The surface brightness distribution of the source, as it would appear in the absence of lensing, will be denoted by $I_{\rm S}(\vec{s})$. These surface brightness distributions are related by the lens equation; as surface brightness is unchanged by lensing,

$$I_{\rm S}(\vec{s}(\vec{r})) = I_{\rm I}(\vec{r}) \ , \ \text{ where } \ \vec{s}(\vec{r}) = \vec{r} - \vec{\nabla}\Phi(\vec{r}). \tag{5.28}$$

It is also useful to introduce the total flux of an image, where the integral is understood to extend only over that region where that image has non-zero flux,

$$S_{\rm I} = \int d^2r \ I_{\rm I}(\vec{r}). \tag{5.29}$$

Similarly, the flux of the source as it would be in the absence of lensing is

$$S_{\rm S} = \int d^2s \ I_{\rm S}(\vec{s}). \tag{5.30}$$

The centroid of the lensed emission, also referred to as the image centroid, is

$$\vec{r}_o = \frac{1}{S_{\rm I}} \int d^2r \ \vec{r} I_{\rm I}(\vec{r}). \tag{5.31}$$

If the emission were not lensed, the centroid would be the source centroid,

$$\vec{s}_o = \frac{1}{S_{\rm S}} \int d^2s \ \vec{s} I_{\rm S}(\vec{s}). \tag{5.32}$$

The extent of the lensed emission is described by the quadrupole moment of each image,

$$W^{\rm I}{}_{ij} = \frac{1}{S_{\rm I}} \int d^2r \ I_{\rm I}(\vec{r})(r_i - r_{oi})(r_j - r_{oj}). \tag{5.33}$$

Similarly the width of the source as it would be in the absence of lensing is

$$W^{\rm S}{}_{ij} = \frac{1}{S_{\rm S}} \int d^2s \ I_{\rm S}(\vec{s})(s_i - s_{oi})(s_j - s_{oj}). \tag{5.34}$$

5.4.2 Approximation appropriate to slightly extended sources

Attention will now be focused on the effect of source extent on "point" modeling and how one corrects for it. The analysis will be performed for "slightly extended" sources.

By "slightly extended" sources we mean sources for which $(\vec{r}-\vec{r}_o)$ is in some sense small for $\vec{r}$ in the region where an image has non-zero flux, where $\vec{r}_o$ is that image's centroid. Actually, what is relevant is that $(\vec{r}-\vec{r}_o)$ times the magnification gradient be small. One can think of this as requiring that the change in magnification across the image size be small. More quantitatively, it is assumed that $(\vec{r}-\vec{r}_o)\cdot\vec{\nabla}\,\boldsymbol{M}^{-1}\big|_{\vec{r}_o}$ is small for $\vec{r}$ in the region of non-zero surface brightness. For convenience we denote this small quantity as being $\mathcal{O}(\epsilon)$, thus

$$(\vec{r}-\vec{r}_o)\cdot\vec{\nabla}\,\boldsymbol{M}^{-1}\big|_{\vec{r}_o}=\mathcal{O}(\epsilon). \tag{5.35}$$

It follows also that $(\vec{r}-\vec{r}_o)\cdot\vec{\nabla}\det\boldsymbol{M}^{-1}\big|_{\vec{r}_o}=\mathcal{O}(\epsilon)$. We note that $(r_i-r_{oi})(r_j-r_{oj})\partial_i\,\boldsymbol{M}^{-1}\big|_{\vec{r}_o}\,\partial_j\,\boldsymbol{M}^{-1}\big|_{\vec{r}_o}=\mathcal{O}(\epsilon^2)$.

We assume that if the first derivative of the magnification matrix is small then the second derivation is smaller. In other words, we assume that

$$(r_i-r_{oi})(r_j-r_{oj})\partial_i\partial_j\,\boldsymbol{M}^{-1}\big|_{\vec{r}_o}=\mathcal{O}(\epsilon^2), \tag{5.36}$$

from which it follows that $(r_i-r_{oi})(r_j-r_{oj})\partial_i\partial_j\det\boldsymbol{M}^{-1}\big|_{\vec{r}_o}=\mathcal{O}(\epsilon^2)$.

In the approximations in the following sections, only the more significant terms will be kept. The terms that are higher order in these small quantities will be dropped. This approximation we will call the "slightly-extended" approximation.

Note that the assumptions that go into the "slightly-extended" approximation are quite different from the assumptions that go into the "source-plane" approximation, (sections 5.2.1 and 5.3.1). For the "slightly-extended" approximation it is assumed that the magnification changes little over the extent of each image; this could be a poor approximation for extended sources, even if the deflector model were perfect.

On the other hand, for sufficiently point-like sources, it is a fine approximation, even when the deflector model itself is poor. This is in contrast to the "source-plane" approximation which assumes that the change in magnification is small between the observed image location and the best-fit model's image location; the validity of the "source-plane" approximation depends on the model being a sufficiently good fit.

5.4.3 Relation between source flux and image flux for slightly extended sources

Starting from the integral for the source flux, $S_S = \int d^2s\ I_S(\vec{s})$ one can change variables from $\vec{s}$ to $\vec{r}$, getting $S_S = \int d^2r\ I_I(\vec{r}) \det \boldsymbol{M}^{-1}(\vec{r})$. One then Taylor expands the determinant around the image centroid $\vec{r}_o$. Dropping the terms that are higher order in the derivatives of the magnification, one finds

$$S_S = S_I \left\{ \underbrace{\det \boldsymbol{M}^{-1}(\vec{r}_o)}_{\mathcal{O}(1)} + \underbrace{\frac{1}{2} W^I{}_{ij} \partial_i \partial_j \det \boldsymbol{M}^{-1}(\vec{r}_o)\Big|_{\vec{r}_o}}_{\mathcal{O}(\epsilon^2)} + \mathcal{O}(\epsilon^3) \right\}. \tag{5.37}$$

The $\mathcal{O}(\epsilon)$ term vanishes because the expansion is done around the centroid of the image flux.

Unless the second derivative of the magnification matrix is significant over the region of the image flux, the "point" modeling flux transformation should be adequate, $S_S \approx S_I \det \boldsymbol{M}^{-1}(\vec{r}_o)$. This is true even when we are including the first order corrections, the $\mathcal{O}(\epsilon)$ corrections, to account for slightly extended sources.

5.4.4 Relating the image centroid to the image-plane location of the source centroid

Starting from the integral for the source centroid, $\vec{s}_o = \frac{1}{S_S} \int d^2s\ \vec{s} I_S(\vec{s})$ one can again change variables from $\vec{s}$ to $\vec{r}$, getting $\vec{s}_o = \frac{1}{S_S} \int d^2r\ I_I(\vec{r}) \left\{ \det \boldsymbol{M}^{-1}(\vec{r}) \vec{s}(\vec{r}) \right\}$. One then Taylor expands the term in curly braces around the image centroid $\vec{r}_o$, and uses for source flux S_S the expression above in equation 5.37. Dropping the terms of higher

order in the expansion parameter, one finds:

$$s_{oi} - s_i(\vec{r}_o) = \underbrace{W^{\mathrm{I}}{}_{jk}\left(\frac{1}{2}\partial_k\, M^{-1}{}_{ij}\Big|_{\vec{r}_o} + M^{-1}{}_{ij}\Big|_{\vec{r}_o}\,\partial_k \ln\det \boldsymbol{M}^{-1}\Big|_{\vec{r}_o}\right)}_{\mathcal{O}(\epsilon)\times\mathcal{O}(\text{source size})} + \mathcal{O}(\epsilon^2)\times\mathcal{O}(\text{source size}). \tag{5.38}$$

This is the discrepancy in the source plane between the source centroid $\vec{s}_o$ and the image centroid mapped back to the source plane, $\vec{s}(\vec{r}_o)$. Note that this is $\mathcal{O}(\epsilon)$. The discrepancy goes to zero for point-like sources, as expected. The discrepancy also goes to zero if there is no magnification gradient over the source size $\sqrt{W^{\mathrm{I}}{}_{ij}}$. However there is a discrepancy if the source is somewhat extended and there is a change in magnification over the extent of the source. The discrepancy is a first order effect, and any extension to "point" modeling to take source size into account must account for this.

The expression for the discrepancy in the image plane can also be found. Taylor expansion gives $\vec{r}(\vec{s}_o) - \vec{r}_o = \boldsymbol{M}(\vec{r}_o)(\vec{s}_o - \vec{s}(\vec{r}_o)) + \mathcal{O}((\vec{s}_o - \vec{s}(\vec{r}_o))^2)$. Since $(\vec{s}_o - \vec{s}(\vec{r}_o))$ is itself $\mathcal{O}(\epsilon)$ (see equation 5.38), the $\mathcal{O}((\vec{s}_o - \vec{s}(\vec{r}_o))^2)$ term is $\mathcal{O}(\epsilon^2)$ and can be dropped if we are keeping terms only to $\mathcal{O}(\epsilon)$. Therefore

$$r_p(\vec{s}_o) - r_{po} = \underbrace{W^{\mathrm{I}}{}_{jk}\left(\frac{1}{2}M_{pi}\partial_k M^{-1}{}_{ij} + \delta_{pj}\partial_k \ln\det \boldsymbol{M}^{-1}\right)\Big|_{\vec{r}_o}}_{\mathcal{O}(\epsilon)\times\mathcal{O}(\text{source size})} + \mathcal{O}(\epsilon^2)\times\mathcal{O}(\text{source size}). \tag{5.39}$$

This is the displacement in the image plane between the image of the source centroid $\vec{r}(\vec{s}_o)$ and the image centroid $\vec{r}_o$. This displacement is roughly the image width multiplied by the fractional change in the magnification over the extent of the image. If, for the sake of illustration, it is assumed that the image extent $W^{\mathrm{I}}{}_{ij}$ and the magnification matrix $M^{-1}{}_{ij}$ are both diagonal in some basis, and that the only non-zero gradient term is $\partial_x M^{-1}{}_{xx}$ then the displacement is $\vec{r}_o - \vec{r}(\vec{s}_o) = -\frac{3}{2} w_x \hat{x}\,(w_x \partial_x M^{-1}{}_{xx}/M^{-1}{}_{xx}) \approx$

$\frac{3}{2} \times$ x-width $\times \frac{\Delta \text{magnif}}{\text{magnif}}$ where w_x is the rms width of the image in the x-direction. As one expects, the image centroid is displaced in the x-direction if the magnification increases in the x-direction.

5.4.5 The "slightly-extended" corrections to "point" modeling

When the lens images are somewhat extended and there are magnification gradients at the image locations, the "point" modeling introduces slight errors as shown above. It is desirable to correct these effects to $\mathcal{O}(\epsilon)$, the first order in the gradient effects. This corrected chi-squared is called the "slightly-extended" corrections to the "point" modeling approximation.

The relation between source flux and image flux is only affected at second order, $\mathcal{O}(\epsilon^2)$, in the magnification gradients. Therefore the "point" modeling relation can be used unchanged, $S_S = S_I \det M^{-1}(\vec{r}_o)$. Thus the "point" modeling χ^2_{flux} (equation 5.20) and $\chi^2_{\substack{\text{flux} \\ \text{srcplane}}}$ (equation 5.24) carry over unchanged to the "slightly-extended" approximation:

$$\chi^2_{\substack{\text{flux} \\ \text{extended-source}}} = \chi^2_{\text{flux}}, \tag{5.40}$$

and

$$\chi^2_{\substack{\text{flux} \\ \text{srcplane,extended-source}}} = \chi^2_{\substack{\text{flux} \\ \text{srcplane}}}. \tag{5.41}$$

The relation between source position and image position needs a correction at first order, $\mathcal{O}(\epsilon)$, in the magnification gradients. In "point" modeling we have been treating the image-centroid and the source-centroid as being the same — under the lens mapping. As seen in equations 5.38 and 5.39 these actually differ by $\mathcal{O}(\epsilon)$ for slightly extended sources in the presence of magnification gradients. However these same expressions, 5.38 and 5.39, allow one to make corrections to the "point" modeling, to eliminate the $\mathcal{O}(\epsilon)$ discrepancy and to leave only an $\mathcal{O}(\epsilon^2)$ discrepancy. These corrections use the measured image extents, $W^I{}_{l\alpha,ij}$, and the inverse of the model magnification matrix and its derivatives evaluated at the measured image locations,

$\boldsymbol{M}^{-1}(\vec{r}_{l\alpha})$ and $\vec{\nabla}\,\boldsymbol{M}^{-1}\big|_{\vec{r}_{l\alpha}}$. One corrects from the measured location of the image centroid ($\vec{r}_o \equiv \vec{r}_{l\alpha}$, or $\vec{s}(\vec{r}_o) \equiv \vec{s}(\vec{r}_{l\alpha})$) to the location of the source centroid ($\vec{s}_o \equiv \vec{s}_l$, or $\vec{r}(\vec{s}_o) \equiv \vec{r}_\alpha(\vec{s}_k)$) since the source centroids are the model source parameters. This correction needs to be done for all the images so that the various images may be compared with each other.

In the actual chi-squared as calculated in the image plane ($\chi^2_{\text{positions}}$ equation 5.6) the expression $(\vec{r}_\alpha(\vec{s}_l) - \vec{r}_{l\alpha})$ enters into the chi-squared. This compares the image-plane position corresponding to the model source centroid, $\vec{r}_\alpha(\vec{s}_l)$, to the observed image centroid $\vec{r}_{l\alpha}$. This observed position should be corrected by the difference between the image-of-the-source-centroid and the centroid-of-the-image. In short, $\vec{r}_{l\alpha}$, should be replaced by

$$\vec{r}_{l\alpha} \longrightarrow \vec{r}\,'_{l\alpha} = \vec{r}_{l\alpha} + \hat{e}_p \left\{ W^{\mathrm{I}}{}_{l\alpha,jk} \left(\frac{1}{2} M_{pi} \partial_k M^{-1}{}_{ij} + \delta_{pj} \partial_k \ln \det \boldsymbol{M}^{-1} \right) \Big|_{\vec{r}_{l\alpha}} \right\}. \qquad (5.42)$$

where the correction term is from equation 5.39. Note that l denotes which source $1, \ldots, N_l$, that α denotes which image of the source $1, \ldots, N_\alpha$, that i, j, k, and p are indices running over the x- and y-components, and that $\hat{e}_p$ is a basis vector $\hat{x}$ or $\hat{y}$. The position chi-squared is then

$$\chi^2_{\substack{\text{positions} \\ \text{extended-source}}} = \sum_{\substack{\text{sources} \\ l}} \sum_{\substack{\text{images} \\ \alpha}} (\vec{r}_\alpha(\vec{s}_l) - \vec{r}\,'_{l\alpha}) \cdot \mathbf{S}^{-1}_{l\alpha} \cdot (\vec{r}_\alpha(\vec{s}_l) - \vec{r}\,'_{l\alpha}). \qquad (5.43)$$

If the source-plane approximation is being used then it is the expression $\vec{s}_l - \vec{s}(\vec{r}_{l\alpha})$ that appears in the expression for the chi-squared, $\chi^2_{\substack{\text{positions} \\ \text{srcplane}}}$, equation 5.13. This compares the model source centroid, $\vec{s}_l$, to the observed image centroid as mapped to the source plane, $\vec{s}(\vec{r}_{l\alpha})$. This mapping of the observed position to the source plane should also be corrected for the difference between the source-centroid and the

source-plane-position-of-image-centroid,

$$\begin{aligned}\vec{s}(\vec{r}_{l\alpha}) \longrightarrow \vec{s}\,'(\vec{r}_{l\alpha}) &= \vec{s}(\vec{r}_{l\alpha}) + \hat{e}_i \left\{ W^{\mathrm{I}}{}_{l\alpha,jk} \left(\frac{1}{2} \partial_k M^{-1}{}_{ij} + M^{-1}{}_{ij} \partial_k \ln \det \boldsymbol{M}^{-1} \right) \Big|_{\vec{r}_{l\alpha}} \right\} \\ &= \vec{r}_{l\alpha} - \vec{\nabla} \Phi(\vec{r}_{l\alpha}) \\ &\quad + \hat{e}_i \left\{ W^{\mathrm{I}}{}_{l\alpha,jk} \left(\frac{1}{2} \partial_k M^{-1}{}_{ij} + M^{-1}{}_{ij} \partial_k \ln \det \boldsymbol{M}^{-1} \right) \Big|_{\vec{r}_{l\alpha}} \right\} .\end{aligned} \tag{5.44}$$

This can be thought of as a correction to the deflection angle. The source-plane approximation to the chi-squared is then

$$\chi^2_{\substack{\text{positions} \\ \text{srcplane,extended-source}}} = \sum_{\substack{\text{sources} \\ l}} \sum_{\substack{\text{images} \\ \alpha}} \left(\vec{s}_l - \vec{s}\,'(\vec{r}_{l\alpha}) \right) \boldsymbol{M}(\vec{r}_{l\alpha}) \cdot \mathbf{S}^{-1}_{l\alpha} \cdot \boldsymbol{M}(\vec{r}_{l\alpha}) \left(\vec{s}_l - \vec{s}\,'(\vec{r}_{l\alpha}) \right) . \tag{5.45}$$

Note that since $\vec{s}\,'(\vec{r}_{l\alpha})$ does not depend on $\vec{s}_l$, the analytic minimization of the chi-squared over model source parameters $\vec{s}_l$ is not affected. Equation 5.15, which gives the model positions $\vec{s}_l$ that minimize the chi-squared, is changed only by the replacement of $\vec{s}(\vec{r}_{l\alpha})$ by $\vec{s}\,'(\vec{r}_{l\alpha})$. Equation 5.27, that gives the model fluxes that minimize the chi-squared, remains unchanged.

Note that this "slightly-extended" correction to the "point" modeling technique requires that the third derivatives of the model potential, $\partial_i \partial_j \partial_k \Phi(\vec{r})$, be calculated.

5.5 Beyond "point" modeling

The "point" modeling discussed above utilized only a single position and integral flux for each distinct component of each lens image. The rationale for this method was to capture just the most important information about the lens system thus allowing there to be a computationally quick way of ruling out bad models for the deflector. Even the corrections to "point" modeling for slightly extended sources used basically the same small subset of information about the lens system. In some cases it will be desirable to use more of the information contained in the surface brightness distribution.

5.5.1 Using source major axis, minor axis, and orientation

The logical next step beyond treating each image as a point is to acknowledge its finite extent. That is, one measures the second moment of the sky brightness distribution for each image and uses these second moments as constraints.

$$W^{\mathrm{I}}{}_{l\alpha,ij} = \frac{1}{S_{\mathrm{I}}} \int d^2r \; I_{\mathrm{I}}(\vec{r})(r_i - r_{l\alpha,i})(r_j - r_{l\alpha,j}). \tag{5.46}$$

The source and the image widths are related by

$$W^{\mathrm{S}}{}_l = \boldsymbol{M}^{-1}(\vec{r}_{l\alpha}) W^{\mathrm{I}}{}_{l\alpha} \boldsymbol{M}^{-1}(\vec{r}_{l\alpha}) + \mathcal{O}(\epsilon) \times \mathcal{O}(\text{source size}^2). \tag{5.47}$$

(This is a matrix equation. There is a 2×2 symmetric matrix $W^{\mathrm{I}}{}_{l\alpha}$ for each image α of each source l.) Therefore the image widths can be used as modeling constraints provided that the source widths $W^{\mathrm{S}}{}_l$ are added as additional model parameters.

Astronomical source sizes are often characterized by specifying major axis length, minor axis, and orientation angle. The source and image widths can alternately be described by specifying the major axis, minor axis, and orientation angle of each image if that is preferred over the second moments.

The source widths give three constraints, rms x-width $\sqrt{W^{\mathrm{I}}{}_{l\alpha,xx}}$, rms y-width $\sqrt{W^{\mathrm{I}}{}_{l\alpha,yy}}$, and xy-covariance $W^{\mathrm{I}}{}_{l\alpha,xy}$, for each image α of each source l. This provides $3 \times N_l \times N_\alpha$ additional constraints on the model. For each image l one must also include in the model three additional model parameters: rms x-width $\sqrt{W^{\mathrm{S}}{}_{l,xx}}$, rms y-width $\sqrt{W^{\mathrm{S}}{}_{l,yy}}$, and xy-covariance $W^{\mathrm{S}}{}_{l,xy}$, for a total of $3 \times N_l$ additional model parameters. Thus including the widths of the images increases that number of constraints of the deflector by a net of $3 \times N_l \times (N_\alpha - 1)$.

The only difficulty in formulating the chi-squared expression to use the source extent as a constraint is obtaining an estimate of the measurement error on each of the axes and the orientation angle, and estimating the covariance in these measurement errors. In general the measurement errors in the major axis, minor axis, orientation angle, and flux will be correlated with each other. The errors in x-position and y-

position, while correlated with each other, will be uncorrelated with the errors in the major axis, the minor axis, the orientation, and the flux. See appendix A.

The effect of magnification gradients on the mapping from image to source widths (equation 5.47) shows up at $\mathcal{O}(\epsilon)$. The effect depends on the third moment of the image flux. For the α^{th} image of the l^{th} source, the effect is

$$\begin{aligned} W^{\mathrm{S}}{}_{l,ij} &= M^{-1}{}_{im}(\vec{r}_{l\alpha}) M^{-1}{}_{jn}(\vec{r}_{l\alpha}) W^{\mathrm{I}}{}_{l\alpha,mn} \\ &\quad + T^{\mathrm{I}}{}_{l\alpha,mnk} \Big\{ M^{-1}{}_{im} M^{-1}{}_{jn} \partial_k \ln \det \boldsymbol{M}^{-1} \\ &\qquad + \frac{1}{2} M^{-1}{}_{im} \partial_k M^{-1}{}_{jn} + \frac{1}{2} M^{-1}{}_{jn} \partial_k M^{-1}{}_{im} \Big\} |_{\vec{r}_{l\alpha}} \\ &\quad + \mathcal{O}(\epsilon^2) \times \mathcal{O}(\text{source size}^2), \end{aligned} \tag{5.48}$$

where the third moment of an image's flux around its centroid $\vec{r}_o$ is

$$T^{\mathrm{I}}{}_{ijk} = \frac{1}{S_{\mathrm{I}}} \int d^2r \; I_{\mathrm{I}}(\vec{r})(\vec{r}_i - \vec{r}_{oi})(\vec{r}_j - \vec{r}_{oj})(\vec{r}_k - \vec{r}_{ok}). \tag{5.49}$$

As the third moments are likely to be poorly constrained, and as the errors on the second moments will be larger than the error on the image centroids, it seems sensible to ignore the effects on magnification gradients on the source widths when using the source widths as constraints. Besides, for symmetrical sources, the third moments are zero.

5.5.2 LensClean

In some cases one wants to use all the surface brightness information available in constraining models of the deflector's potential, despite the computational cost. This is particularly desirable for lens systems in which there is extended flux that is not adequately described by the parameters that describe a Gaussian: position, flux, width, and orientation.

The LensClean algorithm was developed to do this (Kochanek & Narayan 1992). It deconvolves the beam pattern from the map of the lens system, mapping the clean

components through the current lens model and removing all the lensed images of the clean components simultaneously. The deflector model parameters are adjusted to minimize the residual errors caused by inconsistency between the map and the model. A more sophisticated version, Visibility LensClean (Ellithorpe *et al.* 1996), subtracts the lensed clean components directly from the interferometer visibility data rather than from the source map. These LensClean methods require that a CLEAN be done for each set of deflector model parameter values which are tried in the course of finding the best-fit parameter values. For maps in which even the ordinary CLEAN without lensing (Hogbom 1974; Clark 1980; Schwab 1984) takes hours on a workstation[2], with models having an eight-to-ten-or-more dimensional parameter space to explore[3], the computational cost of LensClean is prohibitive.

5.6 Confidence limits on model parameters

It is desirable to know not only what the best-fit model parameters are, but also whether the model adequately explains the data, and if so how well the model parameters are constrained. An especially lucid and accessible introduction to this subject is given in Press *et al.* (1992), where it is made clear under what conditions one may use what techniques, although some of the results are stated without proof.

The "point" modeling technique described above requires measured positions and error estimates on the measured positions for the identifiable components in the emission. The position errors estimated from fitting elliptical Gaussians to the maps of MG J0414+0534, (table 3.5), contain only the statistical error due to the map noise (see appendix A). They do not take into account the systematic error introduced by the deconvolution (CLEAN) algorithm. In fact, the errors introduced by the deconvolution are not well understood. Due to this unknown error introduced by the deconvolution, (1) it is expected that the statistical errors are underestimates of the total error, and (2) the errors are not known to be Gaussian.

[2]such as the observation of MG J0414+0534 described in this thesis

[3]For this MG J0414+0534 observation, the simpler models can be ruled out using "point" modeling techniques.

That the size of the errors is not well known prevents an absolute measure of goodness of fit of the deflector models. The standard method of estimating goodness of fit, that the reduced χ^2 be near 1, requires that the measurement errors be known, that they be Gaussian, and that the model be adequately approximated by a linearizable model in the region of interest. In that case the minimum of the chi-squared is distributed as a chi-squared with $n-m$ degrees of freedom, where m is the number of fitted model parameters and n is the number of measured data points or constraints. If the data yields a minimized chi-squared that is is unlikely to have been drawn from the model, then the model may be rejected.

We would however like to say something about goodness of fit, rejecting at least those models which are grossly wrong. To this end, it is assumed that the deconvolution algorithm will not displace components by a whole beam width. More precisely, a Gaussian is fitted to each component in the clean map, and this Gaussian is used as the error distribution. The clean map consists of the clean component model of the source, convolved with a Gaussian which is the size of the central lobe of the beam, and added to the residual map noise. Therefore this error estimate is just the beam size for point sources, and is the beam size convolved with the actual source width for extended sources. As crude as this upper limit on the position errors is, the simplest lens models (less than 9 parameters) are all rejected by this criterion (see table 7.2).

Using the convolved image sizes as estimates of the position errors gives an overestimate of the errors. Using only the statistical errors on the image centroids underestimates the errors. A better estimate of the errors, between these upper and lower limits, is to add the deconvolved image size in quadrature with the statistical errors on the image centroids, since this make some allowance for image shifts due to deconvolution errors.

For those models which are not rejected by these crude goodness-of-fit criteria, we would like to know how well the model parameters are constrained.

The concept of confidence limits on model parameters is as follows: If a measurement were repeated many times, and fitted parameters $\vec{a}_i$ were found for each measurement, then any region of any shape within with $p\%$ of the measurements

would fall is a $p\%$ confidence region. If the best-fit model is a good model, then simulated data drawn from the best-fit model, with an error distribution the same as the actual error distribution, result in simulated model parameters $\vec{a}_i^S$ which are assumed to be distributed approximately the same as the hypothetical repeated measurements $\vec{a}_i$. Once one knows the shape of the distribution of the $\vec{a}_i^S$'s, one can find confidence regions for any subset of the parameters containing the specified $p\%$ of the distribution.

Were the error distribution known well enough to simulate, whether it were Gaussian or not, then a Monte-Carlo technique could be used. Using the best-fit model, simulated data sets could be generated, and errors generated from the known error distribution could be added to the simulated data. These simulated data sets then could be processed in the same way as the actual data set, yielding "best-fit" parameter values for each simulated data set. This distribution of these simulated results would give one the confidence interval on one's fitting results.

However, our error distribution is not well enough known to use this technique. Instead we make a few assumptions:

1. It is assumed that the errors are Gaussian and that their size is given by the statistical error on the centroid positions added in quadrature with the deconvolved image sizes.

2. It is assumed that the parameterized model is linearizable in the region of interest. This is sensible since any well-behaved function is linear locally.

3. It is assumed that the parameterized model is indeed the one from whose distribution the data were drawn. A typical data set drawn from this distribution should have $\chi^2_{\substack{\text{min}\\\text{expected}}} = n - m$, where n is the number of data points and m is the number of fitted model parameters. Actually, $\chi^2_{\substack{\text{min}\\\text{expected}}}$ is drawn from a chi-squared distribution with $n - m$ degrees of freedom, which has mean $n - m$ and variance $2(n - m)$. It is seen in practise (table 7.2) that the better-fitting models for the VLBA observation of MG J0414+0534 all have $\chi^2_{\text{min}} < n - m$, indicating that the values of χ^2_{min} are artificially low or equivalently that the er-

ror estimate is an overestimate of the position errors. Therefore the confidence regions resulting from them will be over large.

Then to calculate confidence limits on the model parameters, we use the following technique which is suitable for Gaussian errors and linearizable models. The confidence region for ν interesting parameters out of the m fitted parameters can be found as follows: Let $\chi^2_\nu(\vec{a})$ be the χ^2 with the ν interesting parameters held fixed and minimized over the $m-\nu$ remaining parameters. Then the ν-dimensional confidence region is the region where $\chi^2_\nu(\vec{a}) < \chi^2_{\rm min} + \Delta\chi^2_{\substack{\nu \\ \rm theoretical}}$. The level $\Delta\chi^2_{\substack{\nu \\ \rm theoretical}}$ is a function of the number of interesting parameters ν and the desired probability level $p\%$. For example, for one interesting parameter, $\Delta\chi^2_{\substack{\nu=1 \\ \rm theoretical}} = 1$, 2.71, and 4 for confidence regions of 68.3%, 90%, and 95.4%, respectively. For two interesting parameters, $\Delta\chi^2_{\substack{\nu=2 \\ \rm theoretical}} = 2.30$, 4.61, and 6.17 for these same confidence levels. Therefore, to find each parameter's confidence limits, that parameter is stepped away from its value at the minimum, while the chi-squared is minimized over all the other parameters, until the minimized chi-squared reaches the level $\chi^2_{\rm min} + \Delta\chi^2_{\substack{\nu=1 \\ \rm theoretical}}$.

Chapter 6

Choosing a Model for the Deflector Mass Distribution or Potential

In modeling a gravitational lens system one needs a model of the deflector: its mass distribution, its Newtonian gravitational potential, or its 2-D lensing potential, as described in chapter 4. Indeed in the thin lens approximation, only the 2-D lensing potential, proportional to the integral along the line of sight of the Newtonian gravitational potential, is actually constrained.

A parameterized form for the deflector is chosen, $\Phi_{\{a\}}(\vec{r})$ depending on some number N_a of model parameters $\{a\}$. The model parameters are tuned to find the best fit model of its type. Even if one chooses to parameterize the mass distribution instead of parameterizing the potential directly, one can still think of the model as a model of the 2-D lens potential, since the 2-D lens potential is calculable from the mass distribution, and only the 2-D lens potential enters into the expression for the deflection of light.

The form of the potential may be motivated by a desire to have a model that is a reasonable shape for a galaxy mass distribution. Alternately it may be motivated by an attempt to be sensitive to those features of the deflector mass distribution which lensing is expected to constrain.

6.1 How many parameters can a model have?

The complexity of the model is limited by the number of observational constraints on the modeling.

For point-modeling a lens-system that has N_α images of N_l background sources, without using flux information, there are $2 \times N_\alpha \times N_l$ constraints – the right ascension and declination of each image. From this must be subtracted the $2 \times N_l$ model source positions: right ascension and declination for each model source. This leaves $2 \times N_l \times (N_\alpha - 1)$ constraints available to constrain deflector model parameters.

For a typical quad lens, with four images of a single background source, this leaves only six constraints to constrain the deflector potential. As the very simplest deflector models have five parameters, such a system is not much over-constrained.

Using image fluxes as constraints increases the net number of constraints available for modeling the deflector. There are now $3 \times N_\alpha \times N_l$ model constraints: x, y, and flux for each image. However x, y, and flux for each model source, $3N_l$ parameters, must be included in the source model. This leaves $3 \times N_l \times (N_\alpha - 1)$ constraints available to constrain deflector model parameters. This is a significant improvement in the number of available constraints. Is there any reason why one would not use the flux information? There are three principal reasons why one would ignore the flux information. Firstly, as in the case of MG J0414+0534, there may be significant discrepancies in the flux ratios at two different wavelengths — causing doubt as to which fluxes to trust. In the case of MG J0414+0534, these may be due to microlensing, which is expected to be a more significant effect in the optical than in the radio as the optical flux is assumed to come from a smaller region than the radio flux. Another explanation for the flux ratio discrepancy is that there may be obscuration by dust in one part of the deflector galaxy. One may want to assume that such microlensing and obscuration effects are unimportant at radio wavelengths. Alternately one may simply dispense with fluxes, and use only the positions which are not distorted by these effects. A second reason to ignore the flux information in modeling is that the source may be variable in flux. Variability combined with the time delay could result

No. of constraints on deflector model when using specified modeling constraints						
Lens Type			Constraint type			
	N_l	N_α	position	position & flux	position & galaxy x,y	position, flux & galaxy x,y
any	N_l	N_α	$2N_l(N_\alpha-1)$	$3N_l(N_\alpha-1)$	$2+2N_l(N_\alpha-1)$	$2+3N_l(N_\alpha-1)$
double	1	2	2	3	4	5
quad (PG B1115+080)	1	4	6	9	8	11
double (B0957+561)	6	2	12	18	14	20
quad (MG J0414+0534)	4	4	24	36	26	38

Table 6.1: The number of constraints on the deflector model when the specified type of constraints are used. These are given for lens systems with various numbers of source objects N_l and with various numbers of images N_α of each source. In parentheses are names of lens systems that fall within these classes. (B0957+561: Garrett *et al.* 1994)

in the observed flux in one image corresponding to a flux minimum while the flux in another image corresponds to a flux maximum. Thirdly, the positions may be so much better constrained than the fluxes that it is expected that the fluxes have little contribution to the chi-squared; this is the case for the for VLBA observation of MG J0414+0534 described in this thesis.

One may also use direct observation of the deflector in modeling the deflector: x-position, y-position, velocity dispersion, ellipticity of isophotes, orientation, etc. However, instead of using the luminous matter to construct a model of the deflector's mass, one would prefer to constrain the deflector's mass based on direct gravitational effects — and then compare with the observed properties of the luminous matter. Yet in a case where there is too little information in the lens images, and where a deflector model is needed to calculate a time delay to compare with an observed time delay, such information will be used. However for MG J0414+0534, where the images positions have errors of well under a milliarcsecond (chapter 3 of this thesis) and the galaxy position has 50 mas errors (Falco *et al.* 1997) we chose not the use the galaxy position as a constraint.

The very simplest deflector models have at minimum five parameters: x- and y-

position of the deflector, the strength of the lens, the amount of ellipticity or shear, and the orientation of the ellipticity or shear. (The strength of the lens is variously specified by Einstein ring radius, deflector mass, or velocity dispersion.) One can see (table 6.1) that for a simple double lens one is forced to use constraints from direct observation of a luminous deflector even to attempt to model the system. For a quad lens, if one neither trusts the flux measurements nor has observations of the deflector then there are barely enough constraints (six) for a simple model (five parameters). So one must use flux or deflector positions if one wants to consider more complicated models.

For lenses that have multiple background objects, such as knots in a jet, the number of available constraints is much larger, and one can hope to learn more about the deflector's mass distribution.

6.2 Taylor-expanded multipole model: An expansion in terms expected to be of decreasing significance

6.2.1 Motivation

One way to model the deflector potential is to use a physically motivated potential shape that has no more free parameters than observational constraints. A different approach, used here, is to tailor the model choice to be sensitive to what can be learned from the gravitational lens, rather than to what one would wish *a priori* to learn.

The deflections and magnifications in a gravitational lens system provide constraints on the gradient and higher derivatives of the lens potential only at the locations of the lensed images. In other words, there are constraints on $\vec{\nabla}\Phi(\vec{r})$ and $\partial_i\partial_j\Phi(\vec{r})$ only at those locations $\vec{r}$ where there is flux. One may wish to know the behavior of the potential at the center of the mass distribution, but there are no

constraints there — except perhaps for some limits from the absence of additional images above a certain flux level. One may wish to know the behavior of the potential at the very far edges of the deflector galaxy. However, there is only information on the behavior of the potential at the locations of the lensed images.

For quad lenses, or for Einstein rings, all the lensed flux lies roughly equidistant from the center of the deflector. For MG J0414+0534 all the VLBA components lie within ± 0.23 arcsec of a ring of radius 1.16 arcsec around the measured galaxy position of Falco *et al.* (1997). Letting r equal the radial distance to each image and letting b be this ring radius, then $\left|\frac{r-b}{b}\right| < 0.20$ for all the images of MG J0414+0534. For such a system, Taylor-expanding the radial dependence of the potential in powers of $\left|\frac{r-b}{b}\right|$ is a natural choice. Successive terms become successively less important: $\left|\frac{r-b}{b}\right|^3 < 1\%$ for all the images of MG J0414+0534. This expansion was suggested by Kochanek (1991).

Since the lens potential is a function of both angular location and radial location, (r, θ), some way of choosing the important contributions to the angular dependencies must be selected. The 2-D Poisson equation is separable, leading to the multipole expansion, so a multipole expansion in angle is used. This separability allows one to relate the multipole components of the potential term by term to the corresponding multipole components of the mass distribution, which is convenient for gaining physical insight into the deflector system. The assumption is made that the mass distribution varies smoothly with angle so the higher order angular multipole terms may be neglected.

6.2.2 Using a multipole expansion in angle to solve the 2-D Poisson equation

The 2-D lens potential, $\Phi(\vec{r})$, is related to the surface mass density of the deflector, $\sigma(\vec{r})$, by a two dimensional Poisson equation

$$\nabla^2_{\vec{r}} \Phi(\vec{r}) = \frac{2\sigma(\vec{r})}{\sigma_{\rm crit}}, \tag{6.1}$$

where $\sigma_{\rm crit}$ is constant for any particular lens, and where $\vec{r}$ is a two-dimensional angular vector, (see chapter 4).

In the case of the three dimensional Poisson equation it is useful, and standard, to expand the potential field (the left-hand side of the equation) and the source term (the right-hand side of the equation) in terms of a complete set of orthonormal basis functions which make the Poisson equation separable. This is the multipole expansion. See, for example, Jackson (1975) chapter 3 for electrostatics or Binney & Tremaine (1987) section 2.4 for the Newtonian gravitational field. The same method is readily, and even more simply, used in the two dimensional case.

To find the basis functions in which the Poisson equation is separable, one attempts to find solutions to the homogenous Poisson equation $\nabla^2_{\vec{r}} f(\vec{r}) = 0$ which are separable, $f(r,\theta) = R(r)\Theta(\theta)$. In (r,θ) coordinates the Poisson equation is

$$0 = \left\{ \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \right\} R(r)\Theta(\theta). \tag{6.2}$$

Separating the variables,

$$\frac{r^2}{R(r)} \left\{ \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} \right\} R(r) = \mu = -\frac{1}{\Theta(\theta)}\frac{\partial^2}{\partial \theta^2}\Theta(\theta), \tag{6.3}$$

where the constant μ has been introduced for convenience, and where different values of μ will correspond to different solutions of the Poisson equation. For $\mu < 0$, letting $\mu = -m'^2$, it is seen that $\frac{\partial^2}{\partial \theta^2}\Theta(\theta) = m'^2\Theta(\theta)$ requiring that $\Theta(\theta)$ be a linear combination of $\cosh(m'\theta)$ and $\sinh(m'\theta)$. Only solutions on a circle are of interest, for which $\Theta(\theta) = \Theta(\theta + 2\pi)$, so these hyperbolic solutions may be ignored. For $\mu \geq 0$ we may let $\mu = m^2$, so that $\frac{\partial^2}{\partial \theta^2}\Theta(\theta) = -m^2\Theta(\theta)$. Thus $\Theta(\theta)$ is a linear combination of $\cos m\theta$ and $\sin m\theta$. We note in passing that the left-hand side of equation 6.3 can be solved to find the solutions $R(r)$ for each angular function solution.

These angular solutions form a complete and orthonormal set of basis functions over the interval from 0 to 2π. The orthonormality relations are, for m and n positive

integers:

$$\begin{aligned}
\delta_{mn} &= \int_0^{2\pi} d\theta \frac{\sin m\theta}{\sqrt{\pi}} \frac{\sin n\theta}{\sqrt{\pi}} \quad \text{for } m \text{ and } n \text{ positive integers} \\
\delta_{mn} &= \int_0^{2\pi} d\theta \frac{\cos m\theta}{\sqrt{\pi}} \frac{\cos n\theta}{\sqrt{\pi}} \quad \text{for } m \text{ and } n \text{ positive integers} \\
1 &= \int_0^{2\pi} d\theta \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} \\
0 &= \int_0^{2\pi} d\theta \frac{\cos m\theta}{\sqrt{\pi}} \frac{\sin n\theta}{\sqrt{\pi}} \quad \text{for } m \text{ and } n \text{ positive integers} \\
0 &= \int_0^{2\pi} d\theta \frac{1}{\sqrt{2\pi}} \frac{\cos m\theta}{\sqrt{\pi}} \quad \text{for } m \text{ a positive integer} \\
0 &= \int_0^{2\pi} d\theta \frac{1}{\sqrt{2\pi}} \frac{\sin m\theta}{\sqrt{\pi}} \quad \text{for } m \text{ a positive integer.}
\end{aligned} \tag{6.4}$$

The completeness relation is,

$$\delta_{2\pi}(\theta - \theta') = \lim_{N\to\infty} \left\{ \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} + \sum_{m=1}^{N} \left\{ \frac{\cos m\theta}{\sqrt{\pi}} \frac{\cos m\theta'}{\sqrt{\pi}} + \frac{\sin m\theta}{\sqrt{\pi}} \frac{\sin m\theta'}{\sqrt{\pi}} \right\} \right\} \tag{6.5}$$

where $\delta_{2\pi}(\theta - \theta')$ is defined as having the following behavior when integrated against any sufficiently smooth function $f(\theta)$, namely, $f(\theta') = \int_0^{2\pi} d\theta\, f(\theta)\delta_{2\pi}(\theta - \theta')$.

Any function of r and θ may be re-written in terms of these angular basis functions, with coefficients which are functions of r only,

$$f(r,\theta) = f_o(r) + \sum_{m=1}^{\infty} \left\{ f_{mc}(r)\cos m\theta + f_{ms}(r)\sin m\theta \right\}, \tag{6.6}$$

and where the coefficient functions are related to $f(r,\theta)$ by,

$$\begin{aligned}
f_o(r) &= \frac{1}{2\pi} \int_0^{2\pi} d\theta\, f(r,\theta) \\
f_{mc}(r) &= \frac{1}{\pi} \int_0^{2\pi} d\theta\, f(r,\theta)\cos m\theta \\
f_{ms}(r) &= \frac{1}{\pi} \int_0^{2\pi} d\theta\, f(r,\theta)\sin m\theta.
\end{aligned} \tag{6.7}$$

As a consequence of the separability of the Poisson equation, the Poisson equation must hold term by term in the multipole expansion. As can be seen, the Laplacian

operator does not mix terms in the multipole expansion.

$$\begin{aligned}\nabla^2 f(r,\theta) &= \left\{\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right\} f_o(r) \\ &+ \sum_{m=1}^{\infty}\left(\cos(m\theta)\left\{\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{m^2}{r^2}\right\} f_{mc}(r)\right) \\ &+ \sum_{m=1}^{\infty}\left(\sin(m\theta)\left\{\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{m^2}{r^2}\right\} f_{ms}(r)\right). \end{aligned} \tag{6.8}$$

Thus each multipole term in the solution of the inhomogenous Poisson equation $\nabla^2 f(r,\theta) = g(r,\theta)$ depends only on the corresponding multipole term in the source term of the Poisson equation:

$$\begin{aligned}\left\{\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{m^2}{r^2}\right\} f_{mc}(r) &= g_{mc}(r) \\ \left\{\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{m^2}{r^2}\right\} f_{ms}(r) &= g_{ms}(r) \\ \left\{\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right\} f_o(r) &= g_o(r) \end{aligned} \tag{6.9}$$

The solutions to the homogenous Poisson equation $\nabla^2 f(r,\theta) = 0$ may readily be found in the multipole expansion from equations 6.9, setting g to zero:

$$\begin{aligned} f_{mc}(r) &= A_{mc} r^m + B_{mc} r^{-m} \\ f_{ms}(r) &= A_{ms} r^m + B_{ms} r^{-m} \\ f_o(r) &= C_o + C_1 \ln r \end{aligned} \tag{6.10}$$

The integration constants A_{mc}, B_{mc}, A_{ms}, B_{ms}, C_o,and C_1, are determined by the boundary conditions.

Since Poisson's equation is linear, we may solve $\nabla^2_{\vec{r}}\Phi(\vec{r}) = \frac{2\sigma(\vec{r})}{\sigma_{\rm crit}}$, by summing up contributions to Φ from matter on circular rings — actually matter on cylindrical shells that, projected along the line of site, form circular rings. For the ring at radius r_o,

$$\nabla^2_{r,\theta}\Phi(r_o; r, \theta) = \frac{2\sigma(r_o,\theta)}{\sigma_{\rm crit}}\delta(r - r_o) \tag{6.11}$$

Angular multipole expansions are done both of $\Phi(r_o; r, \theta)$ and of $\sigma(r_o, \theta)$. Each multipole component of Φ depends only on the corresponding multipole component of σ as shown above. Everywhere except at the ring $r = r_o$, the potential $\Phi(r_o; r, \theta)$ satisfies the homogenous, free-space, Poisson equation, with one set of coefficients A_{mc}, B_{mc}, A_{ms}, B_{ms}, C_o, and C_1, outside the ring, and another set of coefficients inside the ring. The boundary conditions are used to determine the coefficients for each multipole moment from the corresponding multipole moment of $\sigma(r_o, \theta)$. In particular one requires that far from the mass distribution the potential must behave as that of a point mass, $\lim_{r\to\infty} \Phi(r_o; r, \theta) \propto \ln r$. Since there is no matter at the origin, it is required that $\Phi(r_o; r, \theta)$ be well behaved at the origin, that is that it must be continuous and have a continuous first derivative. The divergence theorem is used to relate the discontinuity in the derivative at the circular shell to the mass distribution on the shell. It is found that for $r > r_o$ the multipole components of $\Phi(r_o; r, \theta)$ are

$$\begin{aligned}
\Phi^{>}_{mc}(r_o; r) &= -\frac{r_o}{m}\left(\frac{r}{r_o}\right)^{-m}\frac{\sigma_{mc}(r_o)}{\sigma_{\text{crit}}} \\
\Phi^{>}_{ms}(r_o; r) &= -\frac{r_o}{m}\left(\frac{r}{r_o}\right)^{-m}\frac{\sigma_{ms}(r_o)}{\sigma_{\text{crit}}} \\
\Phi^{>}_{o}(r_o; r) &= C_o(r_o) + 2\frac{\sigma_o(r_o)}{\sigma_{\text{crit}}} r_o \ln r
\end{aligned} \tag{6.12}$$

and for $r < r_o$

$$\begin{aligned}
\Phi^{<}_{mc}(r_o; r) &= -\frac{r_o}{m}\left(\frac{r}{r_o}\right)^{m}\frac{\sigma_{mc}(r_o)}{\sigma_{\text{crit}}} \\
\Phi^{<}_{ms}(r_o; r) &= -\frac{r_o}{m}\left(\frac{r}{r_o}\right)^{m}\frac{\sigma_{ms}(r_o)}{\sigma_{\text{crit}}} \\
\Phi^{<}_{o}(r_o; r) &= C_o(r_o) + 2\frac{\sigma_o(r_o)}{\sigma_{\text{crit}}} r_o \ln r_o
\end{aligned} \tag{6.13}$$

The contribution of all matter located on all the rings may be integrated to give the complete solution to the 2-D Poisson equation.

The solution to the 2-D Poisson equation, for the multipole components of $\Phi(r, \theta)$

in terms of the multipole components of $\sigma(r,\theta)$ is:

$$\begin{aligned}
\Phi_o(r) &= \text{constant} + 2\ln r \int_0^r r'dr' \frac{\sigma_o(r')}{\sigma_{\text{crit}}} + 2\int_r^\infty r'dr' \ln r' \frac{\sigma_o(r')}{\sigma_{\text{crit}}} \\
\Phi_{mc}(r) &= -\frac{r^{-m}}{m}\int_0^r r'^{(1+m)}dr' \frac{\sigma_{mc}(r')}{\sigma_{\text{crit}}} - \frac{r^m}{m}\int_r^\infty r'^{(1-m)}dr' \frac{\sigma_{mc}(r')}{\sigma_{\text{crit}}} \\
\Phi_{ms}(r) &= -\frac{r^{-m}}{m}\int_0^r r'^{(1+m)}dr' \frac{\sigma_{ms}(r')}{\sigma_{\text{crit}}} - \frac{r^m}{m}\int_r^\infty r'^{(1-m)}dr' \frac{\sigma_{ms}(r')}{\sigma_{\text{crit}}}.
\end{aligned} \tag{6.14}$$

This is the same as in Kochanek (1991) but with a different normalization convention.

6.2.3 Physical significance of the multipole components of the surface mass density

The multipole components of the 2-D lensing potential have been related to the multipole components of the projected surface mass distribution of the deflector. It was claimed that this allows physical insight into the shape of the mass distribution generating the potential. In this section the physical significances of the surface mass density's multipole components are shown.

For thinking about the physical meanings of the dipole and higher moments in the multipole expansion, it is convenient to write them in vector or vector-like form. For any function $f(r,\theta)$ for which a multipole expansion has been made, define

$$\vec{f}_m(r) \equiv \hat{x} f_{mc}(r) + \hat{y} f_{ms}(r). \tag{6.15}$$

Then

$$f(r,\theta) = f_o(r) + \hat{r}\cdot\vec{f}_1(r) + \sum_{m=2}^{\infty} (\hat{x}\cos m\theta + \hat{y}\sin m\theta)\cdot\vec{f}_m(r) \tag{6.16}$$

and

$$\vec{f}_m(r) = \frac{1}{\pi}\int_0^{2\pi} d\theta(\hat{x}\cos m\theta + \hat{y}\sin m\theta) f(r,\theta). \tag{6.17}$$

Note that the dipole $\vec{f}_1(r)$ is a genuine vector as defined by its behavior under rotation. The higher multipole moments, for $m \geq 2$, are not genuine vectors, not behaving as

vectors do under rotation. It only takes a rotation by $\frac{2\pi}{m}$ to map $\vec{f}_m(r)$ back onto itself, rather than the 2π rotation needed by a true vector. We note in passing that the mth multipole moment for $m \geq 2$ may be written as a traceless symmetric 2-dimensional rank m tensor.

Each of these multipole components instead of being described by its sine and cosine components may alternately be described by an angle $\psi_{f,m}$ and an amplitude $f_m(r)$, defined so that

$$\begin{aligned} f_{mc}(r) &= f_m(r)\cos m\psi_{f,m}(r) \\ f_{ms}(r) &= f_m(r)\sin m\psi_{f,m}(r), \end{aligned} \tag{6.18}$$

and consequently,

$$f(r,\theta) = f_o(r) + \sum_{m=1}^{\infty} f_m(r)\cos m(\theta - \psi_{f,m}(r)). \tag{6.19}$$

Notice that in general the angle of a multipole moment may vary with radius.

The multipole expansion of the deflector's surface mass density is therefore

$$\sigma(r,\theta) = \sigma_o(r) + \sum_{m=1}^{\infty} \vec{\sigma}_m(r)\cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta), \tag{6.20}$$

where the multipole components are

$$\begin{aligned} \sigma_o(r) &= \frac{1}{2\pi}\int_0^{2\pi} d\theta\sigma(r,\theta) \\ \vec{\sigma}_m(r) &= \frac{1}{\pi}\int_0^{2\pi} d\theta(\hat{x}\cos m\theta + \hat{y}\sin m\theta)\sigma(r,\theta). \end{aligned} \tag{6.21}$$

Note that each multipole component, $\sigma_o(r)$ and $\vec{\sigma}_m(r)$, at radius r depends only on the surface mass in an infinitesimally narrow annulus of radius r centered at the origin. The multipole components are zero at radius r if there is no mass in that annulus. If there is mass in the annulus, then the multipole components describe the distribution of the mass within that annulus.

The monopole moment $\sigma_o(r)$ is the average surface mass density in the infinites-

imally narrow circular annulus at a radius r. Another way of saying the same thing is that the monopole moment is the angularly averaged surface mass density.

$$\sigma_o(r) = \frac{\int_0^{2\pi} d\theta \sigma(r,\theta)}{\int_0^{2\pi} d\theta\, 1} \tag{6.22}$$

The dipole moment vector $\vec{\sigma}_1(r)$ gives the location of the center of mass of a narrow annulus of matter near radius r. This center of mass is

$$\vec{x}_{\substack{\text{com} \\ \text{ring at } r}} = \lim_{\Delta r \to 0} \frac{\int_r^{r+\Delta r} r' dr' \int_0^{2\pi} d\theta \vec{r}' \sigma(r',\theta)}{\int_r^{r+\Delta r} r' dr' \int_0^{2\pi} d\theta \sigma(r',\theta)}, \tag{6.23}$$

so that

$$\frac{1}{r} \vec{x}_{\substack{\text{com} \\ \text{ring at } r}} = \frac{1}{2} \frac{\vec{\sigma}_1(r)}{\sigma_o(r)}. \tag{6.24}$$

Note that since the surface mass density must be non-negative, the center of mass of the ring cannot fall outside of the ring; therefore from equation 6.24 the amplitude of the dipole moment may not be more than twice the average surface mass density of the ring. Also of interest is that the center of mass of all the matter within distance r of the origin is given by the integral of the dipole moment,

$$\vec{x}_{\substack{\text{com} \\ \leq r}} = \frac{\int_o^r r'^2 dr' \vec{\sigma}_1(r')}{r^2 \sigma_{\substack{\text{average} \\ \leq r}}}, \tag{6.25}$$

where $\sigma_{\substack{\text{average} \\ \leq r}}$ is the average surface mass density of the matter within distance r of the origin.

While the monopole term $\sigma_o(r)$ gives the average surface mass density in an annulus at radius r, the $m \geq 1$ multipole terms give the distribution of matter around that annulus. The dipole $\vec{\sigma}_1(r)$ is related to an offset of matter from the origin, as seen above. The quadrupole moment, $m = 2$ is caused by an elongation of the mass distribution. A triangularity or wedge-shaped-ness would give rise to a non-zero $m = 3$ multipole moment. Boxiness or diskiness of the mass distribution would give rise to an $m = 4$ multipole moment. A change of angle with radius, $\psi_{\sigma,m}(r)$ changing with r, can be caused by "twisted" isodensity contours. In particular, a rank m multipole

moment with orientation angle $\psi_{\sigma,m}(r)$, as defined in equation 6.18, could be caused by an excess of mass at radius r in any of the m equally spaced directions $\psi_{\sigma,m}+n\frac{2\pi}{m}$, where $n=0,1,\ldots,m-1$. Equivalently it could be caused by a deficit of mass in the m "opposite" directions $\psi_{\sigma,m}+(n+\frac{1}{2})\frac{2\pi}{m}$. Therefore knowing the multipole moments of the surface mass density allows one to have an intuitive picture of where the mass is that is causing the multipole moment.

Since the surface mass density of the deflector is non-negative, $\sigma(r,\theta) > 0$, there is an upper limit on the amplitude of $\vec{\sigma}_m(r)$. From equations 6.21 and 6.15, $\sigma_{mc}(r) = \frac{1}{\pi}\int_0^{2\pi} d\theta\, \sigma(r,\theta)\cos m\theta$. Therefore $|\sigma_{mc}(r)| \leq \frac{1}{\pi}\int_0^{2\pi} d\theta\, |\sigma(r,\theta)|\,|\cos m\theta| \leq \frac{1}{\pi}\int_0^{2\pi} d\theta\, \sigma(r,\theta) = 2\sigma_o(r)$, using equation 6.21 in the last step. This may be done (for each r) in a coordinate system chosen so that $\sigma_{ms}(r)$ is zero and $\sigma_{mc}(r) = |\vec{\sigma}_m(r)|$. Therefore we have shown that

$$|\vec{\sigma}_m(r)| \leq 2\sigma_o(r). \tag{6.26}$$

The amplitude of the m^{th} multipole moment of the surface mass density ($m \geq 1$) is less than twice the angularly averaged surface mass density. The multipole moment $\vec{\sigma}_m(r)$ attains its maximum amplitude if all the mass in the annulus at radius r is located at a single angle $\psi_{\sigma,m}(r)$ or is divided (not necessarily evenly) between the equally spaced angles $\psi_{\sigma,m}(r)+n\frac{2\pi}{m}$ where $n=0,1,\ldots,m-1$. If some fraction p ($0 \leq p \leq 1$) of the mass in the ring is in point perturbers at these angles and the rest is evenly distributed around the ring, then

$$|\vec{\sigma}_m(r)| = 2p\sigma_o(r). \tag{6.27}$$

If the perturbers are non-pointlike, then the amplitude of $\vec{\sigma}_m(r)$ is lower.

6.2.4 Taylor expansion of the radial dependencies of the coefficients of the multipole terms

For modeling purposes one may chose to limit oneself to using only the lower multipole components of the mass distribution, as these are sufficient to describe fairly well a smoothly varying mass distribution. From equation 6.14 it is seen that this is the same as limiting oneself to using only the lower multipole components of the 2-D lens potential.

However, the multipole components have coefficients that are arbitrary functions of radius: $\Phi_o(r)$ and $\vec{\Phi}_m(r)$. The relevant part of the radial dependence must be parameterized in just a few parameters in order to be suitable for use in lens modeling. The lens flux gives constraints on $\vec{\nabla}\Phi$ and $\partial_i\partial_j\Phi$ only at the locations of the flux. For Einstein rings and quad lenses this flux is located near the lens system's characteristic ring radius. Since all the constraints on Φ are near the ring radius, Kochanek (1991) suggests Taylor expanding about the ring radius.

To Taylor expand around the Einstein ring radius, one must first define what is meant by the ring radius for a non-circularly-symmetric system. First define an Einstein "curve" for a non-circularly-symmetric system as a closed contour, all of the points of which map back to the same point in the source plane. If there is such a closed curve $\vec{r}(\lambda)$, such that $\vec{s}(\vec{r}(\lambda)) = \vec{s}_o = \text{constant}$ for all λ, then by the divergence theorem, the average surface mass density in the region bounded by the closed curve is the critical surface mass density σ_{crit}. For a circularly symmetric system, this Einstein "curve" is a circle, and the Einstein ring radius is simply its radius, b, which satisfies the relation $b = \frac{d}{dr}\Phi(r)\Big|_{r=b}$. For a non-circularly-symmetric system, the ring radius may be defined using the behavior of the monopole term of the potential, $b = \frac{d}{dr}\Phi_o(r)\Big|_{r=b}$. It can be shown that the average surface mass density, due to all the multipole components, within this radius b is simply the critical surface mass density σ_{crit} — the same average mass density as within the "true" Einstein curve. If the surface mass density falls with radius, then the "true" Einstein curve will lie in some places at $r < b$ and at other places at $r > b$.

For purposes of constructing a model, we will Taylor expand each multipole coefficient, $\Phi_o(r)$ and $\vec{\Phi}_m(r)$ around the Einstein ring radius b, and use as the model the most significant terms in the expansion in powers of $\rho \equiv \left(\frac{r-b}{b}\right)$. The coefficients of these terms will be the model parameters. Therefore we must know how these coefficients relate to the underlying mass distribution, so that we know what the lensing says about the mass distribution. This will be worked out below. First we will name the terms in the expansion.

The potential, as written in a multipole expansion, is

$$\Phi(r,\theta) = \Phi_o(r) + \sum_{m=1}^{\infty} \vec{\Phi}_m(r) \cdot (\hat{x}\cos m\theta + \hat{y}\sin m\theta). \tag{6.28}$$

For each term in the multipole expansion, the radial dependences of $\vec{\Phi}_m(r)$ and $\Phi_o(r)$ are Taylor expanded. The potential is therefore a sum of terms, each term specified by m and t, where m identifies which multipole component and t identifies which term in the Taylor expansion.

$$\Phi(r,\theta) = \sum_{m=0}^{\infty}\sum_{t=0}^{\infty} \Phi_{mt}(r,\theta) \tag{6.29}$$

The terms in the expansion are

$$\Phi_{\substack{m=0\\t\geq 0}}(r,\theta) = \frac{1}{t!}\left\{ b^t \frac{d^t}{dr^t}\Phi_o(r)\Big|_{r=b} \right\} \rho^t \tag{6.30}$$

$$\vec{\Phi}_{\substack{m\geq 1\\t\geq 0}}(r,\theta) = \frac{1}{t!}\left\{ b^t \frac{d^t}{dr^t}\vec{\Phi}_m(r)\Big|_{r=b} \right\} \rho^t \cdot (\hat{x}\cos m\theta + \hat{y}\sin m\theta) \tag{6.31}$$

The θ- and r-dependence of each term $\Phi_{mt}(r,\theta)$ is shown explicitly in equations 6.30 and 6.31, since $\rho = \left(\frac{r-b}{b}\right)$. The $b^t \frac{d^t}{dr^t}\vec{\Phi}_m(r)\Big|_{r=b}$ factor is just a two-component constant with no r- or θ-dependence. Likewise, the $b^t \frac{d^t}{dr^t}\Phi_o(r)\Big|_{r=b}$ factor is a scalar constant with no r- or θ-dependence. These constants are used as parameters of a potential model. These parameters will be related to the underlying mass distribution.

The expressions for the multipole moments of the potential as functions of the multipole components of the surface mass density are give in equation 6.14. The

derivatives as evaluated at the ring radius are:

$$\frac{1}{b^2}\Phi_o(b) = \text{constant} \tag{6.32}$$

$$\frac{1}{b^2} b \frac{d}{dr}\Phi_o(r)\bigg|_{r=b} = 1, \text{ by the definition of the ring radius} \tag{6.33}$$

$$f_{t=2} \equiv \frac{1}{b^2} b^2 \frac{d^2}{dr^2}\Phi_o(r)\bigg|_{r=b} = -1 + \frac{2\sigma_o(b)}{\sigma_{\rm crit}} \tag{6.34}$$

$$f_{t=3} \equiv \frac{1}{b^2} b^3 \frac{d^3}{dr^3}\Phi_o(r)\bigg|_{r=b} = 2 - \frac{2\sigma_o(b)}{\sigma_{\rm crit}} + 2\, b \frac{d}{dr}\left(\frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\bigg|_{r=b} \tag{6.35}$$

$$f_{t\geq 2} \equiv \frac{1}{b^2} b^t \frac{d^t}{dr^t}\Phi_o(r)\bigg|_{r=b} \quad \text{for general } t,\ t \geq 2 \tag{6.36}$$

and

$$\vec{M}_{\substack{\rm sum \\ m}} \equiv -\frac{1}{b^2}\vec{\Phi}_m(b) = \left\{\vec{A}_m + \vec{B}_m\right\} \tag{6.37}$$

$$m\vec{M}_{\substack{\rm diff \\ m}} \equiv -\frac{1}{b^2} b \frac{d}{dr}\vec{\Phi}_m(r)\bigg|_{r=b} = \left\{m(\vec{A}_m - \vec{B}_m)\right\} \tag{6.38}$$

$$\vec{F}_{\substack{m \\ t=2}} \equiv -\frac{1}{b^2} b^2 \frac{d^2}{dr^2}\vec{\Phi}_m(r)\bigg|_{r=b} = \left\{m(m-1)\vec{A}_m - m(-m-1)\vec{B}_m - \frac{2\vec{\sigma}_m(b)}{\sigma_{\rm crit}}\right\} \tag{6.39}$$

$$\vec{F}_{\substack{m \\ t=3}} \equiv -\frac{1}{b^2} b^3 \frac{d^3}{dr^3}\vec{\Phi}_m(r)\bigg|_{r=b} = \Bigg\{m(m-1)(m-2)\vec{A}_m - m(-m-1)(-m-2)\vec{B}_m + \frac{2\vec{\sigma}_m(b)}{\sigma_{\rm crit}} - 2\, b \frac{d}{dr}\frac{\vec{\sigma}_m(r)}{\sigma_{\rm crit}}\bigg|_{r=b}\Bigg\} \tag{6.40}$$

$$\vec{F}_{\substack{m\geq 1 \\ t\geq 2}} \equiv -\frac{1}{b^2} b^t \frac{d^t}{dr^t}\vec{\Phi}_m(r)\bigg|_{r=b} \quad \text{for general } m \text{ and } t, \text{ for } m \geq 1 \text{ and } t \geq 2 \tag{6.41}$$

The quantity $\vec{A}_m$ is attributable to the mth mass multipole of the mass exterior to the ring radius,

$$\vec{A}_m = \frac{b^{m-2}}{m}\int_b^\infty r'^{(1-m)} dr' \frac{\vec{\sigma}_m(r')}{\sigma_{\rm crit}}. \tag{6.42}$$

The quantity $\vec{B}_m$ is attributable to the mass interior to the ring radius,

$$\vec{B}_m = \frac{1}{m}\frac{1}{b^{m+2}}\int_0^b r'^{(1+m)} dr' \frac{\vec{\sigma}_m(r')}{\sigma_{\rm crit}}. \tag{6.43}$$

The reason for the special treatment in naming the coefficients of the constant and linear terms in the expansion of the $m \geq 1$ multipole, $\vec{M}_{\underset{m}{\text{sum}}}$ and $\vec{M}_{\underset{m}{\text{diff}}}$, is clear since they are related to the sum and difference respectively of the effects of the mass located exterior to the ring radius and the mass located interior to the ring radius.

These parameters b, f_t for $t \geq 2$, $\vec{M}_{\underset{m}{\text{sum}}} = \hat{x}M_{\underset{mc}{\text{sum}}} + \hat{y}M_{\underset{ms}{\text{sum}}} = M_{\underset{m}{\text{sum}}}(\hat{x}\cos m\psi_{\underset{m}{\text{sum}}} + \hat{y}\sin m\psi_{\underset{m}{\text{sum}}})$ for $m \geq 1$, $\vec{M}_{\underset{m}{\text{diff}}} = \hat{x}M_{\underset{mc}{\text{diff}}} + \hat{y}M_{\underset{ms}{\text{diff}}} = M_{\underset{m}{\text{diff}}}(\hat{x}\cos m\psi_{\underset{m}{\text{diff}}} + \hat{y}\sin m\psi_{\underset{m}{\text{diff}}})$ for $m \geq 1$, and $\vec{F}_{mt} = \hat{x}F_{mtc} + \hat{y}F_{mts} = F_{mt}(\hat{x}\cos m\psi_{F_{mt}} + \hat{y}\sin m\psi_{F_{mt}})$ for $m \geq 1$ and $t \geq 2$, will be used as parameters in the model. The multipole parameters can be specified either as cosine and sine components, for example F_{mtc} and F_{mts}, or as amplitude and angle, for example F_{mt} and $\psi_{F_{mt}}$. The cosine and sine parameterization is more robustly handled by minimization algorithms, since the potential is linear in these parameters and thus the chi-squared is roughly quadratic in them, at least far from the minimum. However the parameter error estimates on the amplitude-angle parameterization are more useful for insight into the location of the mass, since the amplitude is affected by the mass sheet degeneracy (see below) while the angle is not affected.

Written in terms of these parameters, the potential is,

$$\begin{aligned}\frac{1}{b^2}\Phi(r,\theta) &= \text{const} + \rho + \frac{1}{2}\rho^2 f_2 + \sum_{t=3}^{\infty}\frac{1}{t!}\rho^t f_t \\ &\quad - \sum_{m=1}^{\infty}\left(\left\{\vec{M}_{\underset{m}{\text{sum}}} + \rho m\vec{M}_{\underset{m}{\text{diff}}} + \sum_{t=2}^{\infty}\frac{1}{t!}\rho^t\vec{F}_{mt}\right\}\cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta)\right).\end{aligned} \tag{6.44}$$

The constant term in the expansion of the monopole can be ignored, as it just adds an overall constant level to the potential. The linear term in the expansion of the monopole sets the ring radius and does not have any other adjustable coefficient. Two more model parameters which are hidden in this way of writing the potential are g_x and g_y, the location of the origin. This origin should be centered on the deflector in order for the multipole components of the surface mass density to be readily understandable; section 6.2.5 below shows how to enforce this when fitting the model to a lens.

6.2.5 Effects of lensing degeneracies on the multipole-Taylor model

Certain parameters in this Multipole-Taylor expansion, equation 6.44, must be fixed due to degeneracies between the parameters. When a degenerate parameter is changed in value, the other parameters with which it is degenerate can be adjusted in compensation so that there is no change in the model predictions for the lensed emission. Therefore such degeneracies must be removed before a minimization algorithm can minimize a goodness-of-fit measure.

Special attention is needed for the $t = 0$ and $t = 1$ parts of the dipole term,

$$\begin{aligned} \frac{1}{b^2}\Phi_{\substack{m=1\\t=0}} + \frac{1}{b^2}\Phi_{\substack{m=1\\t=1}} &= \left\{-\vec{M}_{\substack{\text{sum}\\m=1}} - \rho\vec{M}_{\substack{\text{diff}\\m=1}}\right\} \cdot (\hat{x}\cos\theta + \hat{y}\sin\theta) \\ &= \left\{-\frac{r}{b}\vec{A}_1 + (\rho - 1)\vec{B}_1\right\} \cdot (\hat{x}\cos\theta + \hat{y}\sin\theta). \end{aligned} \tag{6.45}$$

$\vec{B}_1$ is the contribution to the dipole term of the matter interior to the ring radius. By equations 6.43 and 6.25, this is simply the center-of-mass of the matter interior to the Einstein ring radius,

$$\vec{B}_1 = \frac{1}{b^3\sigma_{\text{crit}}}\int_0^b r'^2 dr' \vec{\sigma}_1(\vec{r}') = \frac{1}{b}\vec{x}_{\substack{\text{com}\\\text{matter within ring}}}. \tag{6.46}$$

It is desirable for the center of mass of the matter interior to the ring to be at the origin. This is done by fixing the two parameters $\vec{B}_1 = 0$, and allowing the location of the origin (g_x, and g_y) to be fitted in the modeling. Note that one may fit for either $\vec{B}_1$ or for the deflector location (g_x, g_y), but not for both simultaneously, due to a degeneracy.

Once $\vec{B}_1$ has been fixed to zero,

$$\frac{1}{b^2}\Phi_{\substack{m=1\\t=0}} + \frac{1}{b^2}\Phi_{\substack{m=1\\t=1}} = -\frac{1}{b}\vec{A}_1 \cdot \vec{r}, \tag{6.47}$$

where $\vec{A}_1$ is the contribution to the dipole of mass exterior to the ring. This is precisely the prismatic degeneracy, one of several transformations to the potential that have no

effect on the relative locations and brightnesses of the lensed flux. (These degeneracies were pointed out by Falco, Gorenstein, and Shapiro (Falco *et al.* 1985; Gorenstein *et al.* 1988) and are discussed in section 4.1.6 of this thesis.) This term may be ignored without affecting image locations, image fluxes, or the time delay. Therefore, $\Phi_{\substack{m=1\\t=0}}$ and $\Phi_{\substack{m=1\\t=1}}$ will be fixed to zero in the deflector model.

The other degeneracy to be considered is the mass sheet degeneracy. The transformation from the true deflector potential Φ to an alternate model potential Φ',

$$\Phi(\vec{r}) \longrightarrow \Phi'(\vec{r}) = (1-\kappa)\Phi(\vec{r}) + \frac{1}{2}\kappa|\vec{r}|^2, \tag{6.48}$$

leaves the image locations and fluxes unchanged. This is equivalent to reducing the surface mass density by multiplying it by a factor $(1-\kappa)$ and adding in a sheet of uniform surface mass density $\kappa\sigma_{\rm crit}$,

$$\sigma(\vec{r}) \longrightarrow \sigma'(\vec{r}) = (1-\kappa)\,\sigma(\vec{r}) + \kappa\sigma_{\rm crit}. \tag{6.49}$$

Adding such a mass sheet to the true deflector potential Φ and scaling the potential, as in equation 6.48, gives an expression for this model potential Φ' that predicts the same image locations and fluxes as does the true potential Φ.

$$\begin{aligned}
\frac{1}{b^2}\Phi(\vec{r}) \longrightarrow \frac{1}{b^2}\Phi'(\vec{r}) \;=\; & \underbrace{\left\{(1-\kappa)\text{constant} + \frac{1}{2}\kappa\right\}}_{\text{constant}'} + \rho + \frac{1}{2}\rho^2 \underbrace{\{(1-\kappa)f_2 + \kappa\}}_{f'_2} \\
& + \sum_{t=3}^{\infty} \frac{1}{t!}\rho^t \underbrace{(1-\kappa)f_t}_{f'_t} \\
& - \sum_{m=2}^{\infty} \underbrace{(1-\kappa)\vec{M}_{\substack{\rm sum\\ m}}}_{\vec{M}'_{\substack{\rm sum\\ m}}} \cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta) \\
& - \sum_{m=2}^{\infty} \rho m \underbrace{(1-\kappa)\vec{M}_{\substack{\rm diff\\ m}}}_{\vec{M}'_{\substack{\rm diff\\ m}}} \cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta) \\
& - \sum_{m=1}^{\infty}\sum_{t=2}^{\infty} \frac{1}{t!}\rho^t \underbrace{(1-\kappa)\vec{F}_{mt}}_{\vec{F}'_{mt}} \cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta)
\end{aligned} \tag{6.50}$$

The constant offset has been altered, but this is immaterial since adding a constant to the potential has no effect. The transformation must leave the ring radius unchanged, therefore the coefficient of ρ in the monopole part of the potential is unchanged. The quadratic monopole term, which depends on the surface mass density at the ring radius, $f_2 = -1 + \frac{2\sigma_o(b)}{\sigma_{\rm crit}}$, has been changed,

$$f_2 \longrightarrow f_2' = (1-\kappa)\, f_2 + \kappa, \tag{6.51}$$

because the surface mass density at the ring radius has been changed. All the coefficients of the other terms in the model, $f_{t\geq 3}$, $\vec{M}_{\substack{\rm sum \\ m}}$, $\vec{M}_{\substack{\rm diff \\ m}}$, and $\vec{F}_{mt}$, have been scaled by a factor $1-\kappa$. These simultaneous adjustments of f_2 and all the $f_{t\geq 3}$, $\vec{M}_{\substack{\rm sum \\ m}}$, $\vec{M}_{\substack{\rm diff \\ m}}$, and $\vec{F}_{mt}$ parameters leave the model predictions unchanged.

This factor $1-\kappa$ is unconstrainable in lens modeling. Thus one of the degenerate parameters must be fixed to some value, rather than being fitted. The parameter f_2' will be fixed, and the value to which it is fixed may be incorrect. (Fixing f_2' is equivalent to fixing the surface mass density at the ring radius to the possibly incorrect value $\sigma_o'(b) = \frac{1}{2}\sigma_{\rm crit}(1+f_2')$.) Therefore all the fitted values of the other parameters will be wrong by a factor $(1-\kappa) = \frac{1-f_2'}{1-f_2} = \left(1 - \frac{\sigma_o'(b)}{\sigma_{\rm crit}}\right)/\left(1 - \frac{\sigma_o(b)}{\sigma_{\rm crit}}\right)$. For example, the fitted value $\vec{F}'_{mt}$ will be related to the true value $\vec{F}_{mt}$ by $\vec{F}'_{mt} = (1-\kappa)\vec{F}_{mt}$. See table 6.2 where the value of $1-\kappa$ is tabulated for several possible radial profiles for the "true" and the model deflector potentials.

Most simple is to fix $f_2' = 0$, which is the value appropriate for the radial profile of a singular isothermal sphere. Then the model to be fitted is the model with $\vec{M}_{\substack{\rm sum \\ m=1}}$, $\vec{M}_{\substack{\rm diff \\ m=1}}$, f_2', and the additive constant set to zero:

$$\begin{aligned}
\frac{1}{b^2}\Phi'(\vec{r}) \;=\; & \rho + \sum_{t=3}^{\infty} \frac{1}{t!}\rho^t f_t' \\
& - \sum_{m=2}^{\infty} \left\{ \vec{M}'_{\substack{\rm sum \\ m}} + \rho m \vec{M}'_{\substack{\rm diff \\ m}} \right\} \cdot (\hat{x}\cos m\theta + \hat{y}\sin m\theta) \\
& - \sum_{m=1}^{\infty}\sum_{t=2}^{\infty} \frac{1}{t!}\rho^t \vec{F}'_{mt} \cdot (\hat{x}\cos m\theta + \hat{y}\sin m\theta).
\end{aligned} \tag{6.52}$$

Scaling factor $(1-\kappa)$			
	Value to which $\sigma'_o(b)$ and f'_2 are fixed for modeling		
	point mass: $\sigma'_o(b)=0$	singular isothermal sphere: $\sigma'_o(b)=\sigma_{\rm crit}/2$	mass sheet: $\sigma'_o(b)=\sigma_{\rm crit}$
True deflector monopole	$f'_2=-1$	$f'_2=0$	$f'_2=1$
point mass: $\Phi_o(r)=b^2\ln r$ $f_2=-1$, $\sigma_o(b)=0$	1	1/2	0
singular isothermal sphere: $\Phi_o(r)=br$ $f_2=0$, $\sigma_o(b)=\sigma_{\rm crit}/2$	2	1	0
mass sheet: $\Phi_o(r)=\frac{1}{2}r^2$ $f_2=1$, $\sigma_o(b)=\sigma_{\rm crit}$	∞	∞	1
power law: $\Phi_o(r)=\frac{b^2}{2\alpha}\left(\frac{r}{b}\right)^{2\alpha}$ $\alpha\to 0$, point mass $\alpha=1/2$, isothermal $\alpha\to 1$, mass sheet $f_2=2\alpha-1$, $\sigma_o(b)=\alpha\sigma_{\rm crit}$	$\frac{2}{2(1-\alpha)}$	$\frac{1}{2(1-\alpha)}$	$\frac{0}{2(1-\alpha)}$
general case: f_2, $\sigma_o(b)$	$\frac{2}{2(1-\frac{\sigma_o(b)}{\sigma_{\rm crit}})}$	$\frac{1}{2(1-\frac{\sigma_o(b)}{\sigma_{\rm crit}})}$	$\frac{0}{2(1-\frac{\sigma_o(b)}{\sigma_{\rm crit}})}$

Table 6.2: Scaling factors due to the interaction of the mass sheet degeneracy with a mismatch between true and model monopole profiles. This shows the factor $1-\kappa$ by which the Multipole-Taylor parameters $f_{t\geq 3}$, $\vec{M}_{\underset{m}{\rm sum}}$, $\vec{M}_{\underset{m}{\rm diff}}$, and $\vec{F}_{mt}$ are scaled, if there is a mismatch between the actual value of the surface mass density at the location of the ring radius, $\sigma_o(b)$, and the value to which this is fixed for the purpose of modeling, $\sigma'_o(b)$.

The column at left lists several possibilities for the true deflector potential monopole, $\Phi_o(r)$, with the corresponding true values of the surface mass density at the ring radius $\sigma_o(r)$ and of the quadratic coefficient f_2. The three columns at right give the scaling factors $(1-\kappa)$ that there will be if these potentials are modeled using, respectively, a point mass monopole profile, a singular isothermal sphere monopole profile, and (as a limiting case) a mass sheet with critical surface mass density.

For lenses in which the $t\geq 3$ terms do little to constrain the radial dependence of the potential, there is a near degeneracy between all model monopole profiles, which affects the fitted amplitudes of the $m\geq 1$ multipoles and the time delay. For example, in such a case, all multipole amplitudes are twice as large when using a point mass profile for the model than when using a singular isothermal sphere profile (Kochanek 1991), and the model predicted time delay is twice as long — whatever the actual radial profile of the potential.

Note that this rescaling does not affect the angles of the multipole coefficients. However it does make it more difficult to understand the amplitudes of the $m \geq 1$ multipole components. The relationship of these rescaled parameters to the actual surface mass distribution σ and the true potential Φ of the deflector is:

$$f'_3 = 1 - b\frac{d}{dr}\ln\left(1 - \frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\Big|_{r=b} \tag{6.53}$$

$$\begin{aligned} f'_{t\geq 3} &= (1-\kappa) f_{t\geq 3} \\ &= (1-\kappa)\,\frac{1}{b^2} b^t \frac{d^t}{dr^t}\Phi_o(r)\Big|_{r=b} \quad \text{for general } t,\ t \geq 3 \end{aligned} \tag{6.54}$$

and

$$\vec{M}'_{\substack{\rm sum \\ m\geq 2}} = (1-\kappa)\vec{M}_{\substack{\rm sum \\ m\geq 2}} = (1-\kappa)\left\{\vec{A}_m + \vec{B}_m\right\} \tag{6.55}$$

$$m\vec{M}'_{\substack{\rm diff \\ m\geq 2}} = (1-\kappa)m\vec{M}_{\substack{\rm diff \\ m\geq 2}} = (1-\kappa)\left\{m(\vec{A}_m - \vec{B}_m)\right\} \tag{6.56}$$

$$\vec{F}'_{\substack{m \\ t=2}} = +(1-\kappa)\left\{m(m-1)\vec{A}_m - m(-m-1)\vec{B}_m - \frac{2\vec{\sigma}_m(b)}{\sigma_{\rm crit}}\right\} \tag{6.57}$$

$$\begin{aligned} \vec{F}'_{\substack{m \\ t=3}} = (1-\kappa)\Big\{ & m(m-1)(m-2)\vec{A}_m \\ & -m(-m-1)(-m-2)\vec{B}_m \\ & +\frac{2\vec{\sigma}_m(b)}{\sigma_{\rm crit}} - 2\,\frac{d}{d\vec{r}_b}\frac{\vec{\sigma}_m(b)}{\sigma_{\rm crit}}\Big|_{r=b}\Big\} \end{aligned} \tag{6.58}$$

$$\begin{aligned} \vec{F}'_{\substack{m\geq 1 \\ t\geq 2}} &= (1-\kappa)\vec{F}_{\substack{m\geq 1 \\ t\geq 2}} \text{ for general } m \text{ and } t, \text{ for } m \geq 1 \text{ and } t \geq 2 \\ &= -(1-\kappa)\,\frac{1}{b^2} b^t \frac{d^t}{dr^t}\vec{\Phi}_m(r)\Big|_{r=b} \end{aligned} \tag{6.59}$$

6.2.6 Each term in the multipole-Taylor model discussed

For the purposes of lens modeling we need expressions for the first and second[1] derivatives of the model potential, $\vec{\nabla}_{\vec{r}}\Phi'$ and $\overleftarrow{\nabla}_{\vec{r}}\vec{\nabla}_{\vec{r}}\Phi'$. Thus for each term $\Phi'_{mt}(r,\theta)$ the first

[1]It is convenient to use a dyadic notation. (See, e.g., Goldstein (1980) chapter 5.) The antecedent vector of a dyad $\vec{r}\vec{r}$ will be written as $\vec{r}$ and the consequent as $\vec{r}$ as a reminder that this is a dyad. However when the antecedent and the consequent are unit vectors, the dyad will be written, for example, as $\hat{r}\hat{r}$.

and second derivatives must be found. The first and second derivatives will be written in terms of $\hat{r}$, $\hat{\theta}$, $\hat{r}\hat{r}$, $\hat{\theta}\hat{\theta}$, and $(\hat{r}\hat{\theta} + \hat{\theta}\hat{r})$. To be explicit, their components in Cartesian coordinates are shown.

$$\hat{r} = \hat{x}\cos\theta + \hat{y}\sin\theta = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \tag{6.60}$$

$$\hat{\theta} = -\hat{x}\sin\theta + \hat{y}\cos\theta = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \tag{6.61}$$

$$\vec{\nabla}_{\vec{r}} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix} \tag{6.62}$$

$$\hat{r}\hat{r} = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix} \tag{6.63}$$

$$\hat{\theta}\hat{\theta} = \begin{pmatrix} \sin^2\theta & -\cos\theta\sin\theta \\ -\cos\theta\sin\theta & \cos^2\theta \end{pmatrix} \tag{6.64}$$

$$\hat{r}\hat{\theta} + \hat{\theta}\hat{r} = \begin{pmatrix} -2\cos\theta\sin\theta & (\cos^2\theta - \sin^2\theta) \\ (\cos^2\theta - \sin^2\theta) & 2\cos\theta\sin\theta \end{pmatrix} \tag{6.65}$$

$$\overleftarrow{\nabla}_{\vec{r}}\vec{\nabla}_{\vec{r}} = \begin{pmatrix} \frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y^2} \end{pmatrix} \tag{6.66}$$

The constant term in the Taylor expansion of the monopole term of the potential is:

$$\frac{1}{b^2}\Phi'_{\substack{m=0 \\ t=0}}(r,\theta) = \text{constant}' \tag{6.67}$$

$$\frac{1}{b^2}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m=0 \\ t=0}}(r,\theta) = 0 \tag{6.68}$$

$$\frac{1}{b^2}b\overleftarrow{\nabla}_{\vec{r}}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m=0 \\ t=0}}(r,\theta) = 0 \tag{6.69}$$

This is just a constant offset in the potential. It may be ignored, since only potential differences are of importance.

The linear term in the Taylor expansion of the monopole term of the potential is:

$$\frac{1}{b^2}\Phi'_{\substack{m=0\\t=1}}(r,\theta) = \rho \tag{6.70}$$

$$\frac{1}{b^2}b\vec{\nabla}_{\hat{r}}\Phi'_{\substack{m=0\\t=1}}(r,\theta) = \hat{r} \tag{6.71}$$

$$\frac{1}{b^2}b\overleftarrow{\nabla}_{\hat{r}}b\vec{\nabla}_{\hat{r}}\Phi'_{\substack{m=0\\t=1}}(r,\theta) = \hat{\theta}\hat{\theta}\left\{\frac{1}{1+\rho}\right\} \tag{6.72}$$

This term in the expansion has no adjustable parameters besides the Einstein ring radius itself. Therefore this term sets the ring radius, b. The ring radius is the radius within which the average surface mass equals the critical surface mass density $\sigma_{\rm crit}$, therefore this term is sensitive to the average mass density within the ring radius.

This term causes image displacements only in the radial direction. It causes image magnification only in the tangential direction. This term, by itself, is the singular isothermal sphere potential, and if it were the only non-zero term would cause infinite tangential magnification at the ring radius $\rho = 0$.

The higher terms in the Taylor expansion of the monopole term of the potential are:

$$\frac{1}{b^2}\Phi'_{\substack{m=0\\t\geq 2}}(r,\theta) = \frac{1}{t!}\rho^t f'_t \tag{6.73}$$

$$\frac{1}{b^2}b\vec{\nabla}_{\hat{r}}\Phi'_{\substack{m=0\\t\geq 2}}(r,\theta) = \hat{r}\left\{\frac{1}{(t-1)!}\rho^{(t-1)}\right\}\{f'_t\} \tag{6.74}$$

$$\frac{1}{b^2}b\overleftarrow{\nabla}_{\hat{r}}b\vec{\nabla}_{\hat{r}}\Phi'_{\substack{m=0\\t\geq 2}}(r,\theta) = \hat{r}\hat{r}\left\{\frac{1}{(t-2)!}\rho^{(t-2)}\right\}\{f'_t\} + \hat{\theta}\hat{\theta}\left\{\frac{1}{(t-1)!}\frac{\rho^{(t-1)}}{(1+\rho)}\right\}\{f'_t\} \tag{6.75}$$

While these expressions are valid for $t \geq 2$, the quadratic $t = 2$ term in the expansion is unconstrainable due to the mass-sheet degeneracy, as described above. The coefficient of this quadratic $m = 0$, $t = 2$ term depends only on the surface mass density at the ring radius. (See equation 6.34.) Accordingly, the surface mass density at the ring radius is not directly constrained by lens modeling. This is disappointing. Different radial profiles, with the same ring radius, could be distinguished by the surface mass density at the ring radius — were this information available through lens modeling. The first monopole term that does give information on the radial profile, is the cubic

term, $t = 3$, which actually becomes easier to interpret given the mass-sheet degeneracy. (See equations 6.35 and 6.53.) This term depends only on how the surface mass density near the Einstein ring radius falls with radius, being more sensitive to radial falloff when the surface mass density at the ring radius, $\sigma_o(b)$, is close to the critical surface mass density $\sigma_{\rm crit}$ than when $\sigma_o(b)$ is much less than $\sigma_{\rm crit}$:

$$f_3' = \frac{1}{2(1-\frac{\sigma_o(b)}{\sigma_{\rm crit}})} f_3 = 1 - b\frac{d}{dr}\ln\left(1-\frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\Big|_{r=b}. \tag{6.76}$$

This term in the potential is proportional to $\rho^3 = \left(\frac{r-b}{b}\right)^3$, which is less than 1% for each of the images of the quad lens MG J0414+0534, and is expected to be similarly small for other quad lenses and rings, therefore the radial profile of the deflector mass distribution is expected to be quite difficult to constrain for such lenses. If the f_3' parameter can be constrained, then one can draw conclusions about the radial profile: The f_3' parameter is ≤ 1 for surface mass densities that do not increase with radius. For a point mass, $f_3' = 1$. For a singular isothermal sphere, $f_3' = 0$. For a power law potential $\Phi_o(r) = \frac{b^2}{2\alpha}\left(\frac{r^2}{b^2}\right)^\alpha$ then $f_3' = 1 - 2\alpha$. Since α ranges from 0 for a point mass, through $\alpha = 1/2$ for an isothermal model, to $\alpha = 1$ for a mass sheet, then $-1 \leq f_3' \leq 1$ for such a model. For a potential with a core radius, or for any mass distribution in which $\sigma_o(r)$ drops abruptly near the ring radius, f_3' can be even more negative.

Note that all the monopole terms in the expansion can cause only radial displacements of the images, and not tangential displacements, however for $t \geq 2$ the radial displacements are small for small ρ. These higher monopole terms cause radial magnification of images, which for $t \geq 3$ are small for small ρ. These higher monopole terms produce even less tangential magnification than they do radial magnification.

The constant $t = 0$ piece in the Taylor expansion of the dipole and higher multipole terms $m \geq 1$ is:

$$\frac{1}{b^2}\Phi'_{\substack{m\geq 1\\ t=0}}(r,\theta) = \{-1\}\left\{M'_{\substack{\text{sum}\\ mc}}\cos m\theta + M'_{\substack{\text{sum}\\ ms}}\sin m\theta\right\} \tag{6.77}$$

$$\frac{1}{b^2}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t=0}}(r,\theta) \;=\; \hat{\theta}\left\{\frac{m}{1+\rho}\right\}\left\{M'_{\substack{\text{sum}\\ mc}}\sin m\theta - M'_{\substack{\text{sum}\\ ms}}\cos m\theta\right\} \tag{6.78}$$

$$\begin{aligned}
&\frac{1}{b^2}b\overleftarrow{\nabla}_{\vec{r}}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t=0}}(r,\theta)\\
&= \hat{\theta}\hat{\theta}\left\{\frac{m^2}{(1+\rho)^2}\right\}\left\{M'_{\substack{\text{sum}\\ mc}}\cos m\theta + M'_{\substack{\text{sum}\\ ms}}\sin m\theta\right\}\\
&+ (\hat{r}\hat{\theta}+\hat{\theta}\hat{r})\left\{-\frac{m}{(1+\rho)^2}\right\}\left\{M'_{\substack{\text{sum}\\ mc}}\sin m\theta - M'_{\substack{\text{sum}\\ ms}}\cos m\theta\right\}
\end{aligned} \tag{6.79}$$

These formulas are valid for $m \geq 1$, but setting the $m = 1$, $t = 1$ and $m = 1$, $t = 0$ terms to zero is used to force the center-of-mass of the matter inside the ring radius to be at the origin, and to avoid the prismatic degeneracy. Note that $M'_{\substack{\text{sum}\\ mc}}\cos m\theta + M'_{\substack{\text{sum}\\ ms}}\sin m\theta$ can also be written in terms of the amplitude and angle of the multipole moment as $M'_{\substack{\text{sum}\\ m}}\cos m(\theta - \psi_{\substack{\text{sum}\\ m}})$, where the angle is unaffected by the mass sheet degeneracy. Note also that $M'_{\substack{\text{sum}\\ mc}}\sin m\theta - M'_{\substack{\text{sum}\\ ms}}\cos m\theta = M'_{\substack{\text{sum}\\ m}}\sin m(\theta - \psi_{\substack{\text{sum}\\ m}})$.

This first $t = 0$ term of each multipole component results in only tangential, not radial, image displacements, and these displacements can be of significant size, even for small ρ. Mass perturbations, whether interior or exterior to the ring radius, contribute in the same sense to the tangential displacements, since they contribute in the same sense to $\vec{M}_{\substack{\text{sum}\\ m}}$. In particular, perturbing masses, interior or exterior to the ring, located at angles $\psi_{\substack{\text{sum}\\ m}} + n\frac{2\pi}{m}$ for $n = 0, \ldots, m-1$, could produce such a multipole term. For the m^{th} multipole, the displacements of the image from the source, $\vec{\nabla}\Phi$, are displacements away from the angles $\psi_{\substack{\text{sum}\\ m}} + n\frac{2\pi}{m}$ for $n = 0, \ldots, m-1$, the angles at which the perturbing mass may be located.

<table>
<tr><th></th><th>$m = 0$
monopole</th><th>$m = 1$
dipole</th><th>$m \geq 2$
quadrupole, octupole, ...</th></tr>
<tr><td>$t = 0$</td><td>Degenerate
Constant offset to potential</td><td rowspan="2">Degenerate
Fixes location of center of mass of mass interior to ring to be at the origin. (Location of origin is then allowed to vary.) Prismatic degeneracy.</td><td>Multipole-Constant
Sum of exterior and interior contributions to the multipole moment of the mass distribution.</td></tr>
<tr><td>$t = 1$</td><td>Monopole-Linear
Sets Einstein ring radius. Sensitive to total mass within ring.</td><td>Multipole-Linear
Difference of exterior and interior contributions to the multipole moment of the mass distribution.</td></tr>
<tr><td>$t = 2$</td><td>Degenerate
Mass sheet degeneracy. Surface mass density at ring unconstrainable</td><td colspan="2" rowspan="2">Multipole-Taylor
Some dependence on the behavior of the multipole moment of the matter near the ring radius. Must detangle effects of interior and exterior mass.</td></tr>
<tr><td>$t \geq 3$</td><td>Monopole-Taylor
For $t = 3$ gives the falloff of surface mass density with radius, evaluated at the ring radius.</td></tr>
</table>

Table 6.3: Physical significance of terms in the Multipole-Taylor expansion. Certain terms are unconstrainable due to degeneracies as explained in the text. Five functional forms are needed to represent all the other terms in the expansion. These five forms are all that need be coded into a lens modeling program. They can be used with m and t fixed to the appropriate values.

The linear $t = 1$ part in the Taylor expansion of the dipole and higher multipole terms $m \geq 1$ is:

$$\frac{1}{b^2}\Phi'_{\substack{m\geq 1\\ t=1}}(r,\theta) = \{-\rho m\}\left\{M'_{\substack{\text{diff}\\ mc}}\cos m\theta + M'_{\substack{\text{diff}\\ ms}}\sin m\theta\right\} \tag{6.80}$$

$$\begin{aligned}\frac{1}{b^2}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t=1}}(r,\theta) &= \hat{r}\{-m\}\left\{M'_{\substack{\text{diff}\\ mc}}\cos m\theta + M'_{\substack{\text{diff}\\ ms}}\sin m\theta\right\} \\ &+ \hat{\theta}\left\{\frac{\rho m^2}{1+\rho}\right\}\left\{M'_{\substack{\text{diff}\\ mc}}\sin m\theta - M'_{\substack{\text{diff}\\ ms}}\cos m\theta\right\}\end{aligned} \tag{6.81}$$

$$\begin{aligned}&\frac{1}{b^2}b\overleftarrow{\nabla}_{\vec{r}}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t=1}}(r,\theta) \\ &= \hat{\theta}\hat{\theta}\left\{-\frac{m}{(1+\rho)^2}\left(1-(m^2-1)\rho\right)\right\}\left\{M'_{\substack{\text{diff}\\ mc}}\cos m\theta + M'_{\substack{\text{diff}\\ ms}}\sin m\theta\right\} \\ &+ (\hat{r}\hat{\theta}+\hat{\theta}\hat{r})\left\{\frac{m^2}{(1+\rho)^2}\right\}\left\{M'_{\substack{\text{diff}\\ mc}}\sin m\theta - M'_{\substack{\text{diff}\\ ms}}\cos m\theta\right\}\end{aligned} \tag{6.82}$$

These formula are valid for $m \geq 1$, but setting the $m = 1$, $t = 1$ and $m = 1$, $t = 0$ terms to zero is used to force the center-of-mass of the matter interior to the ring radius to be at the origin, and to avoid the prismatic degeneracy.

This second, $t = 1$, term in each multipole component results in primarily radial, not tangential image displacements, since ρ is small. The sense of the image displacements $\vec{\nabla}\Phi$ is inward for $\psi_{\substack{\text{diff}\\ m}} + (n-\frac{1}{4})\frac{2\pi}{m} < \theta < \psi_{\substack{\text{diff}\\ m}} + (n-\frac{1}{4})\frac{2\pi}{m}$, for $n = 0,\ldots,m-1$. That is, the sense of the image displacements changes sign every $\frac{\pi}{m}$ radians around the ring, and the image displacement is inward for $\theta = \psi_{\substack{\text{diff}\\ m}}$.

Mass perturbations interior and exterior to the ring radius contribute in opposite sense to the radial displacements, since they contribute in the opposite sense to $\vec{M}_{\substack{\text{diff}\\ m}}$. In particular, perturbing mass exterior to the ring located at angles $\psi_{\substack{\text{diff}\\ m}} + n\frac{2\pi}{m}$ could produce such a multipole moment, or perturbing mass interior to the ring located at angles $\psi_{\substack{\text{diff}\\ m}} + (n+\frac{1}{2})\frac{2\pi}{m}$ could produce such a multipole moment.

Comparing the first two Taylor terms in the m^{th} multipole, the tangential displacements ($t = 0$) give the overall strength of the m^{th} multipole component, while

the direction of the radial displacements ($t = 1$) indicates whether the mass that is driving this multipole component is interior or exterior to the ring.

Note that the radial displacement can be significant even when ρ is small, thus the size of the radial displacement from multipole terms can be much larger than the size of radial displacement from $t \geq 3$ monopole terms. This makes it hard to constrain $t \geq 3$ monopole terms when the lens images sample only a few different angular locations around the ring (such as for quad lenses); that is, it is difficult to constrain the radial profile of the deflector angularly-averaged mass distribution. Ring lenses may be able to better constrain the $m \geq 2$ multipole terms, to permit the $t \geq 3$ monopole terms to be constrained.

The general term, applying to the quadratic and higher, $t \geq 2$, Taylor terms of the dipole and higher, $m \geq 1$, multipole terms is:

$$\frac{1}{b^2}\Phi'_{\substack{m\geq 1\\ t\geq 2}}(r,\theta) = \left\{-\frac{\rho^t}{t!}\right\}\{F'_{mtc}\cos m\theta + F'_{mts}\sin m\theta\} \tag{6.83}$$

$$\begin{aligned}\frac{1}{b^2}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t\geq 2}}(r,\theta) &= \hat{r}\left\{-\frac{\rho^{(t-1)}}{(t-1)!}\right\}\{F'_{mtc}\cos m\theta + F'_{mts}\sin m\theta\} \\ &+ \hat{\theta}\left\{\frac{m}{t!}\frac{\rho^t}{(1+\rho)}\right\}\{F'_{mtc}\sin m\theta - F'_{mts}\cos m\theta\}\end{aligned} \tag{6.84}$$

$$\begin{aligned}&\frac{1}{b^2}b\overleftarrow{\nabla}_{\vec{r}}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t\geq 2}}(r,\theta) \\ &= \hat{r}\hat{r}\left\{-\frac{\rho^{(t-2)}}{(t-2)!}\right\}\{F'_{mtc}\cos m\theta + F'_{mts}\sin m\theta\} \\ &+ \hat{\theta}\hat{\theta}\left\{-\frac{1}{t!}\frac{\rho^{(t-1)}}{(1+\rho)^2}(t-(m^2-t)\rho)\right\}\{F'_{mtc}\cos m\theta + F'_{mts}\sin m\theta\} \\ &+ (\hat{r}\hat{\theta}+\hat{\theta}\hat{r})\left\{\frac{m}{t!}\frac{\rho^{(t-1)}}{(1+\rho)^2}(t+(t-1)\rho)\right\}\{F'_{mtc}\sin m\theta - F'_{mts}\cos m\theta\}\end{aligned} \tag{6.85}$$

Note that image displacements caused by these terms are small for small expansion parameter ρ.

As for the $t = 0$ and $t = 1$ terms, the angle $\psi_{F_{mt}}$ of the multipole moment can be

related to the location of the mass excess that is causing such a multipole moment. In particular, for $t \leq m$, a mass external to the ring radius, at any of the angles $\psi_{F_{mt}} + n\frac{2\pi}{m}$, for $n = 0, \ldots, m-1$, could cause the multipole moment. For $t > m$, a mass excess external to the ring does not affect this multipole moment. Alternately, for all m and t, a mass excess internal to the ring at angle $\psi_{\underset{m}{\text{diff}}} + n\frac{2\pi}{m}$ for t even, or at $\psi_{\underset{m}{\text{diff}}} + (n + \frac{1}{2})\frac{2\pi}{m}$ for t odd, could drive such a multipole moment.

6.3 Modifications to the Taylor-expanded multipole model

For the $m \geq 1$ multipole components, the first two terms, $t = 0$ and $t = 1$ in the Taylor expansion constrain, respectively, the sum $\vec{M}_{\underset{m}{\text{sum}}}$ and the difference $\vec{M}_{\underset{m}{\text{diff}}}$ of effects due to mass perturbations exterior to and interior to the Einstein ring radius. If the lens models are well enough constrained that both $\vec{M}_{\underset{m}{\text{sum}}}$ and $\vec{M}_{\underset{m}{\text{diff}}}$ are well constrained, the physical meaning of the model becomes clearer if $\vec{M}_{\underset{m}{\text{sum}}}$ and $\vec{M}_{\underset{m}{\text{diff}}}$ are replaced as model parameters by directly using $\vec{A}_m = (\vec{M}_{\underset{m}{\text{sum}}} + \vec{M}_{\underset{m}{\text{diff}}})/2$ and $\vec{B}_m = (\vec{M}_{\underset{m}{\text{sum}}} - \vec{M}_{\underset{m}{\text{diff}}})/2$. Recall that $\vec{A}_m$ (equation 6.42) and $\vec{B}_m$ (equation 6.43) are the m^{th} mass multipole moments of the mass exterior to the ring radius and interior to the ring, respectively.

The mass interior to the ring radius contributes to the m^{th} multipole moment to all orders t in the Taylor expansion. The mass exterior to the ring contributes to the m^{th} multipole moment only to orders $t \leq m$ in the Taylor expansion. See equations 6.39 and 6.40 for the contribution of $\vec{A}_m$ and $\vec{B}_m$ to $\vec{F}_{mt}$. The contribution of $\vec{A}_m$ and $\vec{B}_m$ to the potential can be calculated exactly, and only the effect of the mass near the ring radius be left to Taylor-expand.

When this is done, the potential may be written as:

$$\begin{aligned}\frac{1}{b^2}\Phi(r,\theta) = & \\ & \text{const} + \rho + \frac{1}{2}\rho^2 f_2 + \sum_{t=3}^{\infty}\frac{1}{t!}\rho^t f_t \\ & + \sum_{m=1}^{\infty}\left\{-\left(\frac{r}{b}\right)^m \vec{A}_m - \left(\frac{r}{b}\right)^{-m}\vec{B}_m + \sum_{t=2}^{\infty}\frac{1}{t!}\rho^t \vec{G}_{mt}\right\}\cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta)\end{aligned} \tag{6.86}$$

where the coefficients of the $m \geq 1$, $t \geq 2$ terms have been renamed $\vec{G}_{mt}$ since the effects of $\vec{A}_m$ and $\vec{B}_m$ have been separated out from these terms.

Again the effects of the lensing degeneracies must be taken into effect. A constant offset to the potential has no effect and may be set to zero. As before, the $\vec{B}_1$ may be set to zero to force the center of mass of the matter inside the ring radius to be at

the origin. The $\vec{A}_1$ term is the prismatic degeneracy, the dipole moment due to mass exterior to the Einstein ring; it does not affect the image fluxes or locations, and is therefore unconstrainable. It should therefore be set to zero for modeling. The mass sheet degeneracy again causes the following transformation from the correct model Φ to an alternate model Φ', leaving the image locations and fluxes unchanged:

$$\Phi(\vec{r}) \longrightarrow \Phi'(\vec{r}) = (1-\kappa)\Phi(\vec{r}) + \frac{1}{2}\kappa|\vec{r}|^2. \tag{6.87}$$

In order to use this model for lens modeling, one of the parameters involved in the degeneracy must be fixed; as before the coefficient of the ρ^2 term in Φ' is fixed to zero.

The form of the model suitable for lens modeling is then

$$\begin{aligned}\frac{1}{b^2}\Phi(r,\theta) &= \rho + \sum_{t=3}^{\infty}\frac{1}{t!}\rho^t f'_t \\ &\quad - \sum_{m=2}^{\infty}\left\{(1+\rho)^m\vec{A}'_m + (1+\rho)^{-m}\vec{B}'_m\right\}\cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta) \\ &\quad + \sum_{m=1}^{\infty}\sum_{t=2}^{\infty}\frac{1}{t!}\rho^t\vec{G}'_{mt}\cdot(\hat{x}\cos m\theta + \hat{y}\sin m\theta).\end{aligned} \tag{6.88}$$

The model parameters are related to the true potential Φ and mass distribution σ by:

$$f'_3 = 1 - b\frac{d}{dr}\ln\left(1 - \frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\Big|_{r=b} \tag{6.89}$$

$$f'_{t\geq 3} = (1-\kappa)\,\frac{1}{b^2}b^t\frac{d^t}{dr^t}\Phi_o(r)\Big|_{r=b} \tag{6.90}$$

and

$$\vec{A}'_m = (1-\kappa)\vec{A}_m \tag{6.91}$$

$$\vec{B}'_m = (1-\kappa)\vec{B}_m \tag{6.92}$$

$$\vec{G}'_{\substack{m\\t=2}} = \frac{\vec{\sigma}_m(b)}{\sigma_{\rm crit} - \sigma_o(b)} \tag{6.93}$$

$$\vec{G}'_{\substack{m \\ t=3}} = \frac{1}{\sigma_{\rm crit} - \sigma_o(b)} \left\{ b \frac{d}{dr} \vec{\sigma}_m(r) \Big|_{r=b} - \vec{\sigma}_m(b) \right\} \tag{6.94}$$

$$\vec{G}'_{\substack{m \geq 1 \\ t \geq 2}} = (1-\kappa) \frac{1}{b^2} b^t \frac{d^t}{dr^t} \vec{\Phi}_{\substack{m \\ \text{near ring}}}(r) \Big|_{r=b} \tag{6.95}$$

where

$$\begin{aligned} \vec{\Phi}_{\substack{m \\ \text{near ring}}}(r) = & -\frac{1}{m} \left(\frac{r}{b}\right)^m \int_r^b \left(\frac{r'}{b}\right)^{1-m} \frac{dr'}{b} \frac{\vec{\sigma}_m(r)}{\sigma_{\rm crit}} \\ & -\frac{1}{m} \left(\frac{r}{b}\right)^{-m} \int_b^r \left(\frac{r'}{b}\right)^{1+m} \frac{dr'}{b} \frac{\vec{\sigma}_m(r)}{\sigma_{\rm crit}} \end{aligned} \tag{6.96}$$

Formulas for the derivatives of the monopole terms, as needed in lens modeling, and a discussion of the significance of these terms is given in section 6.2.6. The $m \geq 1$ multipole terms are described below.

The external-mass piece of the $m \geq 1$ multipole term is:

$$\frac{1}{b^2} \Phi'_{\substack{m \geq 1 \\ \text{exterior}}}(r,\theta) = \{-(1+\rho)^m\} \{A'_{mc} \cos m\theta + A'_{ms} \sin m\theta\} \tag{6.97}$$

$$\begin{aligned} \frac{1}{b^2} b \vec{\nabla}_{\vec{r}} \Phi'_{\substack{m \geq 1 \\ \text{exterior}}}(r,\theta) = & \ \hat{r} \left\{-m(1+\rho)^{(m-1)}\right\} \{A'_{mc} \cos m\theta + A'_{ms} \sin m\theta\} \\ & + \hat{\theta} \left\{m(1+\rho)^{(m-1)}\right\} \{A'_{mc} \sin m\theta - A'_{ms} \cos m\theta\} \end{aligned} \tag{6.98}$$

$$\begin{aligned} & \frac{1}{b^2} b \overleftarrow{\nabla}_{\vec{r}} b \vec{\nabla}_{\vec{r}} \Phi'_{\substack{m \geq 1 \\ \text{exterior}}}(r,\theta) \\ & = \hat{r}\hat{r} \left\{-m(m-1)(1+\rho)^{(m-2)}\right\} \{A'_{mc} \cos m\theta + A'_{ms} \sin m\theta\} \\ & + \hat{\theta}\hat{\theta} \left\{m(m-1)(1+\rho)^{(m-2)}\right\} \{A'_{mc} \cos m\theta + A'_{ms} \sin m\theta\} \\ & + (\hat{r}\hat{\theta} + \hat{\theta}\hat{r}) \left\{m(m-1)(1+\rho)^{(m-2)}\right\} \{A'_{mc} \sin m\theta - A'_{ms} \cos m\theta\} \end{aligned} \tag{6.99}$$

The $m = 1$ term is set to zero to avoid the prismatic degeneracy. An $\vec{A}_m$ term can be generated by a mass excess, exterior to the Einstein ring radius, in any of the m equally spaced directions $\psi_{A_m} + n\frac{2\pi}{m}$, for $n = 0, \ldots, m-1$. The displacement of the image away from the source, $\vec{\nabla}\Phi$, is generally a displacement away from these

<table>
<tr><td>$m = 0$
monopole</td><td></td><td>$m = 1$
dipole</td><td>$m \geq 2$
quadrupole,
octupole, ...</td></tr>
<tr><td>Degenerate
Constant offset to potential</td><td>$t = 0$
$\Longleftarrow$

Exterior
$\Longrightarrow$</td><td>Degenerate
Prismatic degeneracy. Dipole moment of the mass exterior to the ring is unconstrainable.</td><td>Multipole-Exterior
Sensitive to the multipole moment of the mass exterior to the Einstein ring.</td></tr>
<tr><td>Monopole-Linear
Sets Einstein ring radius. Sensitive to total mass within ring.</td><td>$t = 1$
$\Longleftarrow$

Interior
$\Longrightarrow$</td><td>Degenerate
Fixes center of mass of the mass interior to ring to be at the origin. (Location of origin is then allowed to vary.)</td><td>Multipole-Interior
Sensitive to the multipole moment of the mass interior to the Einstein ring.</td></tr>
<tr><td>Degenerate
Mass sheet degeneracy. Surface mass density at ring unconstrainable</td><td>$t = 2$</td><td colspan="2" rowspan="2">modified Multipole-Taylor
Depends on the behavior of the multipole moment of the matter near the ring radius: For $t = 2$ gives m^{th} multipole moment of mass in a narrow annulus at the ring radius.

For $t \geq 3$ has more complicated dependence on m^{th} multipole moment of mass in a narrow annulus at the ring radius.</td></tr>
<tr><td>Monopole-Taylor
For $t = 3$ gives the fall-off of surface mass density with radius, evaluated at the ring radius.</td><td>$t \geq 3$</td></tr>
</table>

Table 6.4: Physical significance of terms in the modified Multipole-Taylor expansion. Certain terms are unconstrainable due to degeneracies as explained in the text. Five functional forms are needed to represent all the other terms in the expansion. These five forms are all that need be coded into a lens modeling program. They can be used with m and t fixed to the appropriate values.

perturber locations. The tangential displacements are displacements away from the perturbing mass locations. The radial displacements are inwards, away from the perturber locations, for images within $\pm\frac{1}{4}\left(\frac{2\pi}{m}\right)$ of one of the possible perturber locations. The radial displacements are outwards otherwise.

Looking ahead to section 6.5.3 of this thesis, a point-mass perturber, further from the origin than the furthest lensed image, would contribute to each external multipole:

$$2\left(1-\frac{\sigma_o(b)}{\sigma_{\rm crit}}\right)|\vec{A}'_m| = |\vec{A}_m| = \frac{1}{m}\left(\frac{b}{R}\right)^m\left(\frac{b_{\rm PM}}{b}\right)^2 \tag{6.100}$$

(see equation 6.120) where b is the Einstein ring radius of the primary deflector, R is the distance from the origin (of the primary deflector) to the perturber, and $b_{\rm PM}$ is the Einstein ring radius of the perturber. Note that the effect decreases for increasing m. Constraints on two external multipole moments, $m \neq m'$, caused by an external perturber, would allow the determination of both the distance to the perturber and the Einstein ring radius of the perturber.

The internal-mass piece of the $m \geq 1$ multipole term is:

$$\frac{1}{b^2}\Phi'_{\substack{m\geq 1\\ \text{interior}}}(r,\theta) = \left\{-\frac{1}{(1+\rho)^m}\right\}\{B'_{mc}\cos m\theta + B'_{ms}\sin m\theta\} \tag{6.101}$$

$$\begin{aligned}\frac{1}{b^2}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ \text{interior}}}(r,\theta) &= \hat{r}\left\{\frac{m}{(1+\rho)^{(m+1)}}\right\}\{B'_{mc}\cos m\theta + B'_{ms}\sin m\theta\} \\ &+ \hat{\theta}\left\{\frac{m}{(1+\rho)^{(m+1)}}\right\}\{B'_{mc}\sin m\theta - B'_{ms}\cos m\theta\}\end{aligned} \tag{6.102}$$

$$\begin{aligned}&\frac{1}{b^2}b\overleftarrow{\nabla}_{\vec{r}}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ \text{interior}}}(r,\theta) \\ &= \hat{r}\hat{r}\left\{-\frac{m(m+1)}{(1+\rho)^{(m+2)}}\right\}\{B'_{mc}\cos m\theta + B'_{ms}\sin m\theta\} \\ &+ \hat{\theta}\hat{\theta}\left\{\frac{m(m+1)}{(1+\rho)^{(m+2)}}\right\}\{B'_{mc}\cos m\theta + B'_{ms}\sin m\theta\} \\ &+ (\hat{r}\hat{\theta}+\hat{\theta}\hat{r})\left\{-\frac{m(m+1)}{(1+\rho)^{(m+2)}}\right\}\{B'_{mc}\sin m\theta - B'_{ms}\cos m\theta\}\end{aligned} \tag{6.103}$$

The $m = 1$ term is set to zero to force the center of mass of the mass interior to the ring to be at the origin — this origin, the deflector location, is parameterized by model parameters (g_x, g_y). A $\vec{B}_m$ term can be generated by a mass excess, interior to the Einstein ring radius, in any of the m equally spaced directions $\psi_{B_m} + n\frac{2\pi}{m}$, for $n = 0, \ldots, m-1$. As with the $\vec{A}_m$ term, the displacement of the image away from the source, $\vec{\nabla}\Phi$, is generally a displacement away from these perturber locations. The tangential displacements are displacements away from the perturbing mass locations. The radial displacements are outwards, away from the perturber locations, for images within $\pm\frac{1}{4}\left(\frac{2\pi}{m}\right)$ of one of the possible perturber locations. The radial displacements are inwards otherwise.

The terms in the Taylor expansion of the remainder of each multipole component are:

$$\frac{1}{b^2}\Phi'_{\substack{m\geq 1\\ t\geq 2}}(r,\theta) = \left\{\frac{\rho^t}{t!}\right\}\{G'_{mtc}\cos m\theta + G'_{mts}\sin m\theta\} \tag{6.104}$$

$$\begin{aligned}\frac{1}{b^2}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t\geq 2}}(r,\theta) &= \hat{r}\left\{\frac{\rho^{(t-1)}}{(t-1)!}\right\}\{G'_{mtc}\cos m\theta + G'_{mts}\sin m\theta\} \\ &+ \hat{\theta}\left\{-\frac{m}{t!}\frac{\rho^t}{(1+\rho)}\right\}\{G'_{mtc}\sin m\theta - G'_{mts}\cos m\theta\}\end{aligned} \tag{6.105}$$

$$\begin{aligned}&\frac{1}{b^2}b\overleftarrow{\nabla}_{\vec{r}}b\vec{\nabla}_{\vec{r}}\Phi'_{\substack{m\geq 1\\ t\geq 2}}(r,\theta) \\ &= \hat{r}\hat{r}\left\{\frac{\rho^{(t-2)}}{(t-2)!}\right\}\{G'_{mtc}\cos m\theta + G'_{mts}\sin m\theta\} \\ &+ \hat{\theta}\hat{\theta}\left\{\frac{1}{t!}\frac{\rho^{(t-1)}}{(1+\rho)^2}(t-(m^2-t)\rho)\right\}\{G'_{mtc}\cos m\theta + G'_{mts}\sin m\theta\} \\ &+ (\hat{r}\hat{\theta}+\hat{\theta}\hat{r})\left\{-\frac{m}{t!}\frac{\rho^{(t-1)}}{(1+\rho)^2}(t+(t-1)\rho)\right\}\{G'_{mtc}\sin m\theta - G'_{mts}\cos m\theta\}\end{aligned} \tag{6.106}$$

The $\vec{G}'_{mt}$ terms depend on the behavior of the surface mass distribution in the narrow radial range within which the lensed images are located. The dominant term, $\vec{G}'_{\substack{m\\ t=2}}$, can be generated by an excess of mass, at the ring radius, located in any of the m equally spaced directions $\psi_{G_{mt}} + n\frac{2\pi}{m}$, for $n = 0, \ldots, m-1$, since the coefficient $\vec{G}'_{\substack{m\\ t=2}}$ is

directly related to the corresponding multipole moment of the surface mass density,

$$\vec{G}'_{\substack{m \\ t=2}} = \frac{\vec{\sigma}_m(b)}{\sigma_{\rm crit} - \sigma_o(b)}. \tag{6.107}$$

Non-negativity of the surface mass density limits the amplitude of $\vec{\sigma}_m(b)$ (see equation 6.26), which limits the amplitude of $\vec{G}'_{\substack{m \\ t=2}}$ to

$$|\vec{G}'_{\substack{m \\ t=2}}| \leq \frac{2\sigma_o(b)}{\sigma_{\rm crit} - \sigma_o(b)}. \tag{6.108}$$

This limit ranges from zero for $\sigma_o(b) = 0$ (no mass at the Einstein ring radius) to $+\infty$ for $\sigma_o(b) = \sigma_{\rm crit}$. For a mass distribution with a singular isothermal sphere monopole, the limit is

$$|\vec{G}'_{\substack{m \\ t=2}}|_{\rm SIS} \leq 2. \tag{6.109}$$

For mass distributions more centrally concentrated than a singular isothermal sphere, the limit on $\vec{G}'_{\substack{m \\ t=2}}$ would be lower. If some fraction p of the mass at the Einstein ring radius is located in point-like (single-angle) perturbers at angles $\psi_{G_{m,t=2}} + n\frac{2\pi}{m}$, with the remainder evenly distributed around the ring, then this fraction is

$$p = \frac{|\vec{G}'_{\substack{m \\ t=2}}|}{\left(\frac{2\sigma_o(b)}{\sigma_{\rm crit} - \sigma_o(b)}\right)}. \tag{6.110}$$

If the perturbers were extended in size, then a larger fraction of the mass in the ring would need to be involved in the perturbers to give the same value of $|\vec{G}'_{\substack{m \\ t=2}}|$.

The radial and tangential image displacements caused by this $m \geq 1$, $t \geq 2$ term are small for ρ small. The $t = 2$ term may cause significant radial magnification, even for ρ small, but all the higher terms contribute little to the magnification for small ρ.

6.4 An elliptical variant on the Taylor-expanded multipole model?

Since the Einstein "ring" is approximately an ellipse for lens systems, it would seem sensible to use elliptical coordinates. One could separate the variables in elliptical coordinates to get an analog of the multipole expansion — and then Taylor expand in offsets from the ellipse location. However in elliptical coordinates, the various unconstrainable degeneracies do not coincide nicely with individual terms in the expansion. In cylindrical coordinates as used above, the various unconstrainable degeneracies do coincide nicely with individual terms in the expansion, allowing these degeneracies to be avoided in the modeling.

6.5 Comparisons with other previously-used models

Previous efforts at modeling the gravitational lens MG J0414+0534 have used simple models with few parameters, due to the limited number of modeling constraints available, or due to the computational cost of running LensClean. In generally these models have used physically motivated radial forms, either with an elliptical mass distribution or potential, or with such an elliptical mass distribution or potential truncated at the quadrupole moment.

These potentials may be related to the multipole-Taylor model by showing what the coefficients are in the multipole-Taylor model expansion of each potential. This is useful for relating coefficients of the multipole-Taylor model to physical parameters of the potential. It is also useful for understanding which differences between the models are relevant for lens modeling of rings and quads — and which are not.

A typical deflector model is a 2-D lens potential consisting of a simple monopole term added to a simple quadrupole term. Such models were used by, for example, Kochanek (1991), Witt *et al.* (1995), Ellithorpe (1995), Falco *et al.* (1997), and

Schechter *et al.* (1997). Ellithorpe (1995) also used a surface mass density model which was truncated at the quadrupole term of the surface mass density. However, since each multipole moment of the potential depends only on the corresponding multipole moment of the surface mass density, this is equivalent to truncating the potential at the quadrupole moment.

6.5.1 Monopole terms

Table 6.5 lists several monopole profiles: the singular isothermal sphere, the point mass, the power-law profile, and the power-law profile with core radius. Other physically motivated radial profiles have been used, such as that corresponding to a de Vaucouleurs surface mass density (Ellithorpe 1995), but unless one can constrain the f_3' parameter in the Taylor expansion of the monopole term, little will be gained by using such profiles in modeling ring and quad lenses.

The monopole terms listed in the table are all parameterized with ring radius b. Note that they have differing values of the surface mass density at the ring radius and of the f_3' parameter which is related to the decrease in surface mass density with radius. The surface mass density at the ring radius is unconstrainable due to the mass sheet degeneracy, so for lens modeling purposes, the first significant difference between the radial profiles is the f_3' term. When using a model to fit a potential whose actual surface mass density at the ring radius is $\sigma_o(b)$, the mismatch between $\sigma_{o_{model}}(b)$ and $\sigma_o(b)$ causes a scaling of all the higher multipole parameters and of the time delay. The rescaling factor $(1-\kappa)$ is also tabulated in the table.

Note that this f_3' parameter, if constrainable, will nicely distinguish the continuum of models from a point mass, to a singular isothermal sphere, all the way on out to a mass sheet, as the power law index goes from zero to 1/2 to 1. Note that a core radius or a sudden drop in the surface mass density with radius can make f_3' even more negative that -1.

If one can constrain the f_3' term, one has a chance of learning about the radial profile of the deflector potential, and of predicting the f_2' term that enters into the $(1-\kappa)$ scaling factor, instead of merely fixing $f_2' = 0$. But if one cannot constrain

Model	Potential	$\frac{\sigma_o(b)}{\sigma_{\rm crit}}$	f_3'	$1-\kappa=\frac{1}{2\left(1-\frac{\sigma_o(b)}{\sigma_{\rm crit}}\right)}$
point mass	$b^2 \ln r$	0	1	1/2
singular isothermal sphere	br	1/2	0	1
mass sheet with critical surface mass density	$\frac{1}{2}r^2$	1	-1	∞
power law $\alpha \to 0$, point mass $\alpha = 1/2$, isothermal $\alpha \to 1$, mass sheet	$\frac{b^2}{2\alpha}\left(\frac{r}{b}\right)^{2\alpha}$	α	$1-2\alpha$	$\frac{1}{2(1-\alpha)}$
power law with core radius	$\frac{s^2+b^2}{2\alpha}\left(\frac{s^2+r^2}{s^2+b^2}\right)^{\alpha}$	$\frac{s^2+\alpha b^2}{s^2+b^2}$	$\frac{(1-2\alpha)b^2-3s^2}{s^2+b^2}$	$\frac{1}{2(1-\alpha)}\frac{s^2+b^2}{b^2}$

Table 6.5: Several physically motivated monopole profiles. The first column is the form of the 2-D lens potential. The second column is the surface mass density at the ring radius as a fraction of the critical surface mass density. The multipole-Taylor model parameter $f_3' = 1 - b\frac{d}{dr}\ln\left(1-\frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\Big|_{r=b}$ is the first parameter in that model sensitive to radial profile. Suppose that the deflector of a gravitational lens truly has the monopole potential tabulated in the first column. Suppose that it is fitted with a multipole-Taylor model, with parameter f_2' fixed to zero to fix the mass sheet degeneracy, as described in sections 6.2 and 6.3. The third column lists the fitted value that f_3' would have in that case. In such a case, the fitted multipole-Taylor model predicts a time delay that differs from the actual time delay by a factor $(1-\kappa)$ that arises due to the mass sheet degeneracy. For example, if the deflector truly is a point mass (black hole) and we fit a multipole-Taylor model, then the model predicted time delay will be shorter than the true time delay by a factor of 1/2.

the f_3' term, then there is little point in using more complicated monopole profiles for ring and quad lens systems.

6.5.2 Quadrupole terms

There are several quadrupole terms commonly used to accompany the monopole term.

External quadrupole This is a quadrupole moment due only to mass lying completely outside the Einstein ring radius. In our notation the potential is

$$\Phi_{\rm XQ}(r,\theta) = -r^2 A_2 \cos 2(\theta - \psi_{A_2}) \tag{6.111}$$

where ψ_{A_2} and $\psi_{A_2}+\pi$ are the possible directions to the perturbing mass causing the quadrupole moment. This is sometimes also called an "external shear" and parameterized

$$\Phi_{\rm XS}(r,\theta) = \frac{1}{2}\gamma r^2 \cos 2(\theta - \theta_\gamma) \tag{6.112}$$

where the convention for the angle θ_γ differs by $\pi/2$ radians from that for ψ_{A_2}.

Internal quadrupole This is a quadrupole moment due only to mass lying completely inside the Einstein ring radius. In our notation the potential is

$$\Phi_{\rm NQ}(r,\theta) = -\frac{b^4}{r^2} B_2 \cos 2(\theta - \psi_{B_2}) \tag{6.113}$$

where ψ_{B_2} and $\psi_{B_2}+\pi$ are the possible directions to the mass excess driving this quadrupole moment.

Mixed quadrupole This is a quadrupole due to mass both outside and inside the ring radius. This particular radial dependence is used, because it is obtained by truncating a singular isothermal elliptical potential at the quadrupole term in the multipole expansion.

$$\Phi_{\rm MQ}(r,\theta) = -\frac{1}{2} br\epsilon \cos 2(\theta - \theta_\epsilon) \tag{6.114}$$

Again, one must be careful about factors of two in the definition of ϵ and offsets of $\pi/2$ in the definition of θ_ϵ. Here, θ_ϵ and $\theta_\epsilon+\pi$ are the directions of higher surface mass density.

Recall that the size of the quadrupole moment determines the tangential image displacements. The balance between the internal and external quadrupole moments determines the radial displacement that accompanies the tangential displacement. By using only a single quadrupole piece, with a fixed ratio between internal and external origin for the quadrupole, one fixes the radial displacement relative to the tangential displacement. The value to which it is fixed may not match that of the true potential.

6.5.3 Other potentials

There are a few other potentials that deserve mention.

Singular isothermal elliptical potential

The following elliptical potential was used by Hewitt *et al.* (1992) and Falco *et al.* (1997):

$$\Phi(r,\theta) = b' r \sqrt{1 - \epsilon \cos 2(\theta - \theta_\epsilon)} \tag{6.115}$$

where, for ϵ small, ϵ is related to the axial ratio of the isopotential contours by $\epsilon = 1 - \frac{b_\Phi}{a_\Phi}$ where a_Φ and b_Φ are respectively the major and minor semi-axes of an isopotential contour. Equivalently, for ϵ small, ϵ is related to the axial ratio of the isodensity contours (of the surface mass density) by $3\epsilon = 1 - \frac{b_\sigma}{a_\sigma}$ where a_σ and b_σ are respectively the major and minor semi-axes of an isodensity contour. Note that by symmetry this potential has no odd m multiple components.

For small ellipticities the potential is:

$$\Phi(r,\theta) = br - \frac{1}{2} br\epsilon \cos 2(\theta - \theta_\epsilon) - \frac{1}{16} br\epsilon^2 \cos 4(\theta - \theta_\epsilon) + br\mathcal{O}(\epsilon^3) \tag{6.116}$$

where $b = (1 - \epsilon^2/16 + \mathcal{O}(\epsilon^3))\, b'$. This is a singular isothermal sphere monopole term, plus a "mixed" quadrupole term, with small amounts of a "mixed" $m = 4$ multipole term.

If a deflector potential with such a shape were fitted with our modified Taylor-expanded multipole model, then the fitted values of the amplitudes of the $m = 2$ external and internal multipole moments would be $|\vec{A}'_2| = \frac{3}{8}\epsilon$ and $|\vec{B}'_2| = \frac{1}{8}\epsilon$, so that the ratio of external to internal quadrupole would be $|\vec{A}'_2|/|\vec{B}'_2| = 3$.

Elliptical potential with softness α and core radius s

This potential was suggested by Blandford & Kochanek (1987).

$$\Phi(x,y) = \frac{\text{const}}{\alpha} \left\{ \left[1 + (1-\epsilon) \left(\frac{x}{s}\right)^2 + (1+\epsilon) \left(\frac{y}{s}\right)^2 \right]^\alpha - 1 \right\} \tag{6.117}$$

Keeping only terms to $\mathcal{O}(\epsilon)$, this gives a monopole piece that is the same as the power law with core radius monopole (which is tabulated in table 6.5) and a quadrupole term linear in ϵ with a radial dependence specified by α and s.

Perturbing point mass displaced from the center of the potential

Another approach is to use a multipole expansion for the principal deflector, but to use a point mass or a singular isothermal sphere located some distance away, to account for a secondary deflector object. If there is such an object, and it is not accounted for separately, how does it appear in the multipole-Taylor expansion?

For a point mass M located at a displacement $R(\hat{x}\cos\psi_{PM} + \hat{y}\sin\psi_{PM})$ away from the origin, the surface mass density is

$$\sigma_{PM}(\vec{r}) = \frac{M}{D_L^2 R}\delta(r-R)\delta(\theta-\psi_{PM}), \tag{6.118}$$

where D_L, the angular diameter distance to the deflector, enters the problem because x, y, r, and R are angular distances on the sky. The corresponding 2-D lens potential can be written as a sum over its multipole components

$$\Phi_{PM}(\vec{r}) = \text{const} - \sum_{m=1}^{\infty}\frac{1}{m}\left(\frac{r}{R}\right)^m b_{\rm PM}^2\cos m(\theta-\psi_{PM}), \tag{6.119}$$

where this expression is valid for $|\vec{r}| < R$, and where $b_{\rm PM}^2 = \frac{M}{\pi D_L^2 \sigma_{\rm crit}}$ is the squared Einstein ring radius of the perturber. This expression is adequate as long as this perturbing point mass is further from the origin than any of the lens images.

This potential consists of external multipole moments to all orders m in the multipole expansion. (Compare with equation 6.86.) The $m = 1$ term has no effect on the image positions, due to the prismatic degeneracy. The dominant term is the external quadrupole, with the effects of the higher-order terms being smaller by powers of r/R. For ring and quad lenses, the images are located near the ring radius, $r \approx b$. Therefore for a perturbing mass much further away than the ring radius (b/R small), it should be adequate to represent the effect of the mass by just the first few terms

in the multipole expansion, which have amplitudes (in the notation of section 6.3)

$$A_m = \frac{1}{m}\left(\frac{b}{R}\right)^m \left(\frac{b_{\rm PM}}{b}\right)^2, \tag{6.120}$$

where b is the ring radius of the principal deflector. For a nearby perturber ($b/R \approx 1$), such a truncation of the multipole expansion will not be adequate. Therefore if there is a known nearby perturber, it may be worth explicitly adding to the potential model a point mass at the perturber's location.

Slightly extended perturbing point mass displaced from the center of the potential

How do radial and tangential extent in an external perturber affect the external multipole amplitudes due to the perturber? This can be found by considering a perturber with a center of mass at a radial distance R from the center of the principal deflector, and with surface mass density uniform over a region of radial extent ΔR and tangential extent $R\Delta\theta$. Such a perturber contributes to the external multipole moments of the potential to all orders m in the multipole expansion, with amplitudes

$$A_m = \frac{1}{m}\left(\frac{b}{R}\right)^m \left(\frac{b_{\rm perturber}}{b}\right)^2 \left\{ 1 + \frac{m(m+1)}{24}\left(\frac{\Delta R}{R}\right)^2 - \frac{m(m+1)}{24}(\Delta\theta)^2 + \mathcal{O}\left(\frac{\Delta R}{R}\right)^4 + \mathcal{O}(\Delta\theta)^4 + \mathcal{O}(\Delta\theta)^2\,\mathcal{O}\left(\frac{\Delta R}{R}\right)^2 \right\}, \tag{6.121}$$

Extension in angle causes the multipole amplitudes to fall more rapidly with m; angular extent to a perturber washes out the higher m multipole moments more than the lower m multipole moments, since the surface mass density is being integrated against $\sin m\theta$. In contrast, radial extent increases the multipole amplitudes, with a larger effect for higher m, since these require an integration in radius against a weight $r^{(1-m)}$. The net effect for perturbers with similar radial and tangential extents, $\Delta R = R\Delta\theta$, is that there is no effect through $\mathcal{O}\left(\frac{\Delta R}{R}\right)^3$. Therefore, using the point mass perturber formula (equation 6.120) to estimate the distance to and the Einstein

ring radius of a perturber with small extent, $\Delta R \ll R$, should cause no problem. However this formula will not be valid for significantly extended perturbers, $\Delta R \approx R$, for which the detailed shape of the perturber must be taken into account.

6.6 Effects of the lensing degeneracies on the model predicted time delay

In whatever manner one chooses to make a parameterized model of the potential, the mass sheet degeneracy makes it impossible to constrain the deflector's surface mass density at the Einstein ring radius, $\sigma_o(b)$, from the image locations and fluxes. The value of $\sigma_o(b)$ can be deduced only by making model-dependent assumptions of how it depends on other features of the deflector. These assumptions may be hidden in the choice of the deflector model. For example, choosing a point mass monopole term, makes the assumption that $\sigma_o(b) = 0$. Choosing a singular isothermal monopole term, makes the assumption that $\sigma_o(b) = \sigma_{\rm crit}/2$. With slightly more sophistication, one can attempt to deduce $\sigma_o(b)$ in a model-dependent way from the fall-off of surface mass density with radius, $1 - b\frac{d}{dr}\ln\left(1 - \frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\Big|_{r=b}$. However this latter is expected to be difficult to constrain for quad and ring lenses.

The mismatch between one's model-dependent assumptions and the true value of $\sigma_o(b) = 0$ gives a factor $1 - \kappa$, that enters into the model-predicted time delay:

$$t_{\rm model}(\vec{r}_{l\alpha}) - t_{\rm model}(\vec{r}_{l\beta}) = (1 - \kappa)\left(t_{\rm true}(\vec{r}_{l\alpha}) - t_{\rm true}(\vec{r}_{l\beta})\right). \qquad (6.122)$$

For example, assuming a singular isothermal sphere profile, when the deflector truly is a point mass, results in underestimating the time delay by a factor of two, $1 - \kappa = 1/2$. Assuming a singular isothermal sphere profile when the deflector is much more extended than an isothermal sphere can result in overestimating the time delay by a large factor. This scaling factor $1 - \kappa$ is tabulated in tables 6.2 and 6.5 for various cases. The error in the time delay prediction is a systematic error based on modeling assumptions. It is therefore difficult to quantify.

Chapter 7

Modeling in practise: modeling the VLBA maps of MG J0414+0534

The deflector models developed in chapter 6 were used to model the VLBA observation of MG J0414+0534 described in chapter 3 (figure 3-6) using the modeling techniques of chapter 5. This chapter describes how the modeling was done and what the modeling results were.

7.1 Implementation of the modeling

A program was written to find the best-fit model parameters by minimization of the chi-squared, and to calculate confidence intervals and time delays for the models. The significant features of this implementation of lens modeling are described below.

7.1.1 The chi-squared

Only the image positions, not the image fluxes were used as modeling constraints. For speed of computation, the "source-plane" approximation was used, wherein the observation positions and errors were mapped back to the source plane. The form of the chi-squared used was that of equation 5.13.

For model parameters that minimized the "source-plane" approximation to the

chi-squared, the actual "image-plane" chi-squared was also calculated, using equation 5.6, as a check on the approximation. For the models that were adequate (see below) the "source-plane" chi-squared at its minimum differed from the true "image-plane" chi-squared by no more than 1.5 percent, and typically only by 0.2-0.6%. Since this is much less than the chi-squared increment used in finding the confidence limits on the model parameters, this approximation did not introduce any appreciable error — and saved significant computation time.

The reasons that only positions and not fluxes were used were as follows. Each MG J0414+0534 image, A1, A2, B, or C, consists of four components, "a", "b", "c", and "d", which are somewhat non-collinear. Three non-collinear components would be sufficient to determine the magnification matrix at each image, A1, A2, B, or C, provided that the magnification matrix was constant over the region of the subcomponents "a", "b", "c", and "d". Using the fluxes of each subcomponent merely adds the information about how the determinant of the magnification matrix changes between "a", "b", "c", and "d".

However for MG J0414+0534 there are significant discrepancies in the flux ratios at different wavelengths. It may be that these are due to microlensing — which is expected to be a more significant effect in the optical than in the radio as the optical flux is assumed to come from a smaller region than the radio flux. Alternately, the flux ratio discrepancy may be explained by obscuration by dust in one part of the deflector. One may want to assume that such microlensing and obscuration effects are unimportant at radio wavelengths. Alternately, one may simply dispense with the fluxes and use only the positions for lens modeling.

Given the MG J0414+0534 flux ratio discrepancy, given that the positions are much better constrained than the fluxes for MG J0414+0534, and given the secondary nature of the information that the fluxes provide in modeling this source, it was decided to use only the MG J0414+0534 positions and not the fluxes for the lens modeling.

7.1.2 The implementation of the models

In writing the lens modeling program, a modular approach to deflector potential models was taken. The basic building blocks which were coded into the program could be added together to form a complete potential model simply by listing them in the input file. New potential models (or building blocks) could easily be implemented as long as functions could be written to calculate the potential itself $\Phi(\vec{r})$, its gradient $\vec{\nabla}\Phi(\vec{r})$, and its second derivative matrix $\partial_i\partial_j\Phi(\vec{r})$, as functions of the position $\vec{r}$ in the image plane. (The third derivative tensor $\partial_i\partial_j\partial_k\Phi(\vec{r})$ would also be needed if one were making the corrections for slightly extended sources in the presence of magnification gradients, equations 5.44 and 5.45.)

The basic building blocks that were implemented include several simple models: the singular isothermal sphere "SIS", the point mass "PM", the external shear "XS", and a singular isothermal elliptical potential truncated at the quadrupole moment "SIEP" which is actually a singular isothermal sphere plus a mixed quadrupole. These basic models were described in section 6.5.

The building blocks of the Multipole-Taylor model were also coded. Only five functional forms were needed for this. They are tabulated in table 6.3, and the expressions for $\Phi(\vec{r})$, $\vec{\nabla}\Phi(\vec{r})$, and $\partial_i\partial_j\Phi(\vec{r})$ are given in section 6.2.6.

The building blocks of the modified Multipole-Taylor model, as tabulated in table 6.4, were also coded in. The two monopole forms are the same as for the Multipole-Taylor model. The expressions for $\Phi(\vec{r})$, $\vec{\nabla}\Phi(\vec{r})$, and $\partial_i\partial_j\Phi(\vec{r})$ for the three $m \geq 1$ multipole forms are given in section 6.3.

Necessary for this implementation was the ability to link model parameters between the model building blocks. For example, consider a modified-Multipole-Taylor model of several terms including the linear part of the monopole and the internal part of the quadrupole. Each term has as parameters the ring radius b and the deflector location (g_x, g_y), yet this composite model must have only one ring radius b and one deflector location (g_x, g_y). The chi-squared minimization algorithm allows these parameters to be linked and treated as one. Actually, the implementation allows

arbitrary parameters to be linked: one could link the orientation angle of the internal part of the quadrupole with that of the external part of the quadrupole if so desired.

The lens modeling program also allows parameters to be held fixed during the chi-squared minimization. For example, one might want to fix the deflector location to be that of the optically observed galaxy.

Several of the models have vector or vector-like components: the XS, the SIEP, and the $m \geq 1$ multipole terms of the Multipole-Taylor and modified-Multipole-Taylor models. For example the general $m \geq 1$, $t \geq 2$ term in the modified Multipole-Taylor model has a parameter $\vec{G}'_{mt}$. This can be parameterized by its x- and y-components, $\vec{G}'_{mt} = \hat{x}G'_{mtc} + \hat{y}G'_{mts}$, or by its orientation angle and amplitude: $\vec{G}'_{mt} = G'_{mt}(\hat{x}\cos m\psi_{G_{mt}} + \hat{y}\sin m\psi_{G_{mt}})$. Two forms have been implemented for each model of this type, one using as parameters the cosine and sine components G'_{mtc} and G'_{mts}, the other using the amplitude G'_{mt} and the angle $m\psi_{G_{mt}}$. (Note the factor of m.) For purposes of visualizing the significance of the model parameters, the amplitude-angle form, and error estimates on the amplitude and angle parameters, are more illuminating. For ease in minimization, the Cartesian form is better, since in this form the chi-squared becomes approximately quadratic in the model parameters when the parameter values are far from the chi-squared minimum. Therefore for those vector-like model parameters, when neither the amplitude nor the angle was fixed or linked to another parameter, the modeling program converted the model to its Cartesian form before minimization and converted back to its original form before displaying the output.

7.1.3 The minimization algorithms

The best-fit model is determined by finding the model parameters that minimize the chi-squared.

The model parameters are of two sorts: the model source positions and the model deflector parameters. For fixed values of the model deflector parameters, the minimization over the model source positions can be done analytically. The expression for the model source positions that minimize the chi-squared (with the model deflector

parameters held fixed) is given in equation 5.15. For each set of deflector parameters examined, the minimization over the source parameters was done analytically. Therefore the dimensionality of the parameter space to be explored by a general-purpose minimization algorithm was reduced from $(2 \times N_l) + N_a$, to merely N_a, where N_a is the number of deflector parameters and N_l is the number of model sources. For MG J0414+0534 where $N_l = 4$ this reduces by 8 dimensions the dimensionality of the parameter space. This is a significant aid to any minimization algorithm used to minimize the chi-squared.

Three different algorithms were used for finding the minima of the models' chi-squareds within the deflector-parameter space: the Powell direction set method which does line minimizations successively in various directions (Press *et al.* 1992), the downhill simplex method which is also known as the "amoeba" (Press *et al.* 1992), and a variant of simulated annealing combining a downhill simplex with some probability of taking uphill steps (Press *et al.* 1992). Before discussing the relative merits of these algorithms for lens modeling, the problems encountered in chi-squared minimization will be briefly discussed.

The search for the global minimum of a general function in a multi-dimensional parameter space is plagued by two problems. Firstly the algorithm may find a local minimum and stop. There may be a lower minimum some distance away in parameter space, but if the algorithm has not sampled that portion of parameter space, and has found a minimum, the true global minimum may go undetected.

Secondly, the algorithm may stray into a region of parameter space in which the function value decreases asymptotically approaching a limiting value as some parameter becomes arbitrarily large. A concrete example of this is a model containing a deflector at a fixed location and a secondary deflector some distance away whose position is allowed to vary. As the secondary deflector's position is moved arbitrarily far from the primary deflector, the effect of this secondary deflector asymptotically approaches that of an external shear. The symptoms of this problem appear in two different forms: The minimization algorithm may hit a maximum iteration limit without converging. Alternately, the function value may become so close to its asymptotic

limit, within the convergence tolerance of the algorithm, that a "minimum" is reported — but this "minimum" contains parameters that are so large as to be not physically meaningful, for example, a secondary deflector hundreds of arcseconds away from the primary.

For complicated deflector models, with $\gtrsim$11 parameters, minimization algorithms that walk downhill to the first encountered minimum, such as the Powell direction set method or the downhill simplex method, tend to find and become stuck in local minima. Once the Powell method reaches a local minimum, restarts do not lift it out of this minimum because all directions are uphill. In contrast, for the downhill simplex, restarts starting with a finite-sized simplex in some cases allowed the simplex to reach outside of small local minima, and make further progress downhill to a lower minimum. For one particular 11-parameter model, for example, 90 amoeba restarts were required before a final minimum was reached.

From this behavior of the minimization algorithms, one may conclude that the chi-squared surfaces for moderately complicated models have a number of local minima at higher values of chi-squared than the true local minimum. Indeed, one would expect a local minimum at deflector parameter values that cause some but not all of the model images to be aligned with the observed image positions.

The simulated annealing method was introduced because of the local minimum problem, and it was found that it dealt well with the small local minima. In very few cases did a restart from its best-found value find a lower minimum, and the need could be reduced by "cooling" at a slower rate.

Even the simulated annealing algorithm did not handle well the case of two different minima well-separated in parameter space. All our modeling started with a guess of the ring radius based on previous modeling work and a guess of the galaxy position based on the HST observations of Falco *et al.* (1997). Using simulated annealing with a high starting temperature (10 times the starting chi-squared) sometimes ended in a much higher local minimum with the model galaxy position well away from the optically observed galaxy position (which was not being used as a constraint). A high starting temperature thus introduced the possibility of escaping from the physically

reasonable portion of parameter space and becoming stuck in a deep local minimum far from the desired global minimum. Therefore a moderately low starting temperature was used (1% of the starting chi-squared) which allowed the exploration of the region of parameter space near the physically motivated best guesses of the galaxy location and the ring radius. The starting guesses for all $m \geq 1$ multipole moments were zero.

7.1.4 Error estimates and confidence limit calculation

Three different error estimates were used for the errors on the image positions. For a crude upper limit on the position error of each image, the fitted image size was used. This is the actual image size convolved with the VLBA beam width. Deconvolution errors, and centroid shifts due to magnification gradients, should both be less than this amount. The error estimate is referred to as "fit-size-as-errs" in some tables. A lower limit on the position error of each image is the statistical error on the fitted centroid due to thermal noise in the map, "centroid-errs". A better estimate of the position error of each image is to add the statistical error on the centroid position in quadrature with the deconvolved image size. Magnification gradients can cause image centroid shifts of the order of the image width times the fractional change in magnification over the image size. This error estimate ("deconv-size-n-centroid-errs") makes allowance for this as well as for deconvolution error.

The confidence limits on each model parameter were determined by stepping the parameter away from the chi-squared minimum, while minimizing over all other parameters, until a certain $\Delta\chi^2$ was obtained. The method and the reason for using this method are described in section 5.6 of this thesis. Nominal 68.3% confidence limits were found.

7.1.5 The time delays and error estimates on the time delays

The time delay between two images of the same source is given in equation 4.20. For modeling purposes it is convenient to introduce a dimensionless version of the time

delay by which image α of source l lags image β of source l:

$$\Delta\tau_{l,\beta\alpha} = \left(\frac{1}{2}|\vec{\nabla}_{\vec{r}_{l\alpha}}\Phi(\vec{r}_{l\alpha})|^2 - \frac{1}{2}|\vec{\nabla}_{\vec{r}_{l\beta}}\Phi(\vec{r}_{l\beta})|^2 - \Phi(\vec{r}_{l\alpha}) + \Phi(\vec{r}_{l\beta})\right). \tag{7.1}$$

This dimensionless time delay depends on the modeled potential, but does not depend on the assumed values of the cosmological parameters. It must be multiplied by a cosmological factor

$$\frac{(1+z_L)}{c}\frac{D_L D_S}{D_{LS}} \tag{7.2}$$

to convert the "dimensionless time delay" into a time delay. For MG J0414+0534, with $z_s = 2.639$ (Lawrence *et al.* 1995b), and assuming that $z_l = 0.5$, $\Omega_o = 1$, $\Omega_{\Lambda o} = 0$, and $H_o = 75 h_{75}$km/sec/Mpc then[1] this factor is

$$\frac{(1+z_L)}{c}\frac{D_L D_S}{D_{LS}} = \frac{2.845 \times 10^{12}}{h_{75}}\text{days} \tag{7.3}$$

Dimensionless time delays are computed from the best-fit model parameters for each model.

To compute error estimates on the time delay from the error estimates on the model parameters by standard error-propagation methods would require that the covariance of the model parameters be known — and these covariances were not computed. Instead the following method was used. Time delays were computed at the upper and lower confidence limits for each model parameter. The most extreme of these time delays were taken to be error estimates on the time delays. This method takes the covariance between model parameters into account, since at the confidence limit for a model parameter, the other parameters have been adjusted to minimize the chi-squared.

Keep in mind that these error estimates on the time delay do not include the systematic errors that may result from the interaction of the mass sheet degeneracy

[1]For a flat universe dominated by a cosmological constant, ($\Omega_o = 0$, $\Omega_{\Lambda o} = 1$). this cosmological factor is 3.2% higher, $\frac{(1+z_L)}{c}\frac{D_L D_S}{D_{LS}} = 2.937 \times 10^{12}\, h_{75}^{-1}$ days. The cosmological factor is 16.5% higher, $\frac{(1+z_L)}{c}\frac{D_L D_S}{D_{LS}} = 3.315 \times 10^{12}\, h_{75}^{-1}$ days, assuming an open, empty, no-cosmological-constant universe, $\Omega_o = 0$, $\Omega_{\Lambda o} = 0$.

with an error in the assumed radial profile. This systematic error is discussed in chapter 6, in particular in section 6.6.

7.2 Results from simple models previously used for modeling MG J0414+0534

In section 5.6 it was explained that the convolved size of each lens images (actual size convolved with the beam width) would be used as a crude upper limit on the position error for each image. Even though this is an overestimate of the VLBA positions errors, these position errors are still much smaller than those from VLA or HST observations — and positions could be measured for four subcomponents "a", "b", "c", and "d" which could not be done with the VLA or HST observations. Using this upper limit on the position errors, the simplest lens models for MG J0414+0534 are all rejected. These models include a singular isothermal sphere with external shear, "SIS+XS", as used by Falco *et al.* (1997), Ellithorpe (1995), and others, a point mass with external shear, "PM+XS", used by, for example, Falco *et al.* (1997), and the singular isothermal elliptical potential truncated at the quadrupole moment, "SIEP", used by Ellithorpe (1995) and which is very similar to that used by Hewitt *et al.* (1992) and Falco *et al.* (1997).

$$\Phi_{\rm SIS+XS}(r,\theta) = br + \frac{1}{2}\gamma r^2 \cos 2(\theta - \theta_\gamma) = br - r^2 A_2 \cos 2(\theta - \psi_{A_2}) \quad (7.4)$$

$$\Phi_{\rm PM+XS}(r,\theta) = b^2 \ln r + \frac{1}{2}\gamma r^2 \cos 2(\theta - \theta_\gamma) = b^2 \ln r - r^2 A_2 \cos 2(\theta - \psi_{A_2}) \quad (7.5)$$

$$\Phi_{\rm SIEP}(r,\theta) = br - \frac{1}{2} br\epsilon \cos 2(\theta - \theta_\epsilon) \quad (7.6)$$

As can be seen in table 7.1, the chi-squareds are extremely large. The model image locations are a qualitatively bad fit. This can be seen in figure 7-1 which shows the observed image locations, the model image locations, the ring radius, and the orientation of the mass that would cause this shear, for the SIS+XS model.

The best-fit model parameters are listed in table 7.1. The fitted values are similar

Goodness of fit			
Model	$\chi^2_{\rm min}$	$N_{\rm DOF}$	N_σ
SIS+XS	16803	19	3851
PM+XS	14477	19	3317
SIEP	23567	19	5402

Best-fit model parameters								
Model	g_x	g_y	b	γ	ϵ	$A_2{=}\gamma/2$, or $\epsilon/4$	$2\psi_{A_2}$, or $2\theta_\epsilon$	ψ_{A_2}, or θ_ϵ
SIS+XS	1.116	0.599	1.1849	0.127		0.064	-0.4133	78.16
PM+XS	1.098	0.597	1.1631	0.258		0.129	-0.4146	78.12
SIEP	1.323	0.537	1.1969		0.245	0.061	-0.4353	77.53

Time delays for best-fit model parameters							
Model	$\Delta\tau_{A1aA2a}$	$\Delta\tau_{A1aBa}$	$\Delta\tau_{A1aCa}$	Δt_{A1aA2a}	Δt_{A1aBa}	Δt_{A1aCa}	$\frac{\Delta\tau_{BaA1a}}{\Delta\tau_{A1aCa}}$
	dimensionless			h_{75}^{-1} days			
SIS+XS	3.84e-14	-1.54e-12	1.06e-11	0.109	-4.39	30.1	0.146
PM+XS	7.77e-14	-3.14e-12	2.09e-11	0.221	-8.92	59.3	0.150
SIEP	5.92e-14	-2.17e-12	2.05e-11	0.169	-6.19	58.2	0.106

Table 7.1: Results of fitting simple models to MG J0414+0534.
$N_{\rm DOF}$ is the number of constraints minus the number of fitted model parameters. For datasets drawn from the model, with errors of this size, the minimized chi-squared should have mean $\mu = N_{\rm DOF}$ and standard deviation $\sigma = \sqrt{N_{\rm DOF}}$. The column "$N_\sigma$" in the table shows by how many standard deviations the actual data differs from what would be expected from this model. Note that the convolved size of each lens image (actual size convolved with the beam width) was used to approximate the position error on each image. This gives an upper limit on the position errors, and thus a lower limit on the chi-squareds.
The deflector position (g_x, g_y) is in arcseconds from the position of the correlation center at A1, with g_x increasing to the west and g_y increasing the north. Ring radius b is in arcseconds. Quadrupole parameters ϵ and γ are dimensionless, as in equations 7.4, 7.5, and 7.6. Definitions of the quadrupole parameters differ by factors of 2; an additional column rescales these parameters for ease in comparison with the quadrupole moments of the modified-Multipole-Taylor model. The first angle column is in radians north of west. The second angle column, for comparison with previous modeling, is in degrees east of north.
The time delays are given for the brightest component "a". These are the light arrival times of the A2a, Ba, and Ca images relative to the A1a image. Ba leads and Ca lags. The first three columns are the "dimensionless time delays" described in equation 7.1. The second three columns have been multiplied by the cosmological factor in equation 7.3 to give the delay in days. Note that the time-delays (and the quadrupole moment) are twice as large for the PM monopole as for the SIS monopole.

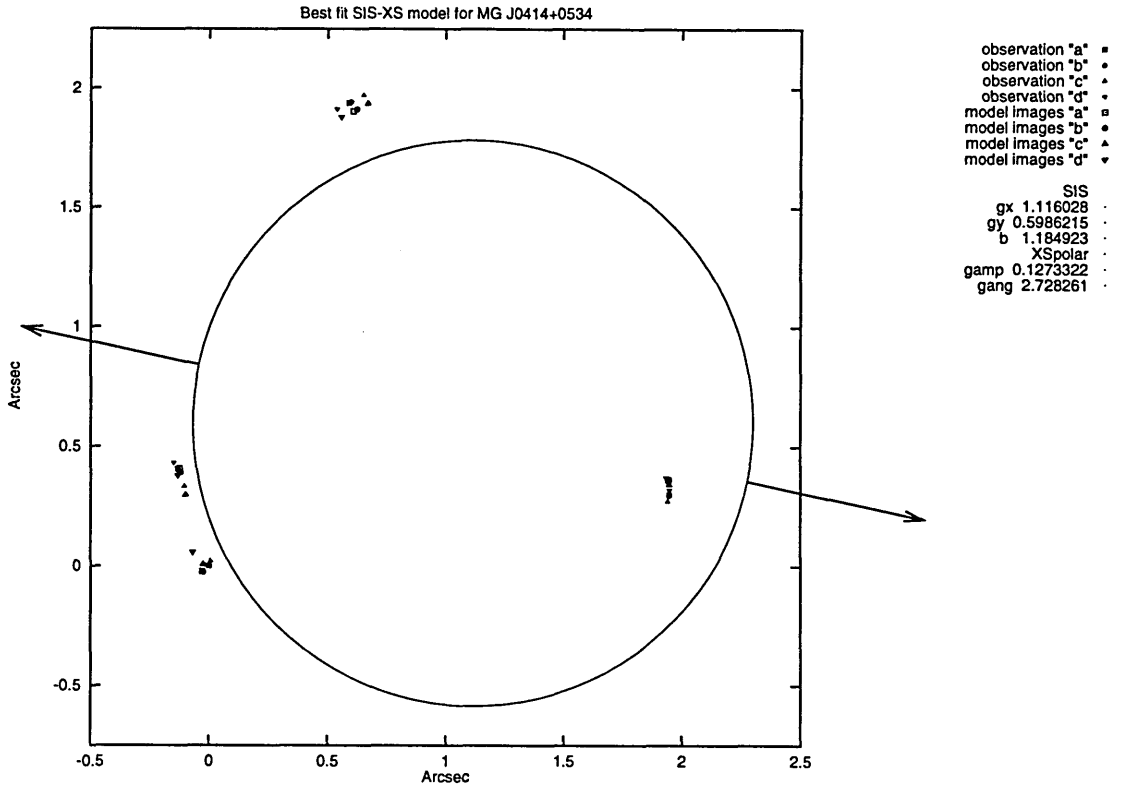

Figure 7-1: Singular isothermal sphere with external shear (SIS+XS) model, fitted to the MG J0414+0534 VLBA 5 GHz map. Observed image locations, model image locations, Einstein ring radius, and the orientation of the mass that would cause this shear are shown. It clearly fails to reproduce the observed image locations.

to those of Falco *et al.* (1997) and of Ellithorpe (1995), but a number of the fitted parameters lie somewhat outside of the previous works' error ranges. This is not surprising for the following reason. The models are inadequate and cannot reproduce all of the features of the observation. Changing the relative weights on the constraints would be expected therefore to change the fitted model parameters — since with one set of weights certain constraints would be well fit, and with other weights other constraints would be well fit.

No confidence limits were calculated since they would be meaningless for a model which fits the data so poorly.

7.3 Results from the Multipole-Taylor model and the modified Multipole-Taylor model

The simple five-parameter models above, with a fixed radial dependence and with a single shear term, were unable to reproduce the approximate image locations. Therefore the modified Multipole-Taylor model was used because it allows one to select, and add to the potential, terms that are expected to have a significant effect on lensing, while ignoring terms with less effect on lensing. These terms are tabulated in table 6.3 for the Multipole-Taylor model and in table 6.4 for the modified Multipole-Taylor model. The only terms that cause shifts in image positions for images at the ring radius are the first ($t = 1$) monopole ($m = 0$) term and the first two terms ($t = 0$ and $t = 1$, or "external" and "internal") of each higher multipole ($m \geq 1$) term.[2] For all other terms the image shifts go to zero for $(r - b)/b = 0$

7.3.1 Selection of a "basic model": terms to include in all models

The only monopole $m = 0$ term that is expected to be readily constrainable for lenses all of whose images lie near the ring radius is the linear $t = 1$ term. This "m0t1" term sets the ring radius and was used in all of the models.

Which $m \geq 1$ terms should be included? The quadrupole moment ($m = 2$) is expected to be the dominant $m \geq 1$ multipole term. Ellipticity in the deflector mass distribution would show up primarily in the quadrupole moment of the deflector. Also, the dominant effect of an external perturber would be to contribute to the quadrupole moment of the potential. Therefore both the external "m2ext" and internal "m2int" pieces of the quadrupole were included in all of the models.

Comparing this m0t1+m2ext+m2int model with the single-shear models of section 7.2 (SIS+XS $\equiv$ m0t1+m2ext and SIEP $\equiv$ m0t1+mixed-quadrupole) we see that

[2]However the first two $m = 1$ terms are not used as model parameters due to degeneracies in the model. See table 6.3 or 6.4.

the addition of the second shear term caused a vast improvement in the fit. The chi-squared dropped by two orders of magnitude. (The chi-squareds of the fitted models are tabulated in table 7.2. The fitted model parameters are listed and shown graphically in figures in appendix B.)

7.3.2 Adding an additional term to the basic model

Higher ($t \geq 2$) terms in the quadrupole, ("m2t2", "m2t3", ...), should cause little displacement in model image locations for images lying near the ring radius, so they are expected to have little effect. The first dipole term ("m1t2") and the next monopole term ("m0t3") also should cause little displacement for images near the ring. Therefore the $m \geq 3$ multipole terms were considered.

The second term in which effects of an external perturber would appear is the external octupole term ("m3ext"). If the perturber were nearby it would appear with lesser amplitude also in the "m4ext" and higher multipole terms. An asymmetry in the mass distribution of the primary deflector would be reflected in an internal octupole term ("m3int"). Diskiness or boxiness of the deflector's mass distribution could cause an $m = 4$ internal multipole term ("m4int"), or if the mass distribution extends past the ring radius a "m4ext" term.

Therefore, to the basic m0t1+m2ext+m2int model, were in turn added each of these terms: m3int, m3ext, m4int, and m4ext. The terms m0t3, m1t2, and m2t2 were also tried, even though they are expected to have less effect when the images lie near the ring radius, since they are the next most significant terms for their respective values of m. No $m \geq 5$ multipole components were tried, as there was little physical motivation for doing so. The m3t0 and m4t0 terms were also used, for a different balance between internal and external multipole effects on those multipole terms (half internal and half external). The chi-squareds are given in table 7.2 and the fitted model parameters are in appendix B.

Each one of these additional terms improved the fit. However, the greatest improvement was caused by an external or a mixed internal-and-external octupole term (m3ext or m3t0), each of which caused another order of magnitude drop in the fitted

Model	$N_{\text{deflector parms}}$	N_{DOF}	Error estimate on image positions errors: "fit-size-as-errs" (an upper limit) $\chi^2_{\min}$	"fit-size-as-errs" (an upper limit) N_σ	"centroid-errs" (a lower limit) $\chi^2_{\min}$	"centroid-errs" (a lower limit) N_σ	"deconv-size-n-centroid-errs" $\chi^2_{\min}$	"deconv-size-n-centroid-errs" N_σ
SIEP	5	19	2.357e4	5.402e3	1.297e8	2.976e7	3.667e5	8.412e4
SIS+XS (m0t1+m2ext)	5	19	1.680e4	3.851e3	1.078e8	2.474e7	1.910e5	4.382e4
PM+XS	5	19	1.448e4	3.317e3	1.012e8	2.321e7	2.080e5	4.771e4
m0t1+m2ext+m2int	7	17	2.485e2	5.615e1	1.636e5	3.967e4	3.996e3	9.651e2
+m3int	9	15	1.356e2	3.113e1	1.143e5	2.950e4	2.022e3	5.181e2
+mm2t2	9	15	1.757e2	4.149e1	9.875e4	2.549e4	1.357e3	3.464e2
+m0t3	8	16	1.108e2	2.371e1	1.040e5	2.601e4	1.438e3	3.554e2
+mm1t2	9	15	1.400e2	3.227e1	1.430e5	3.691e4	8.457e2	2.145e2
+m4ext	9	15	1.011e2	2.224e1	7.958e4	2.054e4	7.493e2	1.896e2
+m4t0	9	15	1.235e2	2.801e1	8.522e4	2.200e4	3.913e2	9.717e1
+m4int	9	15	4.168e1	6.889	5.551e4	1.433e4	1.983e2	4.733e1
+m3ext	9	15	1.619e1	0.308	1.831e4	4.723e3	1.351e2	3.100e1
+m3t0	9	15	8.783	-1.605	1.603e4	4.135e3	7.038e1	1.430e1
+m4ext+m4int	11	13	1.917e1	1.712	3.370e4	9.344e3	1.297e2	3.235e1
+m3t0+m0t3	10	14	8.113	-1.573	1.596e4	4.261e3	6.284e1	1.305e1
+m3t0+m4int	11	13	5.087	-2.195	6.435e3	1.781e3	5.647e1	1.206e1
+m3t0+m4t0	11	13	3.795	-2.553	6.130e3	1.697e3	2.862e1	4.331
+m3t0+m4ext	11	13	2.459	-2.924	3.998e3	1.105e3	1.239e1	-0.1692
+m3t0+mm2t2	11	13	3.285	-2.694	8.421e3	2.332e3	1.179e1	-0.3348
+m3t0+mm1t2	11	13	2.527	-2.905	7.794e3	2.158e3	1.145e1	-0.4291
+m3ext+m4t0	11	13	9.435	-0.989	1.223e4	3.389e3	5.506e1	1.167e1
+m3ext+m4int	11	13	7.374	-1.560	1.050e4	2.908e3	4.732e1	9.519
+m3ext+m0t3	10	14	9.144	-1.298	1.631e4	4.356e3	4.183e1	7.439
+m3ext+mm1t2	11	13	7.254	-1.594	1.528e4	4.235e3	3.612e1	6.413
+m3ext+m3int	11	13	1.654	-3.147	4.176e3	1.155e3	8.960	-1.120
+m3ext+mm2t2	11	13	4.783	-2.279	5.986e3	1.657e3	8.387	-1.279
+m3ext+m4ext	11	13	1.257	-3.257	3.863e3	1.068e3	6.073	-1.921
+m4ext+m4int+m0t3	12	12	1.814e1	1.773	2.871e4	8.285e3	8.107e1	1.994e1
+m3t0+m4int+m0t3	12	12	4.889	-2.053	6.342e3	1.827e3	3.179e1	5.711
+m3t0+m4t0+m0t3	12	12	3.791	-2.370	5.948e3	1.714e3	2.455e1	3.622
+m3t0+m4ext+m0t3	12	12	2.321	-2.794	3.798e3	1.093e3	1.234e1	0.099
+m3t0+mm2t2+m0t3	12	12	1.364	-3.070	2.795e3	8.033e2	9.373	-0.758
+m3t0+mm1t2+m0t3	12	12	2.203	-2.828	5.505e3	1.586e3	8.495	-1.012
+m3ext+mm1t2+m0t3	12	12	6.160	-1.686	1.445e4	4.169e3	3.431e1	6.439
+m3ext+m4t0+m0t3	12	12	3.376	-2.490	7.549e3	2.176e3	1.696e1	1.431
+m3ext+m3int+m0t3	12	12	1.571	-3.011	3.989e3	1.148e3	8.193	-1.099
+m3ext+mm2t2+m0t3	12	12	3.163	-2.551	4.895e3	1.410e3	7.334	-1.347
+m3ext+m4int+m0t3	12	12	2.000	-2.887	4.914e3	1.415e3	6.157	-1.687
+m3ext+m4ext+m0t3	12	12	0.720	-3.256	2.306e3	6.622e2	2.721	-2.679

Table 7.2: Results of fitting modified Multipole-Taylor models to MG J0414+0534.

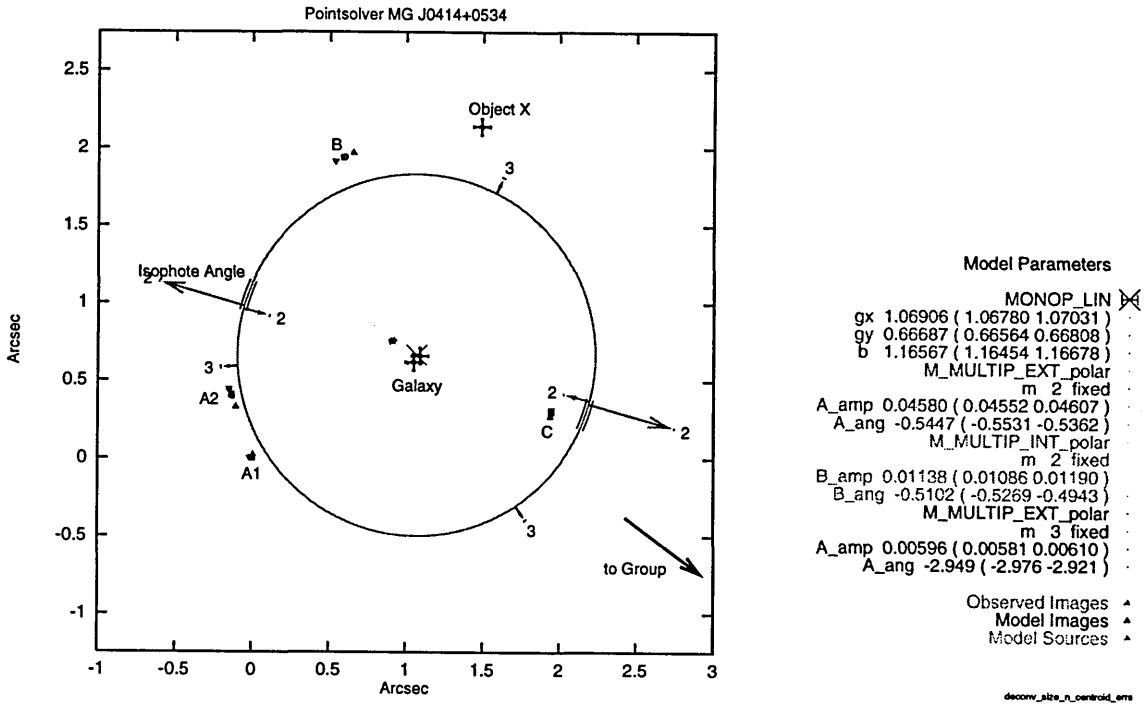

Figure 7-2: Model m0t1+m2ext+m2int+m3ext, fitted to MG J0414+0534. The table at right lists the fitted model parameters with confidence limits. These model parameters are also illustrated graphically. An × marks the model deflector center of mass (g_x, g_y), and a ring is drawn at the model Einstein ring radius b. The error range on the ring radius is shown by additional rings drawn at the upper and lower confidence limits. (In this plot of this model they cannot be distinguished from the Einstein ring itself) Each $m \geq 1$ multipole moment is illustrated by arrows pointing in the possible directions to mass perturbations (excesses) that would cause such a multipole moment. The arrows are labeled with the number m of the multipole moment, and their lengths are proportional to the amplitudes of the multipole moments. Confidence ranges on the angles are indicated by small arcs at the tip of each arrow. Also show on the plot are the optical location of the deflector galaxy (R-band and I-band, relative to the location of image C, from Falco *et al.* (1997)), the position of Object X (Falco *et al.* 1997), the direction to the nearby group of galaxies noted by Falco *et al.* (1997), and the orientation angle of the deflector's optical isophotes (Falco *et al.* 1997) indicated by arcs drawn near the Einstein ring radius. The locations of components "a", "b", "c", and "d" of images A1, A2, B, and C are also shown for reference, but cannot be seen clearly on a plot of this scale. The location of the model sources is also shown, slightly northeast of the deflector location.
In the table, "A_amp" is $|\vec{A}'_m|$ and "A_ang" is $m\psi_{A_m}$ (with $m = 2$) of equation 6.88. Similarly the parameters "B_amp" and "B_ang" give $\vec{B}'_m$, the parameters "G_amp" and "G_ang" (in some other figures) give $\vec{G}'_{mt}$, and the parameter "f" (in some other figures) gives f'_t of equation 6.88. The parameters "msum_amp" and "msum_ang" are the $\vec{M}'_{\text{sum}}$ of equation 6.52. Angles are in radians N of W.

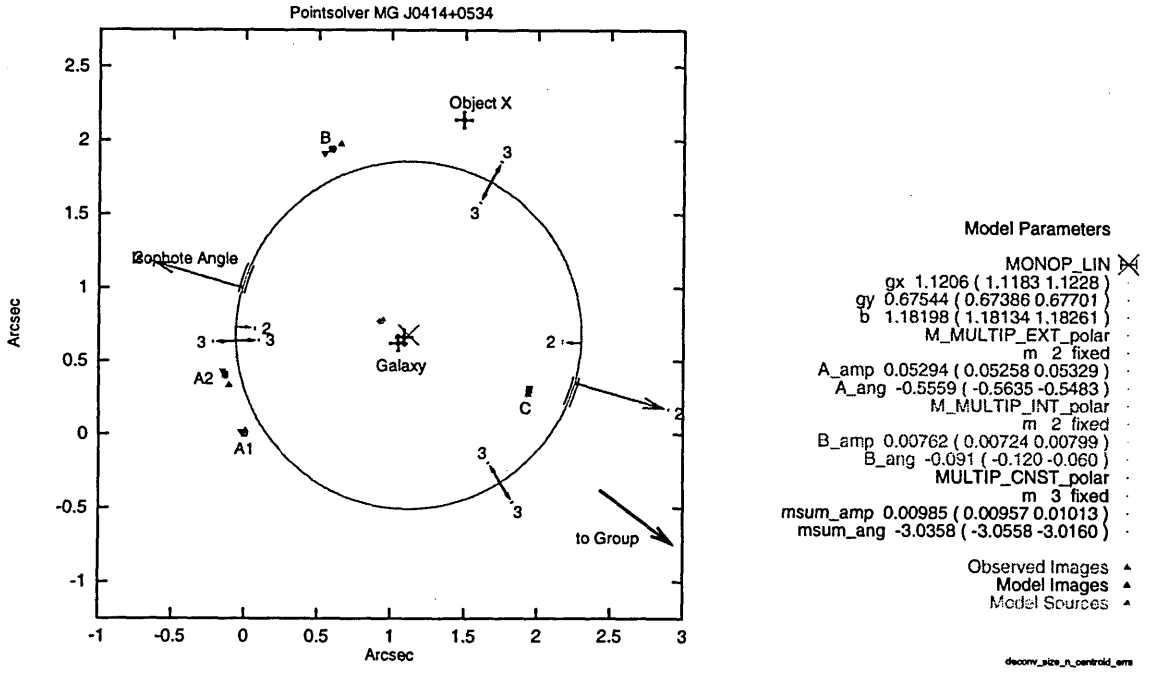

Figure 7-3: Model m0t1+m2ext+m2int+m3t0, fitted to MG J0414+0534. This figure and figure 7-2 show the two best-fitting 9-parameter models. An explanation of the symbols in the plot is given in the caption to figure 7-2.

chi-squared. Only these two models ("+m3ext" and "+m3t0")[3] fit the image positions well enough to satisfy the upper limits on the position errors, as given by the beam size convolved with the image extents. This is consistent with the finding of Ellithorpe (1995) that a 3-θ term improved the model fit. However, neither of the models reproduced the observed positions well enough to satisfy the better estimate of the position errors, the deconvolved image sizes added in quadrature to the statistical errors on the image centroids.

For both the +m3ext and +m3t0 models, the angle of the m2ext term is consistent with the optical isophote angle of the deflector as measured by Falco *et al.* (1997). The angles of the $m = 3$ terms of the two models are almost consistent with each other, and point somewhat to the east of Object X, somewhat to the south of the nearby group of galaxies noted by Falco *et al.* (1997), and somewhat counterclockwise from the isophote orientation. See figures 7-2 and 7-3. The angle of the internal $m = 2$ term is consistent with the optical isophote angle for the +m3ext model, but lies somewhat counterclockwise of the isophote orientation for the +m3t0 model.

7.3.3 Adding two terms to the basic model

Of the models examined above, the models which best fitted MG J0414+0534 had external or mixed-external-and-internal $m = 3$ multipole components: m0t1+m2ext-+m2int+m3ext $\equiv$ "+m3ext" and m0t1+m2ext+m2int+m3t0 $\equiv$ "+m3t0".

To these models were added, one at a time, the next most significant terms in each of the $m = 0, 1, 2, 3$, and 4 multipoles: m0t3; mm1t2; mm2t2; m3ext or m3t0 or m3int; m4ext or m4t0 or m4int. The $m = 0$, $m = 1$, and $m = 2$ terms were tried even though these terms should have little effect for $(r-b)/b$ small. The model +m4ext+m4int was also used to provide, for comparison, another model with the same number of model parameters, all of whose terms can cause significant displacements for images near the ring radius, but which contains no $m = 3$ term. The chi-squared are tabulated in table 7.2 and the fitted model parameters are given in appendix B.

[3] A model name preceded by a "+" means that the basic m0t1+m2ext+m2int terms are also part of the model.

As expected the +m4ext+m4int model was a worse fit than any of the models containing a $m = 3$ multipole term and also having 11 deflector parameters. In this group of models, the models containing the m3ext term were generally better fits than the models containing the m3t0 term. The best-fitting model with this number of deflector parameters was +m3ext+m4ext, (figure 7-4). The +m3ext+m4ext model has the lowest chi-squared, no matter which estimate of the position errors is used. Note that this model is too good of a fit, when using the deconvolved image sizes in quadrature with the statistical errors as a position error estimate. Therefore these somewhat overestimate the position errors.

Note that none of the models adequately fitted the error estimates given by the statistical errors alone. One problem is that these errors do not include the systematic errors in the deconvolution of the beam from the maps. Another problem may well be that magnification gradients are not properly being taken into account — adding deconvolved size to the statistical errors will crudely make allowances for this.

The model with the second-lowest chi-squared value, +m3ext+mm2t2 appears to have an unphysical value for its quadrupole parameter $\vec{G}'_{\substack{m=2\\t=2}}$.[4]

In the best fitting model, the +m3ext+m4ext model, both the internal and external quadrupoles are aligned with the optical isophote angle. The m4ext angle is anti-aligned with the m2ext angle and with the optical isophotes. By this I mean that the direction of mass deficit indicated by the m4ext term, 73.9° ± 1.0° E of N (plus multiples of 90°), is consistent with the optical isophote angle and has a confidence region which overlaps the confidence region of the external quadrupole angle. The m3ext angle, $\psi_{A_{m=3}}$, has been shifted about 13° clockwise relative to the +m3ext model. This is further from Object X, further from the group of galaxies, and closer to the angle of m2ext and the optical isophotes. The m3ext angle is approximately

[4]To have $|\vec{G}'_{\substack{m=2\\t=2}}| = 0.5$ (as this model requires) would require that of the mass in a narrow annulus around the Einstein ring, at least a quarter of it be involved in driving this quadrupole moment — and more if the perturbing mass is not in point-like locations. (This assumes that the mass of the deflector is not more extended than an SIS. See equation 6.110.) A large value of this $\vec{G}'_{\substack{m=2\\t=2}}$ coefficient is needed to cause much of an image displacement, because it has to overcome the small factor $(r - b)/b$. This term is apparently compensating for some of the internal and external quadrupole, which have been displaced relative to their orientations in the +m3ext model.

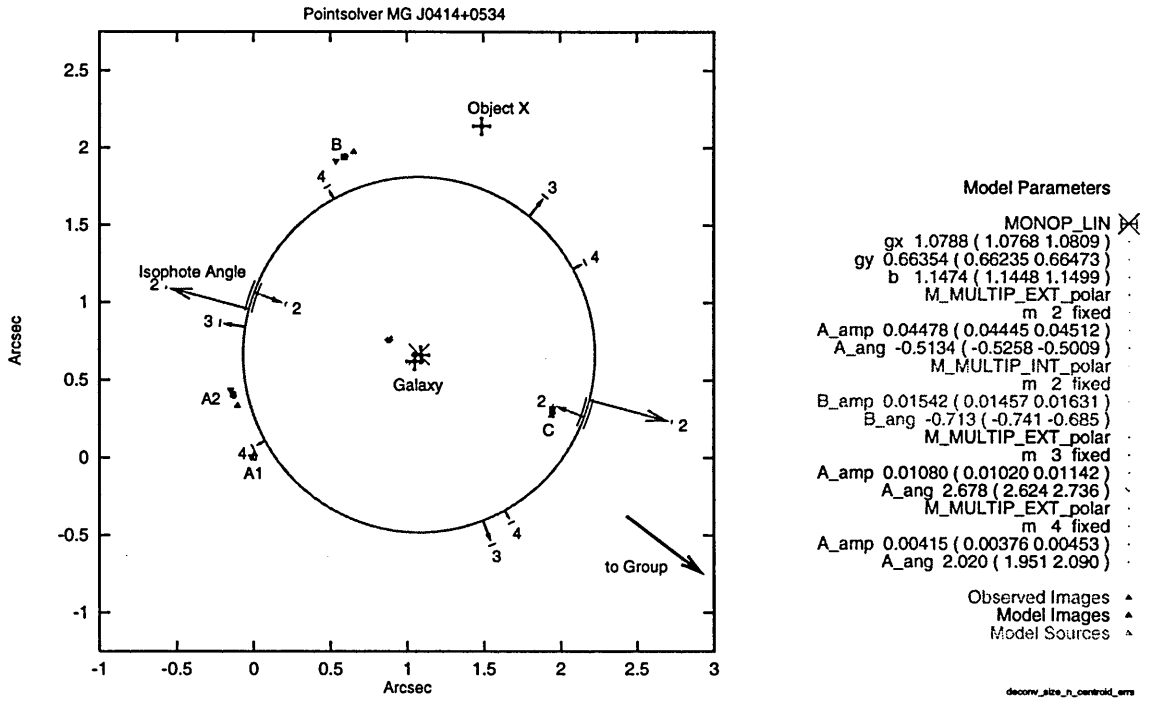

Figure 7-4: Model m0t1+m2ext+m2int+m3ext+m4ext. This is the best-fit 11-parameter model. An explanation of the symbols in the plot is given in the caption to figure 7-2.

aligned with the quadrupole in the east, and with the m4ext angle to the west of south, though the confidence ranges do not overlap.

One could attempt any of the following identifications:

1. The external $m = 2$ and $m = 3$ multipole components might indicate an external perturber to the east. It is only by coincidence that these angles line up with the optical isophotes. If there were such a perturber, and it were a point mass, its Einstein ring radius would be 0.95 ± 0.05 arcseconds and the distance to it would be 3.2 ± 0.2 arcseconds. The $m = 2$ internal quadrupole moment is aligned with the optical isophotes because they are both oriented with the mass of the deflector itself. The external $m = 4$ multipole moment is not explained.

   This explanation seems unlikely because (a) the external quadrupole is so well aligned with the optical isophotes, which is unlikely to happen by chance, and (b) no indication of a perturber in that approximate location is seen in the

I-band and R-band HST maps of Falco *et al.* (1997).

2. The external $m = 3$ and $m = 4$ multipole components might be attributable to an external perturber to the west of south, at about $-65° \pm 4°$ N of W. If this were the case, and the perturber were a point mass, the perturber's Einstein ring radius would be 0.56 ± 0.10 arcseconds and the distance to the perturber would be 2.24 ± 0.24 arcseconds. The perturber would contribute to the external part of the quadrupole term, and the contribution to the external quadrupole term, with amplitude 0.032 ± 0.013, would be significant relative to the dominant part from the primary deflector. Removing such a contribution from the external quadrupole moment would leave a residual external quadrupole of amplitude 0.060 oriented in the direction $2\psi_{A_{m=2}} = 0.052$ radians, or $\psi_{A_{m=2}} = 1.5°$ or 181.5° degrees N of W.

   The orientation of this residual external quadrupole moment is not consistent with the optical isophote angles. Its amplitude is large to be caused be the primary deflector, with the ratio external quadrupole to internal quadrupole of $|\vec{A}_{m=2}|/|\vec{B}_{m=2}| = 3.8$. In contrast, a deflector that followed a singular isothermal elliptical potential would have an external to internal quadrupole ratio of $|\vec{A}_{m=2}|/|\vec{B}_{m=2}| = 3$. Therefore the explanation (2) seems a poor explanation of the modeled multipole components. There is however an object seen in the optical in the HST images of Falco *et al.* (1997) at about 5 arcseconds from the deflector and at an orientation angle of about $-54°$ N of W.

3. The external and internal quadrupole ($m = 2$) moments are both aligned with the optical axis of the deflector because they are both caused by the ellipticity of the primary deflector. This indicates that the mass of the deflector and its ellipticity extends beyond the Einstein ring radius. A deflector that followed a singular isothermal elliptical potential would have an external to internal quadrupole ratio of m2ext/m2int of $|\vec{A}_{m=2}|/|\vec{B}_{m=2}| = 3$, with which this model's value of 2.90 ± 0.17 is consistent; therefore this quadrupole ratio is consistent with the deflector being as extended as an isothermal profile. The ellipticity of

the isopotential contours near the ring radius is $\epsilon_\Phi = 2|\vec{A}_{m=2} + \vec{B}_{m=2}| = 0.120 \pm 0.002$. In contrast the ellipticity of a singular isothermal elliptical potential having an isodensity ellipticity of $\epsilon_G = 0.20 \pm 0.02$ (equal to the mean ellipticity of the fitted isophotes of Falco *et al.* (1997)) would be $\epsilon_\Phi = 0.07$.

The $m = 4$ multipole term is due to a boxiness in the outer regions of the deflector's mass distribution. The $m = 3$ multipole moment is caused by an asymmetry in the outer parts of the deflector's mass distribution, with more mass at one end of the isophote axis than at the other. It is only by coincidence that the $m = 3$ term is sufficiently misaligned with the deflector axis so as to be approximately aligned with the $m = 4$ term in the south.

Note that this model's $m = 4$ external multipole amplitude of $(4.1 \pm 0.4) \times 10^{-3}$ is an order of magnitude larger than the $m = 4$ external amplitude of 6×10^{-4} that a singular isothermal elliptical potential of this ellipticity would have (equation 6.116), and it is oriented in the opposite direction (boxy rather than disky).

Only interpretation (3) adequately explains the best-fit model's parameter values. This best-fit model, +m3ext+m4ext, indicates that neither Object X nor the group of galaxies to the southwest contributes significantly to the lensing.

7.3.4 The deflector's radial dependence

To each of the 5-term, 11-parameter models above, were added a radial dependence ("m0t3") to the monopole term. The minimized chi-squareds are tabulated in table 7.2 and fitted model parameters are given in appendix B.

The fitted values of the radial profile parameter are given in table 7.3. The confidence ranges are quite large. The fitted value of the radial profile parameter shifts depending on what other multipole components are included in the model — the formal confidence ranges do not even all overlap. For mass densities that do not increase with radius, the value of this radial profile parameter must be less than one. For some models the entire confidence range lies within the unphysical region $f_3' > 1$.

Model	f_3'	$\chi^2_{\rm min}$	Model	$\chi^2_{\rm min}$
+m3ext+m0t3	$2.48^{+0.09}_{-0.09}$	4.183e1	+m3ext	1.351e2
+m3t0+m0t3	$0.75^{+0.24}_{-0.26}$	6.284e1	+m3t0	7.038e1
+m4ext+m4int+m0t3	$0.75^{+0.16}_{-0.17}$	8.107e1	+m4ext+m4int	1.297e2
+m3t0+m4int+m0t3	$-2.90^{+0.30}_{-0.25}$	3.179e1	+m3t0+m4int	5.647e1
+m3t0+m4t0+m0t3	$-3.59^{+1.57}_{-0.99}$	2.455e1	+m3t0+m4t0	2.862e1
+m3t0+m4ext+m0t3	$-0.24^{+0.91}_{-2.18}$	1.234e1	+m3t0+m4ext	1.239e1
+m3t0+mm2t2+m0t3	$-9.65^{+3.30}_{-8.50}$	9.373	+m3t0+mm2t2	1.179e1
+m3t0+mm1t2+m0t3	$-1.99^{+1.14}_{-0.98}$	8.495	+m3t0+mm1t2	1.145e1
+m3ext+m4t0+m0t3	$2.49^{+0.18}_{-0.21}$	1.696e1	+m3ext+m4t0	5.506e1
+m3ext+m4int+m0t3	$2.30^{+0.17}_{-0.19}$	6.157	+m3ext+m4int	4.732e1
+m3ext+mm1t2+m0t3	$1.09^{+0.81}_{-0.81}$	3.431e1	+m3ext+mm1t2	3.612e1
+m3ext+m3int+m0t3	$1.20^{+0.89}_{-1.35}$	8.193	+m3ext+m3int	8.960
+m3ext+mm2t2+m0t3	$-0.69^{+0.67}_{-0.58}$	7.334	+m3ext+mm2t2	8.387
+m3ext+m4ext+m0t3	$1.13^{+0.39}_{-0.52}$	2.721	+m3ext+m4ext	6.073

Table 7.3: Radial profile parameter. Fitted values and formal confidence ranges for models fitted to MG J0414+0534 using the "deconv-size-n-centroid-errs" error estimates. For comparison, the models' chi-squareds and the chi-squareds of the corresponding models without the radial profile parameter terms are also listed. Recall that values of $f_3' > 1$ are unphysical, indicating that the surface mass density increases with radius. A point mass would have $f_3' = 1$ and a singular isothermal sphere would have $f_3' = 0$.

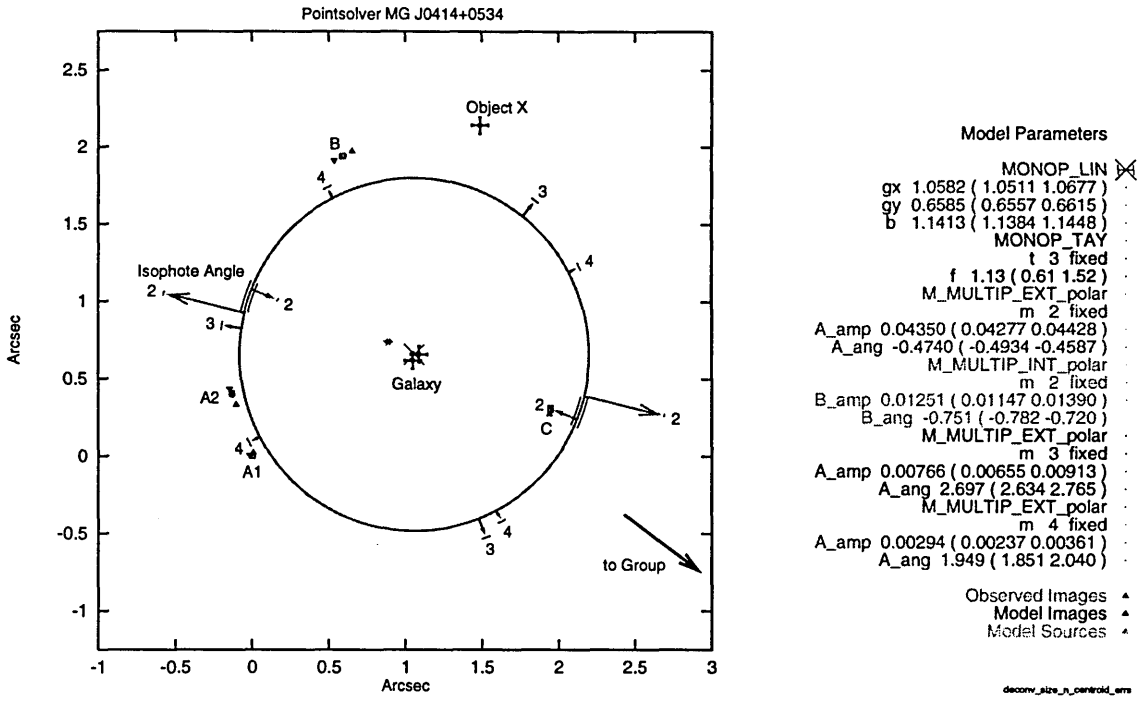

Figure 7-5: Model m0t1+m0t3+m2ext+m2int+m3ext+m4ext. This is the best-fit 12-parameter model that includes a radial profile parameter.

That this parameter's value is difficult to determine by model fitting can be explained. It takes a large change in this parameter's value to cause a small change in the radial positions of images near the ring radius. In contrast, small values of the $m \geq 2$ internal and external multipole components can cause shifts in the radial positions of images near the ring radius. The radial image shifts caused by the m0t3 term do not depend on angular position, whereas the radial image shifts caused by the $m \geq 2$ internal and external multipole moments do depend on angular position. However since for MG J0414+0534 there are only images at a few angular locations, there can be some interplay between the effects of the m0t3 term and of the $m \geq 2$ multipole terms. For example, for the +m3ext, the +m3ext+m4int, and the +m3ext-+m4t0 models, which were not adequate fits, the addition of the radial profile term greatly lowered the chi-squareds to values closer to those of some the models containing different $m \geq 2$ multipole terms, at the cost of unphysically large values of the radial profile parameter (formal lower confidence limit ≥ 2). For MG J0414+0534,

the radial profile parameter is not usefully constrained by lens modeling. This is expected to be the case for other quad lenses, also.

For systems with rings of lensed emission, information is available from imaged flux at all angles. Therefore, for ring systems it remains to be seen whether the radial profile parameter can be constrained.

For quads, the available information on the radial distribution of the matter in the deflector appears to be the balance between the amplitudes of the internal and external multipole moments — for those lenses in which the external multipole components are attributable to the outer regions of the primary deflector rather than to secondary perturbers.

7.3.5 Time delays

Time delays for the 9-parameter, 11-parameter, and 12-parameter deflector models with the lowest chi-squareds are given in tables 7.4 and 7.6. For the 11-parameter and 12-parameter models, only models are included which adequately fit the observation when using the deconvolved image size in quadrature with the statistical centroid errors as the position error estimate (requiring $N_\sigma < 3$ for "deconv-size-n-centroid-errs"). The +m3ext+m4t0+m0t3 and +m3ext+m4int+m0t3 models are excluded because their radial profile parameters, f_3', lie in the unphysical region with the lower confidence limit greater than 2.[5] The +m3ext+mm2t2 model is rejected because it has an apparently unphysically large value of its parameter $\vec{G}'_{\substack{m=2\\t=2}}$[6]

Note that there is a systematic trend that models with higher values of the radial profile parameter f_3' have shorter time delays than models with the radial profile parameter fixed to zero as suitable for a singular isothermal profile.

The best model time delays are from +m3ext+m4ext: $\Delta\tau_{A1aA2a} = 5.15^{+0.19}_{-0.17} \times 10^{-14}$, $\Delta\tau_{A1aBa} = -1.802^{+0.034}_{-0.036} \times 10^{-12}$, $\Delta\tau_{A1aCa} = 1.044^{+0.018}_{-0.017} \times 10^{-11}$, and $\Delta\tau_{BaCa} = 1.225^{+0.017}_{-0.017} \times 10^{-11}$, where the error bars are the formal errors from the

[5]Dimensionless time delays for all models, including the poorly fitting models, are given in appendix B in tables B.1 through B.3.

[6]See note on page 204.

Model	$\Delta\tau_{A1aA2a}$	$\Delta\tau_{A1aBa}$	$\Delta\tau_{A1aCa}$	$\Delta\tau_{BaCa}$
+m3ext	$5.19^{+0.09}_{-0.09}$	$-2.326^{+0.038}_{-0.039}$	$0.866^{+0.009}_{-0.009}$	$1.099^{+0.012}_{-0.012}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0	$5.88^{+0.09}_{-0.09}$	$-2.481^{+0.037}_{-0.037}$	$0.996^{+0.012}_{-0.012}$	$1.244^{+0.016}_{-0.016}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+m4ext	$6.60^{+0.14}_{-0.13}$	$-2.616^{+0.047}_{-0.047}$	$0.987^{+0.032}_{-0.030}$	$1.249^{+0.031}_{-0.029}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+mm2t2	$6.19^{+0.16}_{-0.13}$	$-2.610^{+0.066}_{-0.077}$	$1.133^{+0.020}_{-0.017}$	$1.394^{+0.017}_{-0.016}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+mm1t2	$4.85^{+0.12}_{-0.12}$	$-2.061^{+0.059}_{-0.068}$	$0.838^{+0.027}_{-0.023}$	$1.044^{+0.034}_{-0.029}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m3int	$5.89^{+0.12}_{-0.11}$	$-2.486^{+0.034}_{-0.028}$	$1.069^{+0.060}_{-0.051}$	$1.318^{+0.061}_{-0.052}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+mm2t2[7]	$5.36^{+0.03}_{-0.04}$	$-1.985^{+0.019}_{-0.020}$	$0.811^{+0.015}_{-0.015}$	$1.010^{+0.017}_{-0.017}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m4ext	$5.15^{+0.19}_{-0.17}$	$-1.802^{+0.034}_{-0.036}$	$1.044^{+0.018}_{-0.017}$	$1.225^{+0.017}_{-0.017}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+m4ext+m0t3	$6.80^{+1.72}_{-0.72}$	$-2.694^{+0.273}_{-0.703}$	$1.010^{+0.249}_{-0.092}$	$1.280^{+0.318}_{-0.117}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+mm2t2+m0t3	$7.09^{+5.31}_{-1.74}$	$-2.912^{+0.664}_{-2.092}$	$0.864^{+0.548}_{-0.164}$	$1.155^{+0.757}_{-0.230}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+mm1t2+m0t3	$6.02^{+0.92}_{-0.79}$	$-2.604^{+0.304}_{-0.360}$	$0.928^{+0.093}_{-0.085}$	$1.188^{+0.129}_{-0.116}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m3int+m0t3	$4.93^{+1.11}_{-0.51}$	$-2.133^{+0.161}_{-0.410}$	$0.949^{+0.148}_{-0.093}$	$1.162^{+0.186}_{-0.108}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+mm2t2+m0t3	$5.88^{+0.47}_{-0.51}$	$-2.183^{+0.194}_{-0.174}$	$0.857^{+0.063}_{-0.063}$	$1.075^{+0.077}_{-0.079}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m4ext+m0t3	$4.70^{+0.22}_{-0.17}$	$-1.734^{+0.029}_{-0.029}$	$0.913^{+0.061}_{-0.046}$	$1.086^{+0.062}_{-0.045}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$

Table 7.4: "Dimensionless time delays" for models fitted to MG J0414+0534, using the "deconv-size-n-centroid-errs" position error estimate. The time delays are given for the brightest component "a". Image B leads and image C lags.

[7]Model +m3ext+mm2t2 has an apparently unphysically large value of its parameter $\vec{G}'_{\substack{m=2\\t=2}}$. See note on page 204.

Model	$\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa}$	$\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa}$
+m3ext	$0.268^{+0.003}_{-0.003}$	$0.276^{+0.003}_{-0.003}$
+m3t0	$0.249^{+0.002}_{-0.002}$	$0.256^{+0.002}_{-0.002}$
+m3t0+m4ext	$0.265^{+0.010}_{-0.011}$	$0.274^{+0.011}_{-0.011}$
+m3t0+mm2t2	$0.230^{+0.007}_{-0.008}$	$0.237^{+0.007}_{-0.008}$
+m3t0+mm1t2	$0.246^{+0.002}_{-0.002}$	$0.253^{+0.002}_{-0.002}$
+m3ext+m3int	$0.233^{+0.011}_{-0.011}$	$0.239^{+0.011}_{-0.012}$
+m3ext+mm2t2	$0.245^{+0.004}_{-0.004}$	$0.253^{+0.004}_{-0.004}$
+m3ext+m4ext	$0.173^{+0.006}_{-0.006}$	$0.178^{+0.006}_{-0.006}$
+m3t0+m4ext+m0t3	$0.267^{+0.015}_{-0.012}$	$0.275^{+0.015}_{-0.013}$
+m3t0+mm2t2+m0t3	$0.337^{+0.020}_{-0.017}$	$0.348^{+0.021}_{-0.018}$
+m3t0+mm1t2+m0t3	$0.245^{+0.001}_{-0.001}$	$0.253^{+0.001}_{-0.001}$
+m3ext+m3int+m0t3	$0.225^{+0.013}_{-0.008}$	$0.231^{+0.014}_{-0.008}$
+m3ext+mm2t2+m0t3	$0.255^{+0.016}_{-0.014}$	$0.263^{+0.016}_{-0.015}$
+m3ext+m4ext+m0t3	$0.190^{+0.012}_{-0.012}$	$0.196^{+0.012}_{-0.012}$

Table 7.5: Time delay ratios for models fitted to MG J0414+0534.

Model	Δt_{A1aA2a}	Δt_{A1aBa}	Δt_{A1aCa}	Δt_{BaCa}
	h_{75}^{-1} days	h_{75}^{-1} days	h_{75}^{-1} days	h_{75}^{-1} days
+m3ext	$0.1476^{+0.0024}_{-0.0024}$	$-6.62^{+0.11}_{-0.11}$	$24.6^{+0.2}_{-0.2}$	$31.3^{+0.3}_{-0.3}$
+m3t0	$0.1673^{+0.0024}_{-0.0024}$	$-7.06^{+0.10}_{-0.10}$	$28.3^{+0.4}_{-0.4}$	$35.4^{+0.5}_{-0.5}$
+m3t0+m4ext	$0.1879^{+0.0039}_{-0.0038}$	$-7.44^{+0.13}_{-0.13}$	$28.1^{+0.9}_{-0.9}$	$35.5^{+0.9}_{-0.8}$
+m3t0+mm2t2	$0.1762^{+0.0045}_{-0.0038}$	$-7.43^{+0.19}_{-0.22}$	$32.2^{+0.6}_{-0.5}$	$39.7^{+0.5}_{-0.5}$
+m3t0+mm1t2	$0.1381^{+0.0034}_{-0.0035}$	$-5.86^{+0.17}_{-0.19}$	$23.8^{+0.8}_{-0.6}$	$29.7^{+1.0}_{-0.8}$
+m3ext+m3int	$0.1677^{+0.0034}_{-0.0031}$	$-7.07^{+0.10}_{-0.08}$	$30.4^{+1.7}_{-1.4}$	$37.5^{+1.7}_{-1.5}$
+m3ext+mm2t2	$0.1524^{+0.0009}_{-0.0010}$	$-5.65^{+0.05}_{-0.06}$	$23.1^{+0.4}_{-0.4}$	$28.7^{+0.5}_{-0.5}$
+m3ext+m4ext	$0.1465^{+0.0053}_{-0.0049}$	$-5.13^{+0.10}_{-0.10}$	$29.7^{+0.5}_{-0.5}$	$34.8^{+0.5}_{-0.5}$
+m3t0+m4ext+m0t3	$0.1936^{+0.0489}_{-0.0205}$	$-7.66^{+0.78}_{-2.00}$	$28.7^{+7.1}_{-2.6}$	$36.4^{+9.1}_{-3.3}$
+m3t0+mm2t2+m0t3	$0.2016^{+0.1511}_{-0.0496}$	$-8.29^{+1.89}_{-5.95}$	$24.6^{+15.6}_{-4.7}$	$32.9^{+21.5}_{-6.6}$
+m3t0+mm1t2+m0t3	$0.1495^{+0.0048}_{-0.0056}$	$-6.26^{+0.21}_{-0.12}$	$25.5^{+0.5}_{-0.9}$	$31.8^{+0.6}_{-1.1}$
+m3ext+m3int+m0t3	$0.1402^{+0.0317}_{-0.0146}$	$-6.07^{+0.46}_{-1.17}$	$27.0^{+4.2}_{-2.6}$	$33.1^{+5.3}_{-3.1}$
+m3ext+mm2t2+m0t3	$0.1672^{+0.0135}_{-0.0144}$	$-6.21^{+0.55}_{-0.49}$	$24.4^{+1.8}_{-1.8}$	$30.6^{+2.2}_{-2.2}$
+m3ext+m4ext+m0t3	$0.1338^{+0.0063}_{-0.0049}$	$-4.93^{+0.08}_{-0.08}$	$26.0^{+1.7}_{-1.3}$	$30.9^{+1.8}_{-1.3}$

Table 7.6: Time delays for models fitted to MG J0414+0534. The time delays are given for the brightest component "a". Image B leads and image C lags. This is the same information as in table 7.4, but each dimensionless time delay has been converted into a time delay in days, using a multiplicative factor appropriate for $z_l = 0.5$, $\Omega_o = 1$, $\Omega_{\Lambda o} = 0$, and $H_o = 75 h_{75}$km/sec/Mpc.

model fit. There is more uncertainly due to the uncertainty in the choice of models, with $\Delta\tau_{A1aBa}$ ranging from -2.9×10^{-12} to -1.7×10^{-12}; $\Delta\tau_{A1aCa}$ ranging from 0.84×10^{-11} to 1.13×10^{-11}; and $\Delta\tau_{BaCa}$ ranging from 1.04×10^{-11} to 1.39×10^{-11}; for the models adequately fitting the "deconv-size-n-centroid-errs".[8]

Ratios of $\Delta\tau_{AB}/\Delta\tau_{AC}$ may be used to further distinguish between models, if both time delays are measured. The measured $\Delta\tau_{AB}/\Delta\tau_{AC}$ may be able to confirm that +m3ext+m4ext is the best model for this system, as the ratios $\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa} = 0.173 \pm 0.006$ and $\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa} = 0.178 \pm 0.006$ fall outside the error ranges of all the other models. (See table 7.5. The error ranges of the time delay ratios of the +m3ext+m4ext and +m3ext+m4ext+m0t3 models do overlap, but the latter model is the same as the former model with the addition of a radial profile term.) Of the models that adequately satisfy the "deconv-size-n-centroid-errs" error estimate, the next closest in time delay ratio to our best-fit model, besides +m3ext+m4ext+m0t3, is +m3ext+m3int+m0t3 with $\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa} = 0.225^{+0.013}_{-0.008}$ and $\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa} = 0.231^{+0.014}_{-0.008}$, which are 30% higher than the values for the best-fit +m3ext+m4ext model. The lower limit of 0.217 on $\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa}$ for this model is 18% higher than the upper limit of 0.184 on $\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa}$ for the best-fit model +m3ext+m4ext. Therefore, measuring the time delay ratio $\Delta\tau_{AB}/\Delta\tau_{AC}$ with 18% errors should be adequate to distinguish between our best-fit model +m3ext+m4ext (or its generalization +m3ext+m4ext+m0t3) and the rest of the models.

Using the delays for the brightest radio component "a" whose relative positions are consistent with the relative positions of the optical images, and using an average of the A1 and A2 arrival times under the assumption that the joint A1-A2 light curve will be used in measuring the time delay, then

$$\Delta\tau_{BA} = 1.828 \times 10^{-12} \left(1^{+\underbrace{0.020}_{0.018}}_{-} \pm \underbrace{0.014}_{\substack{\text{A1-A2} \\ \text{difference}}} {}^{+\underbrace{0.616}_{0.038}}_{-} \right), \qquad (7.7)$$

[8]If the model +m3ext+mm2t2 is included despite the apparently unphysical value of its parameter $\vec{G}'_{\substack{m=2 \\ t=2}}$ (see note on page 204) then $\Delta\tau_{A1aCa}$ ranges from 0.81×10^{-11} and $\Delta\tau_{BaCa}$ ranges from 1.01×10^{-11}

$$\Delta\tau_{AC} = 1.042 \times 10^{-11} \left(1 \pm \underbrace{0.017}_{\substack{\text{formal} \\ \text{errors}}} \pm \underbrace{0.002}_{\substack{\text{A1-A2} \\ \text{difference}}} \overset{+0.085}{\underset{-0.198}{\underbrace{}_{\substack{\text{which} \\ \text{model}}}}} \right), \tag{7.8}$$

$$\Delta\tau_{BC} = 1.225 \times 10^{-11} \left(1 \pm \underbrace{0.017}_{\substack{\text{formal} \\ \text{errors}}} \overset{+0.138}{\underset{-0.148}{\underbrace{}_{\substack{\text{which} \\ \text{model}}}}} \right). \tag{7.9}$$

The first error term is due to the formal error on the fit of the +m3ext+m4int model. The second error term on $\Delta\tau_{BA}$ and $\Delta\tau_{AC}$ is just half the difference between the A1 and A2 arrival times. The last error term gives the spread of the time delays over the models that adequately satisfy the "deconv-size-n-centroid-errs" error estimate.[9]

This last uncertainty term can be reduced if the time delay ratio $\Delta\tau_{AB}/\Delta\tau_{AC}$ is measured. A measurement of $\Delta\tau_{AB}/\Delta\tau_{AC}$ to $\sim 2-3\%$, if it confirms the +m3ext+m4int model's time delay ratio, could exclude all the other models, allowing the "which model" uncertainty term to be dropped. A measurement of $\Delta\tau_{AB}/\Delta\tau_{AC}$ to only $\sim 18\%$, assuming it confirms the +m3ext+m4int model's time delay ratio, would exclude all the other models except the generalization of the best-fit 11-parameter model: +m3ext+m3int+m0t3, and for this model we may limit the radial profile parameter to stay within the physical range $f_3' \leq 1$. With this verification of the model form, the model predicted dimensionless time delays would be

$$\left.\begin{aligned}
\Delta\tau_{BA} &= 1.828 \times 10^{-12} \left(1 \overset{+0.020}{\underset{-0.018}{\underbrace{}_{\substack{\text{formal} \\ \text{errors}}}}} \pm \underbrace{0.014}_{\substack{\text{A1-A2} \\ \text{difference}}} \overset{+0.000}{\underset{-0.037}{\underbrace{}_{\substack{\text{which} \\ \text{model}}}}} \right) \\
\Delta\tau_{AC} &= 1.042 \times 10^{-11} \left(1 \pm \underbrace{0.017}_{\substack{\text{formal} \\ \text{errors}}} \pm \underbrace{0.002}_{\substack{\text{A1-A2} \\ \text{difference}}} \overset{+0.000}{\underset{-0.111}{\underbrace{}_{\substack{\text{which} \\ \text{model}}}}} \right) \\
\Delta\tau_{BC} &= 1.225 \times 10^{-11} \left(1 \pm \underbrace{0.017}_{\substack{\text{formal} \\ \text{errors}}} \overset{+0.000}{\underset{-0.100}{\underbrace{}_{\substack{\text{which} \\ \text{model}}}}} \right)
\end{aligned}\right\} \begin{array}{l} \text{Assuming the time delay ratio measured to } \sim 18\% \text{ agrees with the best-fit model.} \end{array} \tag{7.10}$$

These uncertainties in the time delays listed in this chapter do not include the systematic error due to the interaction of the mass sheet degeneracy with the choice of the model angularly-averaged surface mass density at the Einstein ring radius.

[9] If the model +m3ext+mm2t2 is included despite the apparently unphysical value of its parameter $\vec{G}'_{\substack{m=2 \\ t=2}}$ (see note on page 204) then the fractional error on the time delay $\Delta\tau_{AC}$ is -0.223 not -0.198, and the fractional error on the time delay $\Delta\tau_{BC}$ is -0.176 not -0.148.

Unfortunately the $\Delta\tau_{BA}/\Delta\tau_{AC}$ time delay ratio cannot resolve this degeneracy.

The ratio of the model-predicted dimensionless time delay to the true measurable time delay is the cosmology-dependent distance factor

$$\frac{\Delta\tau}{\Delta t} = \frac{c}{(1+z_L)}\frac{D_{LS}}{D_L D_S}. \tag{7.11}$$

The modeled value of $\Delta\tau$ may be compared with a yet-to-be-measured value of Δt (for any pair of images) and the result be used to determine this distance factor and thereby H_o. Assuming that $z_l = 0.5$, $\Omega_o = 1$, and $\Omega_{\Lambda o} = 0$ then[10]

$$H_o = 75\frac{\text{km}}{\text{sec Mpc}}\frac{5.201\,\text{days}}{\Delta t_{BA}}(1-\kappa)\left(1^{+\underbrace{0.020}_{0.018}}_{-}\pm\underbrace{0.014}_{\substack{\text{A1-A2}\\\text{difference}}}{}^{+\underbrace{0.616}_{0.038}}_{-}\right) \tag{7.12}$$

$$\text{formal errors}\qquad \text{which model}$$

$$H_o = 75\frac{\text{km}}{\text{sec Mpc}}\frac{29.64\,\text{days}}{\Delta t_{AC}}(1-\kappa)\left(1\pm\underbrace{0.017}_{\substack{\text{formal}\\\text{errors}}}\pm\underbrace{0.002}_{\substack{\text{A1-A2}\\\text{difference}}}{}^{+\underbrace{0.085}_{0.198}}_{-}\right) \tag{7.13}$$

$$\text{which model}$$

$$H_o = 75\frac{\text{km}}{\text{sec Mpc}}\frac{34.84\,\text{days}}{\Delta t_{BC}}(1-\kappa)\left(1\pm\underbrace{0.017}_{\substack{\text{formal}\\\text{errors}}}{}^{+\underbrace{0.138}_{0.148}}_{-}\right) \tag{7.14}$$

$$\text{which model}$$

where, as explained above, measurement of the time delay ratio may reduce the error due to "which model", and where the factor $(1-\kappa)$ is due to the interaction of an incorrect model value of the surface mass density at the Einstein ring radius, $\sigma_o(b)$, with the mass sheet degeneracy. The modeling was done using an angularly-averaged surface mass density at the Einstein ring appropriate for an isothermal profile, $\sigma_o(b) = \sigma_{\text{crit}}/2$, therefore if the true deflector profile is also an isothermal profile, then $\kappa = 0$.

[10]The Hubble parameter values are 3.2% higher assuming $\Omega_o = 0$, $\Omega_{\Lambda o} = 1$, and 16.5% higher assuming $\Omega_o = 0$, $\Omega_{\Lambda o} = 0$.

Chapter 8

Conclusion

8.1 New observation of MG J0414+0534

Observation of the gravitational lens MG J0414+0534 with the VLBA at 6 cm wavelength revealed substructure that was not resolvable at lower resolutions. This is shown in figure 3-6 and figures 3-7 through 3-10. Each image A1, A2, B, and C, clearly shows four subcomponents "a", "b", "c", and "d". These are named in order of decreasing peak flux in the VLBA maps. Image B, which is highly stretched in the direction of the component separations, shows another subcomponent "e" close to subcomponent "a". Subcomponent "d" is quite extended in all four images A1, A2, B, and C.

Previous VLBI maps showed clearly only two subcomponents. The 5 GHz observation of Ellithorpe (1995) shows both "a" and "d" in images A1, A2, and B, with "a" being the brighter. The 1.7 GHz EVN map of Patnaik & Porcas (1996) also shows A1, A2, and B to have some substructure, but with "d" clearly separated from "a" only in image B, and with "d" having flux closer to that of "a" at 1.7 GHz than at 5 GHz. The secondary component in A2 in the 1.7 GHz map appears to be at the location of component "c".

8.1.1 Correspondence between radio and optical sources

The relative A1, A2, B, and C image locations for subcomponents "a", "b", "c", and "d", were compared with the optical image locations reported by Falco *et al.* (1997). The relative positions of subcomponents "c" and "d" were clearly incompatible with the relative positions of the optical source. The relative positions of subcomponents "a" and "b" are both compatible with the relative positions of the optical sources. (Note however that the Ba-Ca separation in RA is large enough to be marginally inconsistent with the optical.) Therefore if the source of the optical flux is coincident with any of these radio sources, it must be the brightest source "a" or its closest companion "b".

8.1.2 Resolution effects

The total summed flux in a 0.2×0.2 arcsecond box around each image was 0.336 Jy for A1, 0.242 Jy for A2, 0.125 Jy for B, and 0.054 Jy for C. By comparison, a VLA observation at 5 GHz (Katz *et al.* 1997) found fluxes of 0.401 Jy, 0.362 Jy, 0.156 Jy, and 0.058 Jy for A1, A2, B, and C, respectively. The VLBA to VLA flux ratios are 0.84, 0.67, 0.80, and 0.93, respectively. This indicates that a certain amount of the flux has been resolved out of the VLBA maps for images A1 and B, and even more so for image A2. This is unsurprising. Image C which shows the least linear magnification has been least resolved out, and image A2 which shows the most linear magnification has been most resolved out. Most of the flux resolved out has presumably been lost from source "d" which is quite extended. (See figure 3-6.)

8.1.3 Insight into the radio/optical flux ratio discrepancy

An outstanding MG J0414+0534 puzzle is the cause of the radio/optical A1/A2 flux ratio discrepancy, with the flux ratio A1/A2 being 2.5 in R-band and 2.1 in I-band (Falco *et al.* 1997), which is higher than that observed at radio wavelengths. The A1/A2 ratios of the integral fluxes from elliptical Gaussian fits to the image components at 5 GHz were 1.18 for "a", 1.25 for "b", 0.89 for "c", and 1.24 for "d".

Subcomponent "d" is rather extended, so the elliptical Gaussians fit to the peaks of A1d and A2d do not measure all of the flux. The ratio of cleaned flux in A1 (all components) to cleaned flux in A2 is rather higher, 1.43, but a certain amount of flux remains in the residuals of image A2 near A2d, and it is expected that A2d has lost more flux from being resolved out than has A1d.

If the optical source lies at or close to the radio source a or b, as permitted by the relative locations of the images, then the radio/optical flux ratio discrepancy persists as before.

It has been suggested that magnification gradients over the region of the source, with radio and optical flux coming from different regions of the source, might explain the radio-optical A1/A2 flux ratio discrepancy. A slight magnification gradient has been observed over the extent of the source, as seen from the flux ratios, but clearly much less than would be needed to explain the radio-optical flux ratio discrepancy. For magnification gradients to provide an explanation, the optical source would have to be displaced perpendicularly away from the c-b-a-d axis, and there would have to be large magnification gradients in that direction, of which no indication is seen in the VLBA map.

Therefore either microlensing or extinction in the deflector galaxy remains a better explanation for the discrepancy than the magnification-gradient explanation.

8.1.4 Limits on radio emission from a fifth image of the source

Only four lensed images of each radio source were seen; no fifth image was seen. Katz *et al.* (1997) take the flux limits on the fifth image to be 10 times the RMS surface brightness in their maps, finding limits of 6.1, 2.0, 1.2, 2.5, and 3.3 mJy at 1.4, 5, 8, 15, and 22 GHz, respectively. The present observation's map RMS of somewhat over 0.2 mJy/Beam at 5 GHz confirms Katz's finding at 5 GHz but does not allow a better limit to be set.

8.2 Advances in modeling techniques

8.2.1 Advances in "point" modeling techniques

In this thesis the technique of "point modeling", treating subcomponents as point-like objects and using just their positions and perhaps their fluxes as modeling constraints, is justified on the grounds that it permits a computationally quick way of distinguishing between models and rejecting those which are truly a poor fit.

Validity of the "source-plane" approximation to "point" modeling.

The practise of transforming the position errors to the source plane by means of the magnification matrix (equation 5.13), to speed the computation of the chi-squared, is shown in section 5.2.1 to be equivalent to Taylor-expanding the difference between the model source-plane positions and the observed positions (as mapped back to the source-plane) in terms of a small quantity which is roughly the fractional change in magnification between the observed and model image-plane positions — and dropping the higher order terms in the Taylor expansion. Therefore for point sources, the "source-plane" approximation is valid as long as the model is a sufficiently good fit. The source-plane approximation can fail in two cases. The first case is when the deflector model is so poor that the model cannot reproduce the observed positions, but in that situation the model would be rejected anyway. The "source-plane" approximation can fail in a more subtle fashion if the error on the observed image locations is large enough that the magnification changes significantly within the error region. In such a case, the model's chi-squared could be low enough that the model is not rejected, yet the magnification at the model image locations could be significantly different from the magnification at the true image locations.

For the models that adequately fit the image positions from the VLBA 5 GHz observation of MG J0414+0534, the "source-plane" chi-squared at its minimum differed from the true "image-plane" chi-squared by no more that 1.5%, and typically only by 0.2–0.6%. Since this is much less than the chi-squared increment used in finding the confidence limits, the approximation did not introduce any appreciable error — and

saved significant computation time.

Reducing the dimensionality of minimization-algorithm parameter space

In minimizing the chi-squared, the parameter space to be searched has a dimensionality D equal to the number of deflector parameters plus the number of source parameters. It is shown that, for the source-plane approximation, the minimization over the sources parameters (positions and, if used, flux) may be done analytically. The source parameters that minimize the chi-squared for fixed values of the deflector parameters are given in equations 5.15 and 5.27. This significantly reduces the dimensionality of the parameter space to be searched. For four sources, ignoring fluxes, this reduced the dimensionality by 8 dimensions, a significant aid to any minimization algorithm.

Corrections to point modeling for slightly extended sources

For lenses with subcomponents which are not truly point-like, but which are slightly extended, a correction to point modeling may be needed because the centroid of the source may not map to the centroid of the image if there are magnification gradients. The "extended-source" correction to the chi-squared is given in equations 5.42 and 5.43, and the "extended-source" correction to the source-plane approximation to the chi-squared is given in equations 5.44 and 5.45. The expression in equation 5.44 for $\vec{s}'(\vec{r}_{l\alpha})$ is substituted for $\vec{s}(\vec{r}_{l\alpha})$ in equation 5.15 to give the "extended-source" corrections for the source positions that minimize the "source-plane" chi-squared.

The "point" modeling formulas are valid only if the magnification does not change significantly over the extent of each image. These "extended-source" corrections are valid for the case where the magnification does change somewhat over the extent of each image. Note that the validity of these approximations depends on how extended the source is rather than on how good of a fit the model is. For a deflector model which perfectly reproduces the observations, the "point-source" approximation could fail if the source were extended and there were magnification gradients across the images. For somewhat extended sources, the "slightly-extended" approximation accounts for

these magnification gradients; however even this "slightly-extended" approximation could fail if the source were too extended and the magnification gradients too large.

Use of components' major axes, minor axes, and orientation angles as modeling constraints

This is discussed in section 5.5.1 where the relation between source and image extents is given in equation 5.47, and where an expression for the correction for magnification gradients is given in equations 5.48 and 5.49.

Statistical errors on elliptical Gaussian fits

For "point" modeling, it is necessary to know the correlation between the x- and y-position errors. When including fluxes and image extents it is necessary to know the correlations between the measurement errors in these quantities also. The statistical errors, including correlations, are worked out in appendix A, for elliptical Gaussian fits to maps in which the noise is correlated on the scale of the beam.

8.2.2 New models suitable for ring and quad lenses

The surface mass density and the 2-D lens potential of a deflector are related by the 2-D Poisson equation, therefore it is natural to expand both the density and the potential in a multipole expansion in angle, as each multipole component of the potential depends only on the corresponding multipole component of the surface mass density. The question that is not so obvious is how to handle the radial dependence of each multipole component. In electromagnetism, one is interested primarily in fields at radii $r \gg r_{\rm max}$ where $r_{\rm max}$ is the extent of the source terms, the charges and currents; therefore one expands in terms of $r_{\rm max}/r$, discarding all higher order terms. This is inappropriate in lensing since the region of interest is the region where the images are located, which is near the region where the mass is located. Kochanek (1991) suggested expanding in the distance from the ring radius. This is appropriate for quads and rings since all images lie near the ring. This expansion is done, and it is

shown how each, t^{th}, term in the Taylor expansion of each multipole component of the potential relates to the underlying surface mass density. In particular, the constant ($t = 0$) piece in the m^{th} multipole moment depends on the sum of contributions to that multipole moment from mass exterior and mass interior to the ring radius. The next, linear ($t = 1$), piece in the m^{th} multipole moment depends on the difference of contributions to that multipole moment from mass exterior and mass interior to the ring radius. Given the fitted value of a multiple component, this directly indicates the possible directions to the perturbing mass that might be driving that multipole moment.

A modification to the Multipole-Taylor model, which makes the physical meaning of the parameters clearer, is to reparameterize and directly use $\vec{A}_m$ and $\vec{B}_m$, the multipole moments of the surface mass density exterior and interior respectively to the ring (equations 6.42 and 6.43), instead of using their sum and difference as parameters. Therefore the first two terms in the Taylor expansion of the radial dependence are replaced by the "external" and "internal" pieces. Rather than simply taking the first two parameters to be the difference and sum of the linear and constant pieces in the Taylor expansion, once can also include the effects of $\vec{A}_m$ and $\vec{B}_m$ to all orders in the radial expansion parameter, $\rho = (r - b_{\text{E}})/b_{\text{E}}$. This leaves the quadratic and higher ($t \geq 2$) terms depending only on the surface mass density near the ring, that is, between the innermost and outermost images. The meaning of the quadratic ($t = 2$) term is then very clear — it is just the multipole moment of the surface mass density in a circular annulus near the ring radius (equation 6.93). An upper limit on its amplitude (equation 6.108) arises from the upper limit on the amplitudes of $m \geq 1$ multipole moments of the surface mass density (equation 6.26).

This model parameterization makes it very easy to visualize the location of the mass driving the multipole moment — at least the direction in which it lies and whether it is interior or exterior to the Einstein ring.

The strength of a multipole moment caused by an external perturber depends on the strength of the perturber (the perturber's own Einstein ring radius $b_{\text{perturber}}$) as well as on the distance to R the perturber (equations 6.120 and 6.121). Therefore

the distance to an external perturber cannot be determined from a single multipole component. However, in principle, both the distance to a point mass perturber and the perturber's Einstein ring radius can be found from two multipole moments (the m^{th} and m'^{th} multipole moments with $m \neq m'$) if they are sufficiently well constrained.

There are several degeneracies in the mapping from the source plane to the image plane that cause certain transformations of the potential be to completely unconstrained by lensing. ((Falco *et al.* 1985; Gorenstein *et al.* 1988), see discussion in section 4.1.6 of this thesis.) The significance of these degeneracies, as to what they prevent one from learning about the mass distribution of the deflector, becomes clear in this modified Multipole-Taylor deflector model. The prismatic degeneracy simply prevents lens modeling from constraining in any way the dipole ($m = 1$) moment due to mass further from the origin than the furthest lensed image. The first $m \geq 1$ multipole moment in which such mass will show up is the quadrupole ($m = 2$ multipole) moment. (Mass which is further from the origin than the furthest lensed image does not affect the monopole $m = 0$ term either, apart from contributing an additive constant which has no effect, see equation 6.14.)

The other significant degeneracy is the mass sheet degeneracy, which in the terminology of Gorenstein *et al.* (1988) is called the magnification transformation. If the depth of the potential is scaled down by a factor $(1 - \kappa)$ and a mass sheet of surface mass density $\kappa\sigma_{\text{crit}}$ is added to the system to make up the missing amount of convergence, then the image locations and fluxes are unchanged (Falco *et al.* 1985). The consequence of this is that (1) one cannot constrain the angularly-averaged surface mass density of the deflector at the Einstein ring radius, $\sigma_o(b_E)$, and that (2) if one's model differs from the "true" value for the surface mass density at the ring radius, $\sigma_o(b_E)$, then all the model parameters (besides the location and ring radius) will be scaled by a factor that depends on the model-assumed $\sigma_o(b_E)$. More precisely, all parameters in which the model is linear, except for the Einstein ring radius, will be scaled by this scale factor. (These factors are tabulated for various cases in table 6.2.) This scaling also affects the model predicted time delay (Falco *et al.* 1985; Kochanek 1991).

While the angularly-averaged surface mass density at the ring radius, $\sigma_o(b_{\rm E})$, is unconstrainable, the Taylor-expanded multipole model shows, however, to which features of the angularly-averaged radial profile the lens modeling is sensitive. The constant term in the Taylor expansion of the monopole is an additive constant to the potential and therefore has no effect. The linear terms sets the Einstein ring radius. The quadratic term is unconstrainable due to the mass sheet degeneracy. The first sensitivity to radial profile of the angularly averaged surface mass density comes in the cubic term. Its fitted parameter is

$$f_3' = 1 - b_{\rm E}\frac{d}{dr}\ln\left(1 - \frac{\sigma_o(r)}{\sigma_{\rm crit}}\right)\Bigg|_{r=b_{\rm E}}. \tag{8.1}$$

This is sensitive to the falloff in $\sigma_o(r)$ near the Einstein ring radius. If $\sigma_o(r)$ decreases with radius then $f_3' < 1$. However, it is more sensitive to radial falloff when $\sigma_o(b_{\rm E})$ is close to $\sigma_{\rm crit}$ than when $\sigma_o(b_{\rm E})$ is much less than $\sigma_{\rm crit}$.

If this cubic term cannot be constrained by lens modeling for a quad or ring lens, then one may conclude that the angularly-averaged radial profile is unconstrainable. If, however, this term can be constrained, then the surface mass density of the mass at the ring radius can be deduced in a model-dependent way from this radial profile term, and used to fix the mass sheet degeneracy. (See table 6.5.)

In summary, the Multipole-Taylor model and modified Multipole-Taylor model are very suited to modeling rings and quads. They allow one to include the terms needed to accurately reproduce the potential near the ring radius — which is where the constraints (images) are located and where one needs to know the potential in order to calculate the time delay. These models also allow one to exclude terms which only affect the potential at large distances from the ring radius, and which are therefore poorly constrained for rings and quads. These models are, however, not well suited for modeling lenses in which the images are at very different distances from the deflector.

8.3 Modeling results for MG J0414+0534

8.3.1 Modeling results and angular distribution of mass

The positions of components "a", "b", "c", and "d" of images A1, A2, B, and C, from the 5 GHz VLBA observation of MG J0414+0534, were used as constraints for modeling the lens's deflector potential. The source-plane approximation to the point-modeling chi-squared was used, without making corrections for the sources being slightly extended. As position error estimates, the deconvolved (intrinsic) size of each image was added in quadrature to the centroid errors due to the thermal noise in the maps. This makes adequate allowance for the centroid shifts that magnification gradients may induce in extended sources.

The abundance of position constraints, 24 net constraints on the deflector potential which are more than previously available, and the high resolution of the very long baseline interferometry, provide more stringent constraints on the deflector potential than were used in previous modeling work. Accordingly, the simplest 5-parameter deflector models (position, lens strength, and shear) were shown to be inadequate to describe the potential, allowing the investigation of further features of the deflector potential.

Starting from a basic model (position, lens strength, internal quadrupole, and external quadrupole/shear), additional model terms from the Multipole-Taylor and modified Multipole-Taylor models were added in turn, and the fits of the models to MG J0414+0534 were compared. It was clear that the external $m = 3$ multipole and the mixed-external-and-internal $m = 3$ multipole did most to improve the fit. However neither one adequately fit the position within the error estimates.

To these models (position, lens strength, internal and external ($m = 2$) quadrupoles, and external or mixed-external-and-internal $m = 3$ multipole) were added in turn additional terms from the Multipole-Taylor and modified Multipole-Taylor models. Adding any term improved the fit somewhat, and several of the models adequately satisfied the error estimates. Actually, the best-fit model was too good of a fit, with $\chi^2_{\rm min} = 6.1$ for 13 degrees of freedom, indicating, unsurprisingly, that to include the

deconvolved component sizes in the error estimate was to overestimate the errors.

This best-fit 11-deflector-parameter model had both an external $m = 3$ multipole and an external $m = 4$ multipole. The potential was

$$\Phi(x,y) = \Phi(\rho,\theta) = {b_{\rm E}}^2 \left(\rho - \left\{ (1+\rho)^2 \vec{A}'_2 + \frac{1}{(1+\rho)^2} \vec{B}'_2 \right\} \cdot (\hat{x}\cos 2\theta + \hat{y}\sin 2\theta) - \left\{ (1+\rho)^3 \vec{A}'_3 \right\} \cdot (\hat{x}\cos 3\theta + \hat{y}\sin 3\theta) - \left\{ (1+\rho)^4 \vec{A}'_4 \right\} \cdot (\hat{x}\cos 4\theta + \hat{y}\sin 4\theta) \right), \quad (8.2)$$

where the angle θ is given by $\tan\theta = (y-g_y)/(x-g_x)$, and the fractional distance from the Einstein ring radius $b_{\rm E}$ is $\rho = \left(\sqrt{(x-g_x)^2 + (y-g_y)^2} - b_{\rm E}\right)/b_{\rm E}$. See figure 7-4 for an illustration of the fitted model parameters. The fitted model parameters are listed below; the confidence limits are formal 68.3% confidence limits:

deflector positions:		
(W and N of the correlation	$g_x = 1.0788 \pm 0.0020$	arcseconds
center at A1)	$g_y = 0.6635 \pm 0.0012$	arcseconds
ring radius:	$b_{\rm E} = 1.1474^{+0.0025}_{-0.0026}$	arcseconds
internal quadrupole:	$B'_2 = 0.01542^{+0.00089}_{-0.00085}$	
$\vec{B}'_2 = B'_2(\hat{x}\cos 2\psi_{B_2} + \hat{y}\sin 2\psi_{B_2})$	$2\psi_{B_2} = -0.713^{+0.028}_{-0.028}$	radians N of W
	$\psi_{B_2} = 69.57^{+0.81}_{-0.80}$	degrees E of N
external quadrupole:	$A'_2 = 0.04478^{+0.00034}_{-0.00033}$	
$\vec{A}'_2 = A'_2(\hat{x}\cos 2\psi_{A_2} + \hat{y}\sin 2\psi_{A_2})$	$2\psi_{A_2} = -0.513^{+0.013}_{-0.012}$	radians N of W
	$\psi_{A_2} = 75.29^{+0.36}_{-0.35}$	degrees E of N
external $m = 3$ multipole:	$A'_3 = 0.01080^{+0.00062}_{-0.00060}$	
$\vec{A}'_3 = A'_3(\hat{x}\cos 3\psi_{A_3} + \hat{y}\sin 3\psi_{A_3})$	$3\psi_{A_3} = 2.678^{+0.057}_{-0.055}$	radians N of W
	$\psi_{A_3} = 81.15^{+1.09}_{-1.04}$	degrees E of N
external $m = 4$ multipole:	$A'_4 = 0.00415^{+0.00039}_{-0.00038}$	
$\vec{A}'_4 = A'_4(\hat{x}\cos 4\psi_{A_4} + \hat{y}\sin 4\psi_{A_4})$	$4\psi_{A_4} = 2.020^{+0.071}_{-0.069}$	radians N of W
	$\psi_{A_4} = 28.93^{+1.01}_{-0.99}$	degrees E of N
	$45° + \psi_{A_4} = 73.93^{+1.01}_{-0.99}$	degrees E of N

The deflector centroid (g_x, g_y) agrees with the optical galaxy position observed by Falco *et al.* (1997), even though the optical position was not used as a modeling constraint. The directions of the mass excess indicated by both the internal and external quadrupoles are aligned with the optical isophote angle $71° \pm 5°$ E of N measured by Falco *et al.* (1997). The external $m = 4$ multipole is anti-aligned with the optical isophotes and with the quadrupole. That is, the direction of the mass deficit indicated by the $m = 4$ term, $73.9° \pm 1.0°$ E of N, is consistent with the optical isophote angle and is almost consistent with the external quadrupole angle. The $m = 3$ multipole is approximately aligned with the quadrupole in the east and with the $m = 4$ multipole in the south, though the confidence ranges do not overlap.

The internal and external quadrupole moments being so closely aligned with each other and with the optical isophote axis indicates that they are both caused by the ellipticity of the primary deflector itself, and not by some external perturber. This indicates that the mass of the deflector and its ellipticity extend beyond the Einstein ring radius (1.147 arcseconds = 5.44 h_{75}^{-1} kpc).[1] A deflector that followed a singular isothermal elliptical potential would have an external to internal quadrupole ratio of $|\vec{A}_2|/|\vec{B}_2| = 3$, with which this model's value of 2.90 ± 0.17 is consistent. Therefore this quadrupole is consistent with the deflector being as extended as an isothermal profile. The ellipticity of the isopotential contours near the ring radius is $2|\vec{A}_{m=2} + \vec{B}_{m=2}| = 0.120 \pm 0.002$. In contrast the isopotential ellipticity would be 0.07 for a singular isothermal elliptical potential having an isodensity ellipticity of 0.20 ± 0.02 (equal to the mean ellipticity of the fitted optical isophotes of Falco *et al.* (1997)). The $m = 4$ term is due to a boxiness in the outer regions of the deflector's mass distribution. Note that the model's external $m = 4$ multipole amplitude of $(4.1 \pm 0.4) \times 10^{-3}$ is an order of magnitude larger than the $m = 4$ external amplitude of 6×10^{-4} that a singular isothermal elliptical potential of this ellipticity would have, and it is oriented in the opposite direction, boxy rather than disky. The $m = 3$ multipole cannot be attributed to an external perturber (see discussion starting on page 205) and therefore must be caused by an asymmetry in the outer parts of the deflector's mass distribution, with

[1] assuming that $z_l = 0.5$, $\Omega_o = 1$, $\Omega_{\Lambda o} = 0$, and $H_o = 75\, h_{75}$ km/sec/Mpc.

more mass at one end of the isophote axis than at the other. It is only by coincidence that the $m = 3$ term is sufficiently misaligned with the deflector axis so as to be almost as well aligned with the $m = 4$ term to the west of south as it is with the external quadrupole in the east. Note that neither Object X nor the group of galaxies to the southwest is implicated in the lensing.

8.3.2 Radial profile

The first model term that can give information on the radial profile of the angularly-averaged surface mass density is that monopole term which is cubic in the fractional distance ρ of the lensed images from the ring radius. The effect of such a term should be small and its coefficient parameter difficult to constrain since all the images are near the ring radius; nevertheless such a term was added to each 11-parameter deflector model. The fitted parameter values have large error ranges ($\pm$ 0.4 $-$ 0.5 for a radial profile term added to the best-fit 11-parameter model form) compared to the range of interesting values, from $f_3' = 1$ for a point mass to $f_3' = 0$ for a singular isothermal sphere monopole. More problematic is that the value of this parameter is extremely sensitive to which other $m \geq 1$ monopole components are present, with values ranging from -9.7 to 2.3 for 12-parameter models that adequately satisfy the position error estimates. It is clear that useful information on the radial profile of MG J0414+0534 is not obtained from this term. For other quad lenses the radial profile of the monopole term will prove similarly unconstrainable.

That this parameter's value is difficult to determine by model fitting can be explained. It takes a large change in the radial profile parameter to cause a small change in the radial positions of images near the ring radius. In contrast, small values of the $m \geq 2$ internal and external multipole components can cause shifts in the radial positions of images near the ring radius — in particular changes in the balance between internal and external will change the radial image displacements. The radial image shifts caused by the radial profile parameter do not depend on angular position, whereas the radial image shifts caused by the $m \geq 2$ internal and external multipole terms do depend on angular position. For quad lenses there are only images at a few

angular locations, so there can be some interplay between the effects of the radial profile term and of the $m \geq 2$ multiple terms, with large shifts in the radial profile parameter balancing small shifts in the $m \geq 2$ internal and external multipoles.

For systems with rings of lensed emission, information is available from images flux at all angles for constraining the $m \geq 2$ multipoles. Therefore for ring systems it remains to be seen whether the radial profile parameter can be constrained, but it will not be possible to do so unless the $m \geq 2$ internal and external multipoles can be very well constrained. MG J0414+0534's arc may provide more information to constrain MG J0414+0534's angular multipole moments — but only if its location can be very precisely measured. Measurements of its location to the precision of these VLBA observations would be useful.

That the radial profile parameter is so difficult to determine is unfortunate. The radial profile parameter provides information on how the angularly-averaged surface mass density decreases with radius near the Einstein ring radius. Such information on the angularly-averaged radial profile could be used to inform the choice of a model value for the angularly-averaged surface mass density at the ring radius, $\sigma_o(b_{\rm E})$. The quantity $\sigma_o(b_{\rm E})$ is not directly constrainable from lensing, but using a model value that differs from the true value would affect the time delay estimates.

For quad lenses, the available information on the radial distribution of the matter in the deflector is the balance between the amplitudes of the internal and external multipole moments — for those lenses in which the external multipole components are attributable to the outer regions of the primary deflector rather than to secondary perturbers. For MG J0414+0534, the ratio of external to internal quadrupole amplitudes, $|\vec{A}_2|/|\vec{B}_2| = 2.90 \pm 0.17$ is consistent with the value, 3, expected for an isothermal elliptical potential.

8.3.3 Model predicted time delays for use in determining the Hubble parameter

The model predicted time delays were calculated for the fitted models. The B-A, A-C, and B-C "dimensionless" time delays for the best-fit 11-parameter model are:

$$\Delta\tau_{BA} = 1.828 \times 10^{-12} \left(1 \overset{+\underbrace{0.020}_{\text{formal errors}}}{-\underbrace{0.018}_{\text{formal errors}}} \pm \underbrace{0.014}_{\substack{\text{A1–A2} \\ \text{difference}}} \overset{+\underbrace{0.616}_{\text{which model}}}{-\underbrace{0.038}_{\text{which model}}} \right), \tag{8.3}$$

$$\Delta\tau_{AC} = 1.042 \times 10^{-11} \left(1 \pm \underbrace{0.017}_{\text{formal errors}} \pm \underbrace{0.002}_{\substack{\text{A1–A2} \\ \text{difference}}} \overset{+\underbrace{0.085}_{\text{which model}}}{-\underbrace{0.198}_{\text{which model}}} \right), \tag{8.4}$$

$$\Delta\tau_{BC} = 1.225 \times 10^{-11} \left(1 \pm \underbrace{0.017}_{\text{formal errors}} \overset{+\underbrace{0.138}_{\text{which model}}}{-\underbrace{0.148}_{\text{which model}}} \right). \tag{8.5}$$

These time delays are for the brightest subcomponent "a" of each image — it is this component and its fainter nearest neighbor "b" that are consistent with the relative positions of the optical images. The light arrival times at A1 and A2 have been averaged to give the time for "A", under the assumption that the joint A1–A2 light curve will be used in measuring the time delay; the separate A1 and A2 time delays are given in chapter 7. The first error term is the formal (68%) error on the fit of the best-fit model, from the error estimates on the position errors. The second error term is just half the difference between the A1 and A2 light arrival times. The last error term gives the spread of time delays over the models that adequately satisfy the position errors. That the best-fit model is too good of a fit is evidence that the position errors are overestimated. Therefore this range is an overly large estimate of the ranges of time delays from adequate models.

The time delay ratio may be used to further distinguish the best-fit model from the other models that fit the image positions. A measurement of $\Delta\tau_{AB}/\Delta\tau_{AC}$ to $\sim 2-3\%$, if it confirms the best-fit 11-parameter model's time delay ratio, could exclude all the other models, allowing the "which model" uncertainty term to be dropped. Apart from the model which is our best-fit 11-parameter model form plus a radial profile term, the lower limit on $\Delta\tau_{BA}/\Delta\tau_{AC}$ for the model with the next closest time delay

ratio is 18% higher than the upper limit on $\Delta\tau_{BA}/\Delta\tau_{AC}$ for the best-fit model. Such a measurement, assuming it confirms the fitted model's time delay ratio, would exclude all the other models except the generalization of the best-fit 11-parameter model to include a radial profile parameter, for which model we may limit the radial profile parameter to stay within the physical range. With this verification of the model form, the model predicted dimensionless time delays would be

$$\left.\begin{aligned}
\Delta\tau_{BA} =& 1.828\times10^{-12}\left(1^{+\underbrace{0.020}_{\text{formal errors}}}_{-\underbrace{0.018}_{\text{formal errors}}} \pm \underbrace{0.014}_{\text{A1-A2 difference}} {}^{+\underbrace{0.000}_{\text{which model}}}_{-\underbrace{0.037}_{\text{which model}}}\right)\\
\Delta\tau_{AC} =& 1.042\times10^{-11}\left(1 \pm \underbrace{0.017}_{\text{formal errors}} \pm \underbrace{0.002}_{\text{A1-A2 difference}} {}^{+\underbrace{0.000}_{\text{which model}}}_{-\underbrace{0.111}_{\text{which model}}}\right)\\
\Delta\tau_{BC} =& 1.225\times10^{-11}\left(1 \pm \underbrace{0.017}_{\text{formal errors}} {}^{+\underbrace{0.000}_{\text{which model}}}_{-\underbrace{0.100}_{\text{which model}}}\right)
\end{aligned}\right\}\ \text{Assuming the time delay ratio measured to } \sim 18\% \text{ agrees with the best-fit model.} \tag{8.6}$$

These uncertainties in the time delays do not include the systematic error due to the interaction of the mass sheet degeneracy with the choice of the model angularly-averaged surface mass density at the Einstein ring radius. Unfortunately, the time delay ratio $\Delta\tau_{BA}/\Delta\tau_{AC}$ cannot resolve this degeneracy.

The "dimensionless time delay" above is related to the true measurable time delay by a cosmology-dependent distance factor:

$$\frac{\Delta\tau}{\Delta t} = \frac{c}{(1+z_L)}\frac{D_{LS}}{D_L D_S}. \tag{8.7}$$

The modeled value of $\Delta\tau$ may be compared with a yet-to-be measured value of Δt (for any pair of images) and the result be used to determine this distance factor and thereby the Hubble parameter H_o. Assuming that $z_l = 0.5$, $\Omega_o = 1$, and $\Omega_{\Lambda o} = 0$ then[2]

$$H_o = 75\frac{\text{km}}{\text{sec Mpc}}\frac{5.201\,\text{days}}{\Delta t_{BA}}(1-\kappa)\left(1^{+\underbrace{0.020}_{\text{formal errors}}}_{-\underbrace{0.018}_{\text{formal errors}}} \pm \underbrace{0.014}_{\text{A1-A2 difference}} {}^{+\underbrace{0.616}_{\text{which model}}}_{-\underbrace{0.038}_{\text{which model}}}\right) \tag{8.8}$$

[2]The Hubble parameter values are 3.2% higher assuming $\Omega_o = 0$, $\Omega_{\Lambda o} = 1$, and 16.5% higher assuming $\Omega_o = 0$, $\Omega_{\Lambda o} = 0$.

$$H_o = 75 \frac{\text{km}}{\text{sec Mpc}} \frac{29.64\,\text{days}}{\Delta t_{AC}} (1-\kappa) \left(1 \pm \underbrace{0.017}_{\substack{\text{formal}\\\text{errors}}} \pm \underbrace{0.002}_{\substack{\text{A1-A2}\\\text{difference}}} \underbrace{{}^{+\,0.085}_{-\,0.198}}_{\substack{\text{which}\\\text{model}}} \right) \tag{8.9}$$

$$H_o = 75 \frac{\text{km}}{\text{sec Mpc}} \frac{34.84\,\text{days}}{\Delta t_{BC}} (1-\kappa) \left(1 \pm \underbrace{0.017}_{\substack{\text{formal}\\\text{errors}}} \underbrace{{}^{+\,0.138}_{-\,0.148}}_{\substack{\text{which}\\\text{model}}} \right) \tag{8.10}$$

where, as explained above, measurement of the time delay ratio may reduce the error due to "which model", and where the factor $(1-\kappa)$ is due to the interaction of an incorrect model value of the angularly-averaged surface mass density at the Einstein ring radius, $\sigma_o(b_{\rm E})$, with the mass sheet degeneracy. The modeling was done using an angularly-averaged surface mass density at the Einstein ring appropriate for an isothermal profile, $\sigma_o(b_{\rm E}) = \sigma_{\rm crit}/2$. Therefore if the true deflector profile is also an isothermal profile, which is supported by the ratio of the external to internal quadrupole amplitudes, then $\kappa = 0$.

8.4 Perspective

Using time delay measurements to measure the effective distance to a gravitational lens, for use in determining the Hubble parameter, the proportionality constant in the redshift-distance relation, is very appealing for several reasons: (1) The large distances involved, with redshifts of $\sim \frac{1}{2}$ to ~ 3, reduce the effects of peculiar velocities of the observer and of the astronomical object on the deduced Hubble parameter. (2) Uncertainties in the normalization of the cosmic distance scale do not affect this method, since this is a one-step distance measurement that bypasses the cosmic distance ladder. (3) The physics of the time delay phenomenon is well understood: deflection of light in a gravitational field and the slowing of time in a gravitational potential well.

The chief sources of error in the determination of the Hubble parameter using this technique arise from the difficulty in adequately determining the characteristics of the gravitational field that deflects and delays the light. The only other measurements needed are the source and lens redshifts and the time delay itself which can

be measured to high precision if one is fortunate enough to see an abruptly rising or falling flare in the lens source emission. Among these gravitational uncertainties are the effects of density fluctuations along the line of sight (Bar-Kana 1996), effects due to the curvature of space-time or the presence of a cosmological constant, and the difficulty of determining the gravitational potential of the deflector having information on it from only a few lensed image locations.

For the determination of the deflector potential itself, the modeling work on MG J0414+0534 done in this thesis shows that with high precision constraints from VLBI observations of the positions of image subcomponents, most of the error in the model-predicted dimensionless time delays comes from the possibility of error in the selection of the model functional form to use, rather than being due to the imprecision of the positions themselves. Yet, this modeling work on MG J0414+0534 is encouraging: even trying a wide range of model forms, selected so as to explore the effects of the dominant terms in an expansion in small quantities, the error on the model-predicted dimensionless time delay (and therefore on the Hubble parameter due to the lens modeling) is under 15% if the B-C time delay is used to determine the Hubble parameter. The error may be reduced if measurements of the A-B and A-C time delays determine the ratio $\Delta\tau_{BA}/\Delta\tau_{AC}$ to $\sim 18\%$, verifying the best-fit 11-parameter model, and thereby excluding the other model forms except for the generalization of the best-fit 11-parameter model to include a radial profile parameter. This would reduce the uncertainty on the time delays due to the choice of model form, leaving only a $\sim +2\%/-4\%$ error on the determination of the Hubble parameter from the A-B time delay, or a $\sim +2\%/-10\%$ error using the B-C time delay. Measurement of the time delay ratio $\Delta\tau_{BA}/\Delta\tau_{AC}$ to $\sim 2-3\%$ could further reduce the uncertainty due to the choice of model forms.

While modeling difficulties due to the scarcity of information of the deflector potential may prevent this technique from being used for an extremely high precision measurement of the Hubble parameter, a time delay measurement on a single lens — provided its model is as well-constrained as MG J0414+0534's is, can constrain the Hubble parameter to a precision which in interestingly small, given the current

state of the measurements of H_o. Measurements from multiple lenses will serve as checks against modeling errors due to peculiarities in individual lenses. Such lens-time-delay determinations are valuable also for being completely independent of the other techniques for the determination of the Hubble parameter.

If MG J0414+0534's redshift and time delays can be measured, MG J0414+0534 will give a good no-cosmic-distance-ladder measurement of the Hubble parameter.

Appendix A

Errors on elliptical Gaussian fits to maps with correlated noise

In this appendix, using reasonable assumptions, are calculated the correlations in the errors of the fitted parameters, due to thermal noise, when fitting an elliptical Gaussian to a source in a radio-interferometry map.[1] This work takes into account the pixel-to-pixel noise correlations in interferometer maps, extending the work of Condon (1997) and Kogan (1996) who for simplicity neglected the pixel-to-pixel noise correlations in their covariance matrix calculation, and included correlated noise in only an approximate and semi-empirical way.

A.1 Introducing the covariance matrix of the fit of a model to jointly Gaussian random variables

Consider a data set of N data points, z_i for $i = 1, \ldots, N$, for which the measurement errors are normally distributed though not independent. The expected values of the data points, that is, the mean values of the distribution from which they are drawn, are denoted by $\langle z_i \rangle \equiv \overline{z}_i$. The covariance matrix of these jointly Gaussian random variables is denoted by $S_{ij}^{\rm data} = \sigma_{ij}^2 = \langle (z_i - \overline{z}_i)(z_j - \overline{z}_j) \rangle$.

[1] This does not include the deconvolution error.

A model is fitted to the data, $g_i(\vec{a})$, which predicts the $i = 1, \ldots, N$ data values z_i. The model is parameterized by M model parameters, which may be written as an M-dimensional vector $\vec{a}$, or as the vector's components a_α for $\alpha = 1, \ldots, M$. It is assumed that the model may be treated as linear in the model parameters $\vec{a}$ over the region of interest — which is reasonable since any well-behaved function may be approximated as being linear over a small enough region.

To find the best-fit model parameters for a particular data set $\{z_i\}$ drawn from the probability distribution for the data, one minimizes the maximum likelihood statistic with respect to the model parameters. For this case of Gaussian errors, the appropriate maximum likelihood statistic is the familiar χ^2,

$$\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} (z_i - g_i(\vec{a})) \{S^{\text{data}}\}_{ij}^{-1} (z_j - g_j(\vec{a})) . \tag{A.1}$$

Assuming that $g_i(\vec{a})$ is linearizable in the model parameters $\vec{a}$ in the region of interest near the χ^2 minimum, then the M best-fit model parameters $\vec{a}_{\text{fit}}$ that minimize the χ^2 can be found analytically by setting the M derivatives $\frac{d\chi^2}{da_\alpha}$ to zero. As a thought experiment, it is assumed that many datasets $\{z_i\}$ are drawn from the same distribution, and that for each dataset the fitted model parameters $\vec{a}_{\text{fit}}$ that minimize the χ^2 for that realization of the dataset are calculated. Then, for the ensemble of datasets, the mean $\langle \vec{a}_{\text{fit}} \rangle \equiv \overline{\vec{a}_{\text{fit}}}$ and the covariance of the fit results may be computed. The covariance of the fitted parameters may shown to be

$$\left\langle (a_{\text{fit}\alpha} - \overline{a_{\text{fit}\alpha}})(a_{\text{fit}\beta} - \overline{a_{\text{fit}\beta}}) \right\rangle = D^{-1}_{\alpha\beta}, \tag{A.2}$$

where

$$D_{\alpha\beta} \equiv \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{dg_i}{da_\alpha} \{S^{\text{data}}\}_{ij}^{-1} \frac{dg_j}{da_\beta}. \tag{A.3}$$

Note that $D_{\alpha\beta}$ is symmetric, and it has been assumed to be invertible.

$\mathbf{D}^{-1}$ is called the covariance of the fit. It gives the variance and the covariance of the errors in the fitted parameters due to the Gaussian errors in the data points $\{z_i\}$. The covariance matrix is also useful for calculating the expected value and the

standard deviation of functions of the fitted model parameters. For a function of the fitted parameters $f(\vec{a})$, in the region where f is linearizable, the expected value of f is $\langle f(\vec{a})\rangle = f(\langle\vec{a}\rangle)$ and its variance is

$$\left\langle \{f(\vec{a}) - f(\langle\vec{a}\rangle)\}^2 \right\rangle = \left.\frac{df}{da_\alpha}\right|_{\langle\vec{a}\rangle} D^{-1}_{\alpha\beta} \left.\frac{df}{da_\beta}\right|_{\langle\vec{a}\rangle}, \tag{A.4}$$

where the derivatives are evaluated at the expected value of $\vec{a}$.

It is this covariance matrix $\mathbf{D}^{-1}$ that this appendix computes for the parameters of an elliptical Gaussian fitted to a radio interferometry map in which the map noise is correlated from pixel to pixel.

A.2 From discrete to continuous data

When fitting a model to an interferometry map, the amplitude of the flux in the i^{th} pixel is the i^{th} data point z_i, for $i = 1, \ldots, N$. This amplitude can also be written $z(\vec{x}_i)$ where $\vec{x}_i$ is the (x, y) location of the i^{th} pixel. If the pixel spacing is small enough compared with the beam size then the beam is well sampled, and the map values change little from pixel to pixel.[2] In this case it is sensible to approximate the discrete sum over pixels with an integral over pixel locations. Let h be the pixel size, and h^2 be the pixel area; the integral approximation becomes exact in the limit that the pixel size h becomes arbitrarily small. The appropriate correspondence between discrete and continuous is:

i, j	$\longleftrightarrow$	$\vec{x}, \vec{y}$	Identifies the pixel by number/location
Nh^2	$\longleftrightarrow$	A	Total pixel area/integration area
$\sum_{i=1}^{N}$	$\longleftrightarrow$	$\int \frac{d^2x}{h^2}$	Summation/Integration
δ_{ij}	$\longleftrightarrow$	$h^2\delta^{(2)}(\vec{x} - \vec{y})$	Discrete/continuous delta function
f_i	$\longleftrightarrow$	$f(\vec{x})$	A function of the pixel number/location

(A.5)

[2] It is obvious that the source flux will not change significantly over scales much smaller than the beam size. It will be shown below that the effect of the map's thermal noise on the pixel values will not change significantly over scales much smaller than the beam size.

Similarly the definition of an inverse matrix is extended to the continuous case:

$$\sum_{j=1}^{N} M_{ij} M_{jk}^{-1} = \delta_{ik} \longleftrightarrow \int \frac{d^2x'}{h^2} M(\vec{x}, \vec{x}') M^{-1}(\vec{x}', \vec{y}) = h^2 \delta^{(2)}(\vec{x} - \vec{y}). \tag{A.6}$$

Therefore equation A.3, for the inverse covariance matrix of the model parameters of a fitted model, may be converted from a discrete sum over pixels to a continuous integral over pixel locations:

$$D_{\alpha\beta} = \int \frac{d^2x}{h^2} \int \frac{d^2y}{h^2} \left[\frac{\partial g(\vec{x})}{\partial a_\alpha}\right] S^{-1}_{\text{data}}(\vec{x}, \vec{y}) \left[\frac{\partial g(\vec{y})}{\partial a_\beta}\right]. \tag{A.7}$$

For known functional forms of the model $g(\vec{x})$ and the pixel-noise covariance matrix $S^{\text{data}}(\vec{x}, \vec{y})$, the elements $D_{\alpha\beta}$ are straightforwardly computed using Fourier transform techniques.

A.3 Fourier transform techniques

Using the relation that the noise correlation in the map depends only on the vector displacement of one pixel from another (verified below) then the covariance matrix for the pixel noise may be written as

$$S^{\text{data}}(\vec{x}, \vec{y}) = S^{\text{data}}(\vec{x} - \vec{y}). \tag{A.8}$$

In terms of its Fourier transform this is

$$S^{\text{data}}(\vec{x}, \vec{y}) = S^{\text{data}}(\vec{x} - \vec{y}) = \int \frac{d^2k}{(2\pi)^{2/2}} e^{i\vec{k}\cdot(\vec{x}-\vec{y})} \tilde{S}^{\text{data}}(\vec{k}) \tag{A.9}$$

where

$$\tilde{S}^{\text{data}}(\vec{k}) = \int \frac{d^2x}{(2\pi)^{2/2}} e^{-i\vec{k}\cdot\vec{x}} S^{\text{data}}(\vec{x}). \tag{A.10}$$

The inverse of $S^{\rm data}$, needed for equation A.7, is readily found using the Fourier transform:

$$S^{-1}_{\rm data}(\vec{x},\vec{y}) = S^{-1}_{\rm data}(\vec{x}-\vec{y}) = \int \frac{d^2k}{(2\pi)^{2/2}} e^{i\vec{k}\cdot(\vec{x}-\vec{y})} \left[\frac{h^{2\times 2}}{(2\pi)^2}\frac{1}{\tilde{S}^{\rm data}(\vec{k})}\right]. \tag{A.11}$$

The functional form of the model, too, may be Fourier transformed,

$$g(\vec{x}) = \int \frac{d^2k}{(2\pi)^{2/2}} e^{i\vec{k}\cdot\vec{x}}\, \tilde{g}(\vec{k}), \tag{A.12}$$

where

$$\tilde{g}(\vec{k}) = \int \frac{d^2x}{(2\pi)^{2/2}} e^{-i\vec{k}\cdot\vec{x}}\, g(\vec{x}). \tag{A.13}$$

Equations A.11 and A.13 may be used in equation A.7, giving the following expression for the inverse of the model parameters' covariance matrix:

$$D_{\alpha\beta} = \int \frac{d^2k}{(2\pi)^{2/2}} \frac{1}{\tilde{S}^{\rm data}(\vec{k})} \left[\frac{\partial\tilde{g}(\vec{k})}{\partial a_\alpha}\right]^* \left[\frac{\partial\tilde{g}(\vec{k})}{\partial a_\beta}\right]. \tag{A.14}$$

A.4 Covariance matrix for the map noise

For radio interferometry data, in which there is independent Gaussian white noise in each of the real and imaginary parts of each visibility data point, the resultant flux errors in the pixels of an interferometer map (due to this noise) are correlated with each other, and their correlation is given by the beam pattern.

$$S^{\rm data}(\vec{x},\vec{y}) \equiv \langle (z(\vec{x}) - \overline{z}(\vec{x}))\,(z(\vec{y}) - \overline{z}(\vec{y}))\rangle = \mu^2 \frac{I_{\rm beam}(\vec{y}-\vec{x})}{I_{\rm beam}(\vec{0})} \tag{A.15}$$

where μ is the root-mean-square noise in a single pixel and $I_{\rm beam}(\vec{y}-\vec{x})$ is the (dirty) beam pattern. Note that as presumed above, $S^{\rm data}(\vec{x},\vec{y}) = S^{\rm data}(\vec{x}-\vec{y})$.

The beam pattern varies from observation to observation. However, the effect on the model fits to the map, caused by the pixel-to-pixel correlation of the noise, should be mainly dependent on the high degree of correlation between nearby pixels. There should be little effect due to the low degree of correlation between widely separated

pixels. It is assumed that the central portion of the beam is well enough approximated by an elliptical Gaussian, which can be parameterized by a major axis, minor axis, and orientation angle.

$$I_{\rm beam}(\vec{x}) = I_{\rm beam}(\vec{0}) e^{-\frac{1}{2}\vec{x}\cdot\mathbf{B}^{-1}\cdot\vec{x}}, \tag{A.16}$$

where

$$\mathbf{B} = \frac{1}{8\ln 2}\begin{pmatrix} B_{\rm M}^2 \sin^2\phi_{\rm B} + B_{\rm m}^2\cos^2\phi_{\rm B} & -(B_{\rm M}^2 - B_{\rm m}^2)\sin\phi_{\rm B}\cos\phi_{\rm B} \\ -(B_{\rm M}^2 - B_{\rm m}^2)\sin\phi_{\rm B}\cos\phi_{\rm B} & B_{\rm M}^2\cos^2\phi_{\rm B} + B_{\rm m}^2\sin^2\phi_{\rm B} \end{pmatrix}, \tag{A.17}$$

and where $B_{\rm M}$ is the beam major axis, $B_{\rm m}$ is the beam minor axis, and for x increasing to the west and y increasing to the north, $\phi_{\rm B}$ is the angle east of north to the major axis. Thus the covariance matrix for the pixel noise is

$$S^{\rm data}(\vec{x},\vec{y}) = S^{\rm data}(\vec{x}-\vec{y}) = \mu^2 e^{-\frac{1}{2}(\vec{x}-\vec{y})\cdot\mathbf{B}^{-1}\cdot(\vec{x}-\vec{y})}, \tag{A.18}$$

and its Fourier transform (defined in equation A.10) is

$$\tilde{S}^{\rm data}(\vec{k}) = \mu^2\sqrt{\det\mathbf{B}}\; e^{-\frac{1}{2}\vec{k}\cdot\mathbf{B}\cdot\vec{k}}, \tag{A.19}$$

where

$$\sqrt{\det\mathbf{B}} = \frac{B_{\rm M}B_{\rm m}}{8\ln 2}. \tag{A.20}$$

A.5 The fitted model: an elliptical Gaussian

The model to be fitted to the interferometry map is itself an elliptical Gaussian:

$$g(\vec{x}) = Ae^{-\frac{1}{2}(\vec{x}-\vec{x}_o)\cdot\mathbf{G}^{-1}\cdot(\vec{x}-\vec{x}_o)}, \tag{A.21}$$

where

$$\mathbf{G} = \frac{1}{8\ln 2}\begin{pmatrix} \theta_{\rm M}^2 \sin^2\phi_{\rm G} + \theta_{\rm m}^2\cos^2\phi_{\rm G} & -(\theta_{\rm M}^2 - \theta_{\rm m}^2)\sin\phi_{\rm G}\cos\phi_{\rm G} \\ -(\theta_{\rm M}^2 - \theta_{\rm m}^2)\sin\phi_{\rm G}\cos\phi_{\rm G} & \theta_{\rm M}^2\cos^2\phi_{\rm G} + \theta_{\rm m}^2\sin^2\phi_{\rm G} \end{pmatrix}, \tag{A.22}$$

where the six model parameters are

$$\vec{x}_o \equiv (X, Y) = \text{Location of Gaussian to west } (X) \text{ and north } (Y) \quad \text{(A.23)}$$

$$\phi_{\rm G} = \text{Angle east of north to major axis} \quad \text{(A.24)}$$

$$\theta_{\rm M} = \text{Model major axis} \quad \text{(A.25)}$$

$$\theta_{\rm m} = \text{Model minor axis} \quad \text{(A.26)}$$

$$A = \text{Amplitude of Gaussian model} \quad \text{(A.27)}$$

The Fourier transform of this model is

$$\tilde{g}(\vec{k}) = A\, e^{-i\vec{k}\cdot\vec{x}_o} \sqrt{\det \mathbf{G}}\; e^{-\frac{1}{2}\vec{k}\cdot\mathbf{G}\cdot\vec{k}}, \quad \text{(A.28)}$$

where

$$\sqrt{\det \mathbf{G}} = \frac{\theta_{\rm M}\theta_{\rm m}}{8 \ln 2}. \quad \text{(A.29)}$$

A.6 The inverse of the covariance matrix computed for an elliptical Gaussian fitted to an interferometer map with pixel-to-pixel noise correlations

Expression A.14 for the inverse of the covariance matrix requires the Fourier transforms of the model and of the pixel-noise covariance-matrix. These are given in equations A.28 and A.19. Plugging equations A.28 and A.19 into equation A.14, and doing rather a lot of derivatives, integrals, and matrix algebra, one can grind out the expressions for the components $D_{\alpha\beta}$ of the inverse of the model-parameter covariance matrix. The indices α and β run over the model parameters, referring to $X, Y, \phi_{\rm G}, \theta_{\rm M}, \theta_{\rm m}$, and A respectively for $\alpha = 1, \ldots, 6$.

It is found that $\mathbf{D}$ is block diagonal; the position errors are uncorrelated with the

errors in the other fitted model parameters:

$$\mathbf{D} = \left(\begin{array}{cc|cccc} & & 0 & 0 & 0 & 0 \\ \multicolumn{2}{c|}{\mathbf{D}_{\rm pos}} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & & & & \\ 0 & 0 & \multicolumn{4}{c}{\multirow{3}{*}{$\mathbf{D}_{\rm size}$}} \\ 0 & 0 & & & & \\ 0 & 0 & & & & \end{array}\right) \begin{array}{l} \leftarrow X \\ \leftarrow Y \\ \leftarrow \phi_{\rm G} \\ \leftarrow \theta_{\rm M} \\ \leftarrow \theta_{\rm m} \\ \leftarrow A \end{array} \tag{A.30}$$

The matrices $\mathbf{D}_{\rm pos}$ and $\mathbf{D}_{\rm size}$ will be given below. For writing the elements of these matrices, it will be useful to introduce some new notation. The rotation matrix is

$$\mathbf{R} = \begin{pmatrix} -\sin\phi_{\rm G} & -\cos\phi_{\rm G} \\ +\cos\phi_{\rm G} & -\sin\phi_{\rm G} \end{pmatrix}. \tag{A.31}$$

Using the rotation matrix, the matrix describing the model may be rotated to a frame aligned with its axes,

$$\mathbf{G}^o = \mathbf{R}^{-1}\mathbf{G}\mathbf{R} = \frac{1}{8\ln 2}\begin{pmatrix} \theta_{\rm M}^2 & 0 \\ 0 & \theta_{\rm m}^2 \end{pmatrix}. \tag{A.32}$$

The following combination also appears: the model convolved with itself, from which the beam has been deconvolved,

$$\mathbf{G}' = 2\mathbf{G} - \mathbf{B}. \tag{A.33}$$

This same combination, in a frame aligned with the model axes, is

$$\mathbf{G}^{o\prime} = \mathbf{R}^{-1}\mathbf{G}'\mathbf{R} = \mathbf{R}^{-1}\left(2\mathbf{G} - \mathbf{B}\right)\mathbf{R} = 2\mathbf{G}^o - \mathbf{R}^{-1}\mathbf{B}\mathbf{R}. \tag{A.34}$$

The following quantities will also appear:

$$P = \det(\mathbf{G}'\mathbf{G}^{-1}) = \det(\mathbf{2} - \mathbf{B}\mathbf{G}^{-1}), \tag{A.35}$$

$$Q = \mathrm{Tr}\,(\mathbf{G}'\mathbf{G}^{-1}) - 1 = 3 - \mathrm{Tr}\,(\mathbf{B}\mathbf{G}^{-1}). \tag{A.36}$$

Using the quantities defined above the 2×2 position-error portion of the inverse covariance matrix (A.30) is

$$\mathbf{D}_{\mathrm{pos}} = \frac{A^2}{\mu^2} \frac{\det \mathbf{G}}{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}} \, (\mathbf{G}')^{-1}, \tag{A.37}$$

$$= \frac{A^2}{\mu^2} \frac{\det \mathbf{G}}{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}} \, (2\mathbf{G} - \mathbf{B})^{-1}, \tag{A.38}$$

and the 4×4 angle-axes-amplitude portion of the inverse covariance matrix (A.30) is

$$\mathbf{D}_{\mathrm{size}} = \frac{A^2}{\mu^2} \frac{\det \mathbf{G}}{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}} \begin{pmatrix} \frac{\theta_{\mathrm{M}}^2 - \theta_{\mathrm{m}}^2}{8 \ln 2} & 0 & 0 & 0 \\ 0 & \frac{\theta_{\mathrm{M}}}{8 \ln 2} & 0 & 0 \\ 0 & 0 & \frac{\theta_{\mathrm{m}}}{8 \ln 2} & 0 \\ 0 & 0 & 0 & \frac{1}{A} \end{pmatrix} \times$$

$$\begin{pmatrix} \Gamma'_{11}\Gamma'_{22} + 2[\Gamma'_{12}]^2 & \Gamma'_{12}(3\Gamma'_{11} - \Gamma_{11}) & \Gamma'_{12}(3\Gamma'_{22} - \Gamma_{22}) & -\Gamma'_{12} \\ \Gamma'_{12}(3\Gamma'_{11} - \Gamma_{11}) & (\Gamma'_{11} - \Gamma_{11})^2 + 2[\Gamma'_{11}]^2 & (\Gamma'_{11} - \Gamma_{11})(\Gamma'_{22} - \Gamma_{22}) + 2[\Gamma'_{12}]^2 & -\Gamma'_{11} + \Gamma_{11} \\ \Gamma'_{12}(3\Gamma'_{22} - \Gamma_{22}) & (\Gamma'_{11} - \Gamma_{11})(\Gamma'_{22} - \Gamma_{22}) + 2[\Gamma'_{12}]^2 & (\Gamma'_{22} - \Gamma_{22})^2 + 2[\Gamma'_{22}]^2 & -\Gamma'_{22} + \Gamma_{22} \\ -\Gamma'_{12} & -\Gamma'_{11} + \Gamma_{11} & -\Gamma'_{22} + \Gamma_{22} & 1 \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{\theta_{\mathrm{M}}^2 - \theta_{\mathrm{m}}^2}{8 \ln 2} & 0 & 0 & 0 \\ 0 & \frac{\theta_{\mathrm{M}}}{8 \ln 2} & 0 & 0 \\ 0 & 0 & \frac{\theta_{\mathrm{m}}}{8 \ln 2} & 0 \\ 0 & 0 & 0 & \frac{1}{A} \end{pmatrix}, \tag{A.39}$$

where the 2×2 matrixes $\mathbf{\Gamma}'$ and $\mathbf{\Gamma}$ are $\mathbf{\Gamma}' \equiv (\mathbf{G}^{o\prime})^{-1}$ and $\mathbf{\Gamma} \equiv (\mathbf{G}^{o})^{-1}$. To find the covariance matrix $\mathbf{D}^{-1}$, both $\mathbf{D}_{\mathrm{pos}}$ and $\mathbf{D}_{\mathrm{size}}$ must be inverted.

A.7 The covariance matrix computed for an elliptical Gaussian fitted to an interferometer map with pixel-to-pixel noise correlations

When the above equation is inverted, the covariance matrix of the fitted model parameters is found to be

$$\mathbf{D}^{-1} = \left(\begin{array}{c|cccc} \multicolumn{1}{c|}{\multirow{2}{*}{$\mathbf{D}_{\rm pos}^{-1}$}} & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \multicolumn{3}{c}{\multirow{4}{*}{$\mathbf{D}_{\rm size}^{-1}$}} \\ 0 & 0 & & & \\ 0 & 0 & & & \\ 0 & 0 & & & \end{array}\right) \begin{array}{l} \leftarrow X \\ \leftarrow Y \\ \leftarrow \phi_{\rm G} \\ \leftarrow \theta_{\rm M} \\ \leftarrow \theta_{\rm m} \\ \leftarrow A \end{array} \tag{A.40}$$

where the 2×2 covariance matrix for the position errors is

$$\mathbf{D}_{\rm pos}^{-1} = \frac{\mu^2}{A^2} \frac{\sqrt{\det \mathbf{B}} \sqrt{\det \mathbf{G}'}}{\det \mathbf{G}} \mathbf{G}', \tag{A.41}$$

$$= \frac{\mu^2}{A^2} \frac{\sqrt{\det \mathbf{B}} \sqrt{\det \mathbf{G}'}}{\det \mathbf{G}} (2\mathbf{G} - \mathbf{B}), \tag{A.42}$$

and where the 4×4 covariance matrix for the angle-axes-flux errors is

$$\mathbf{D}_{\text{size}}^{-1} = \frac{\mu^2}{A^2} \frac{\sqrt{\det \mathbf{B}} \sqrt{\det \mathbf{G}'}}{\det \mathbf{G}} \begin{pmatrix} \frac{8 \ln 2}{\theta_{\text{M}}^2 - \theta_{\text{m}}^2} & 0 & 0 & 0 \\ 0 & \frac{8 \ln 2}{\theta_{\text{M}}} & 0 & 0 \\ 0 & 0 & \frac{8 \ln 2}{\theta_{\text{m}}} & 0 \\ 0 & 0 & 0 & A \end{pmatrix} \times$$

$$\begin{pmatrix} G_{11}^{o\prime} G_{22}^{o\prime} + [G_{12}^{o\prime}]^2 & G_{12}^{o\prime} G_{11}^{o\prime} & G_{12}^{o\prime} G_{22}^{o\prime} & -G_{12}^{o\prime} Q \\ G_{12}^{o\prime} G_{11}^{o\prime} & \frac{1}{2} [G_{11}^{o\prime}]^2 & \frac{1}{2} [G_{12}^{o\prime}]^2 & \frac{1}{2} (\underbrace{G_{11}^{o}}_{\substack{\text{no} \\ \text{prime}}} P - G_{11}^{o\prime} Q) \\ G_{12}^{o\prime} G_{22}^{o\prime} & \frac{1}{2} [G_{12}^{o\prime}]^2 & \frac{1}{2} [G_{22}^{o\prime}]^2 & \frac{1}{2} (\underbrace{G_{22}^{o}}_{\substack{\text{no} \\ \text{prime}}} P - G_{22}^{o\prime} Q) \\ -G_{12}^{o\prime} Q & \frac{1}{2} (\underbrace{G_{11}^{o}}_{\substack{\text{no} \\ \text{prime}}} P - G_{11}^{o\prime} Q) & \frac{1}{2} (\underbrace{G_{22}^{o}}_{\substack{\text{no} \\ \text{prime}}} P - G_{22}^{o\prime} Q) & \frac{1}{2} (3 + Q^2 - 2P) \end{pmatrix}$$

$$\times \begin{pmatrix} \frac{8 \ln 2}{\theta_{\text{M}}^2 - \theta_{\text{m}}^2} & 0 & 0 & 0 \\ 0 & \frac{8 \ln 2}{\theta_{\text{M}}} & 0 & 0 \\ 0 & 0 & \frac{8 \ln 2}{\theta_{\text{m}}} & 0 \\ 0 & 0 & 0 & A \end{pmatrix} \begin{matrix} \leftarrow \phi_{\text{G}} \\ \leftarrow \theta_{\text{M}} \\ \leftarrow \theta_{\text{m}} \\ \leftarrow A \end{matrix} \tag{A.43}$$

Equations A.41 and A.43 are the main results of this appendix: the covariance matrix for the fitted parameters of an elliptical Gaussian model fitted to a radio-interferometry map which has pixel-to-pixel noise correlated on the scale of the beam. The matrices $\mathbf{B}$, $\mathbf{G}$, $\mathbf{G}^o$, $\mathbf{G}'$, and $\mathbf{G}^{o\prime}$ are defined in equations A.17, A.22, A.32, A.33, and A.34, respectively, and the quantities P and Q are defined in equations A.35, and A.36.

A.8 Variance of the integral flux

The variances of the fitted elliptical Gaussian's orientation angle, major axis length, minor axis length, and peak flux are given by the diagonal elements in the array A.43. The integral flux of the fitted elliptical Gaussian can be computed from its major axis length, minor axis length, and peak flux,

$$I_{\text{flux}} = \frac{\pi \theta_{\text{M}} \theta_{\text{m}} A}{4 \ln 2}. \tag{A.44}$$

The variance of the integral flux can be computed using equation A.4 and the covariance matrix in equation A.43, assuming that the major axis length, minor axis length, and peak flux are all fitted parameters:

$$\frac{\text{Var}\left(I_{\text{flux}}\right)}{I_{\text{flux}}^2} = 2\frac{\mu^2}{A^2}\frac{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}}{\det \mathbf{G}}. \tag{A.45}$$

A.9 Special cases

The expressions above for the covariance matrix (A.40, A.41, A.43), and for the variance of the integral flux (A.45) simplify for special cases.

A.9.1 Model Gaussian fitted using fixed major axis, minor axis, and orientation angle

If the model major axis, minor axis, and orientation angle were held fixed when the model was fitted to the data, then the corresponding rows and columns drop out of the expression (A.30, A.37, A.39) for the inverse of the covariance matrix, and the inverse of the covariance matrix is

$$\mathbf{D} = \frac{A^2}{\mu^2}\frac{\det \mathbf{G}}{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}} \left(\begin{array}{cc|c} & & 0 \\ & (\mathbf{G}')^{-1} & 0 \\ \hline 0 & 0 & \frac{1}{A^2} \end{array}\right) \begin{array}{l} \\ \leftarrow X \\ \leftarrow Y \\ \leftarrow A \end{array} \tag{A.46}$$

The inverse of this is simply found, giving the covariance matrix

$$\mathbf{D}^{-1} = \frac{\mu^2}{A^2}\frac{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}}{\det \mathbf{G}} \left(\begin{array}{cc|c} & & 0 \\ & \mathbf{G}' & 0 \\ \hline 0 & 0 & A^2 \end{array}\right) \begin{array}{l} \leftarrow X \\ \leftarrow Y \\ \leftarrow A \end{array} \tag{A.47}$$

In this case the variances of the peak flux and of the integral flux are both given by

$$\left.\frac{\mathrm{Var}\,(A)}{A^2}\right|_{\theta_\mathrm{M},\theta_\mathrm{m}\text{ fixed}} = \left.\frac{\mathrm{Var}\,(I_\mathrm{flux})}{I_\mathrm{flux}^2}\right|_{\theta_\mathrm{M},\theta_\mathrm{m}\text{ fixed}} = \frac{\mu^2}{A^2}\frac{\sqrt{\det \mathbf{B}}\sqrt{\det \mathbf{G}'}}{\det \mathbf{G}}. \tag{A.48}$$

Note that the variance of the integral flux is half as large as for the case in which the model major axis, minor axis, and orientation angle are fitted parameters (equation A.45).

A.9.2 Beam equal in size to the fitted Gaussian

If the fitted Gaussian is equal in size to the beam, the fitted values of the major axis, minor axis, peak flux, and orientation angle are uncorrelated with each other.

$$\mathbf{D}^{-1} = \frac{\mu^2}{A^2}\left(\begin{array}{cc|cccc} & & 0 & 0 & 0 & 0 \\ & \mathbf{G} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \frac{\theta_\mathrm{M}^2\theta_\mathrm{m}^2}{(\theta_\mathrm{M}^2-\theta_\mathrm{m}^2)^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2}\theta_\mathrm{M}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}\theta_\mathrm{m}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & A^2 \end{array}\right). \tag{A.49}$$

In the case where the major axis, minor axis, and orientation angle have been fitted, and their fitted values match the beam size, then the variance of the integral flux is

$$\frac{\mathrm{Var}\,(I_\mathrm{flux})}{I_\mathrm{flux}^2} = 2\frac{\mu^2}{A^2}. \tag{A.50}$$

If instead, the model major axis and minor axes and orientation angle had been fixed to the beam size, then the variance of the integral flux is

$$\left.\frac{\mathrm{Var}\,(I_{\mathrm{flux}})}{I_{\mathrm{flux}}^2}\right|_{\theta_{\mathrm{M}},\theta_{\mathrm{m}}\ \mathrm{fixed}} = \frac{\mu^2}{A^2}, \tag{A.51}$$

which has the same fractional error as the peak flux.

A.9.3 Beam small with respect to the fitted Gaussian

In the case that the fitted elliptical Gaussian size is much larger than the beam size, the covariance matrix of the fitted model parameters is

$$\mathbf{D}^{-1} = 4\gamma\frac{\mu^2}{A^2}\left(\begin{array}{cc|cccc} \multicolumn{2}{c|}{\multirow{2}{*}{\mathbf{G}}} & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \frac{2\theta_{\mathrm{M}}^2\theta_{\mathrm{m}}^2}{(\theta_{\mathrm{M}}^2-\theta_{\mathrm{m}}^2)^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{\mathrm{M}}^2 & 0 & -\frac{1}{2}A\theta_{\mathrm{M}} \\ 0 & 0 & 0 & 0 & \theta_{\mathrm{m}}^2 & -\frac{1}{2}A\theta_{\mathrm{m}} \\ 0 & 0 & 0 & -\frac{1}{2}A\theta_{\mathrm{M}} & -\frac{1}{2}A\theta_{\mathrm{m}} & A^2 \end{array}\right), \tag{A.52}$$

and the variance of the integral flux is

$$\frac{\mathrm{Var}\,(I_{\mathrm{flux}})}{I_{\mathrm{flux}}^2} = 4\gamma\frac{\mu^2}{A^2}, \tag{A.53}$$

where γ is the ratio of the beam area to the fitted Gaussian's area:

$$\gamma = \frac{\pi B_{\mathrm{M}} B_{\mathrm{m}}/(4\ln 2)}{\pi\theta_{\mathrm{M}}\theta_{\mathrm{m}}/(4\ln 2)}. \tag{A.54}$$

If the beam is very small with respect to the model extent, then the region over which the noise is correlated (which is the beam size) is very small with respect to the model extent. In that case, the covariance matrix should be the same as the covariance matrix for uncorrelated noise. It can be shown that equation A.52 gives the same covariance matrix as is found for the case of no pixel-to-pixel noise correlations,

provided that the area of such pixels equals the beam size used for equation A.52:

$$h^2 = \frac{\pi B_{\rm M} B_{\rm m}}{4 \ln 2}. \tag{A.55}$$

The expression in A.52 combined with the effective pixel size in A.55 gives the same variances as calculated by Condon (1997) for model parameters fitted to maps with uncorrelated noise.

Appendix B

Parameters of Models fitted to MG J0414+0534

The fitted model parameters, for the modified Multipole-Taylor models fitted to MG J0414+0534, are shown graphically in figures B-1 through figure B-37 with the parameter values tabulated alongside each plot. Confidence ranges on the model parameters are given if they were calculated. An explanation of the symbols in the plots is given in figure 7-2

Dimensionless time delays for the models fitted to MG J0414+0534 are tabulated in tables B.1 – B.3. Time delay ratios for the models fitted to MG J0414+0534 are tabulated in tables B.4 – B.6

B.1 Two- or three-term models

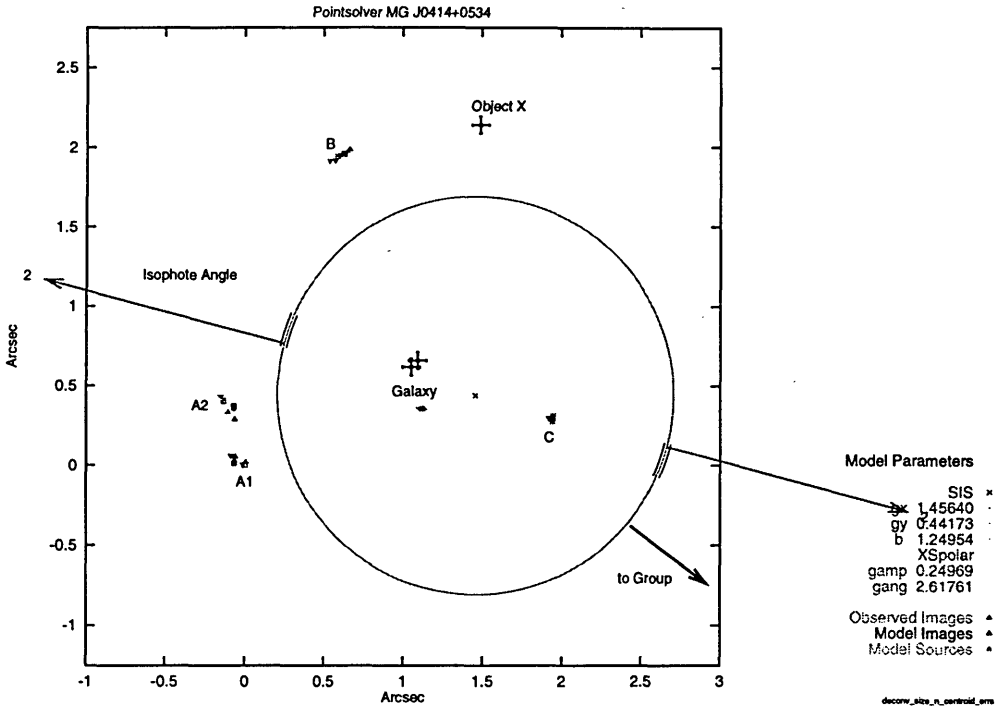

Figure B-1: Model SIS+XS (m0t1+m2ext), fitted to MG J0414+0534.

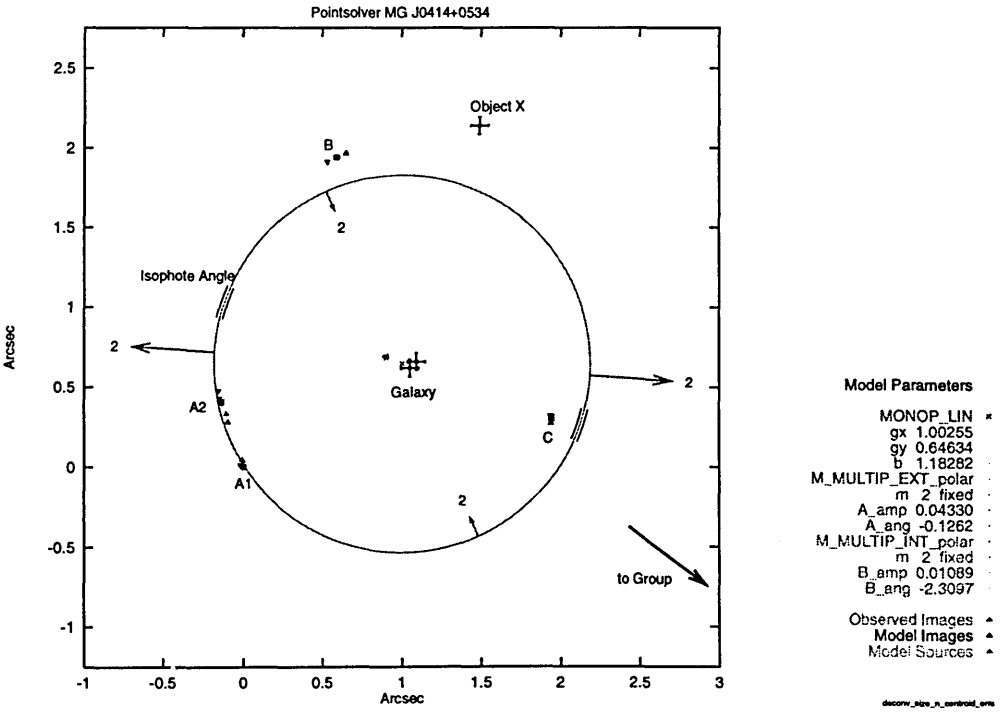

Figure B-2: Model m0t1+m2ext+m2int, fitted to MG J0414+0534.

B.2 Four-term models

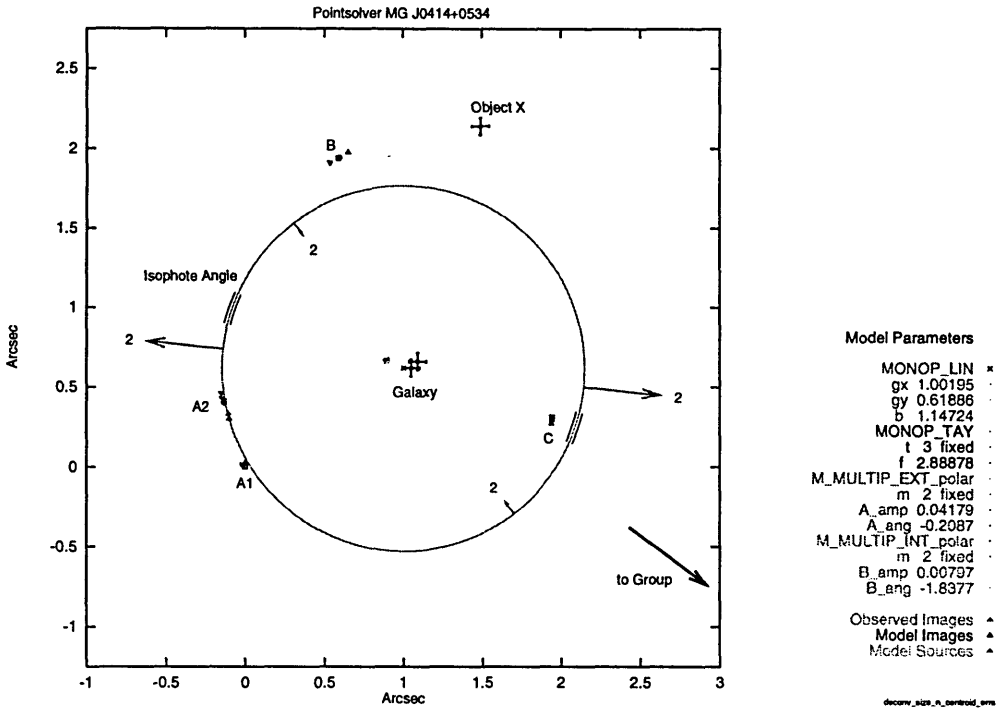

Figure B-3: Model m0t1+m0t3+m2ext+m2int, fitted to MG J0414+0534.

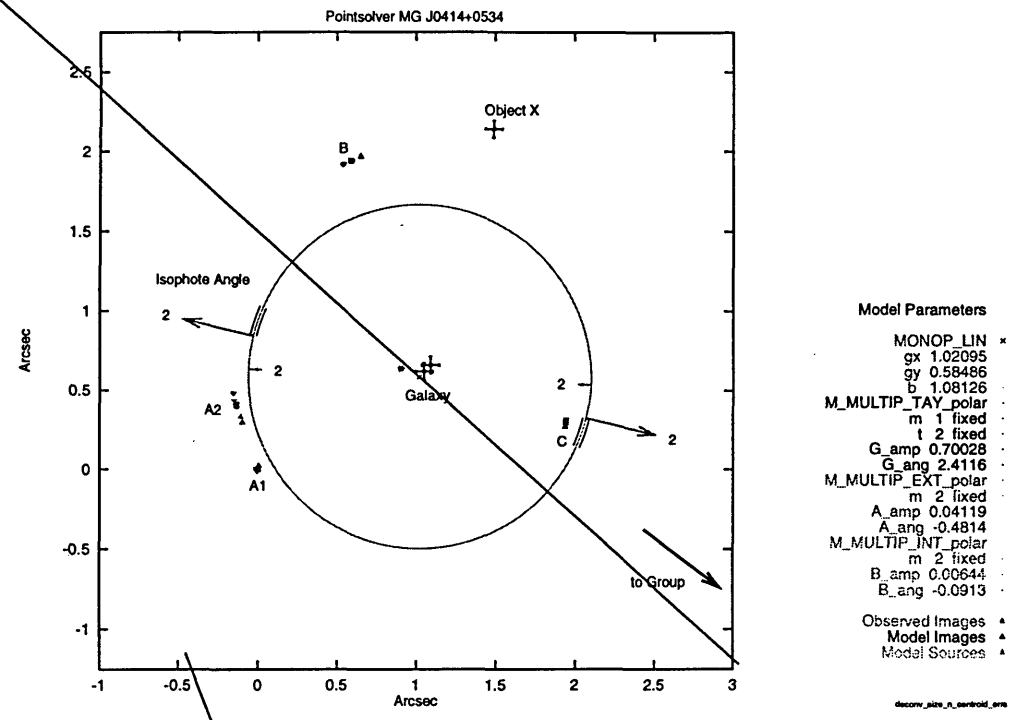

Figure B-4: Model m0t1+mm1t2+m2ext+m2int, fitted to MG J0414+0534.

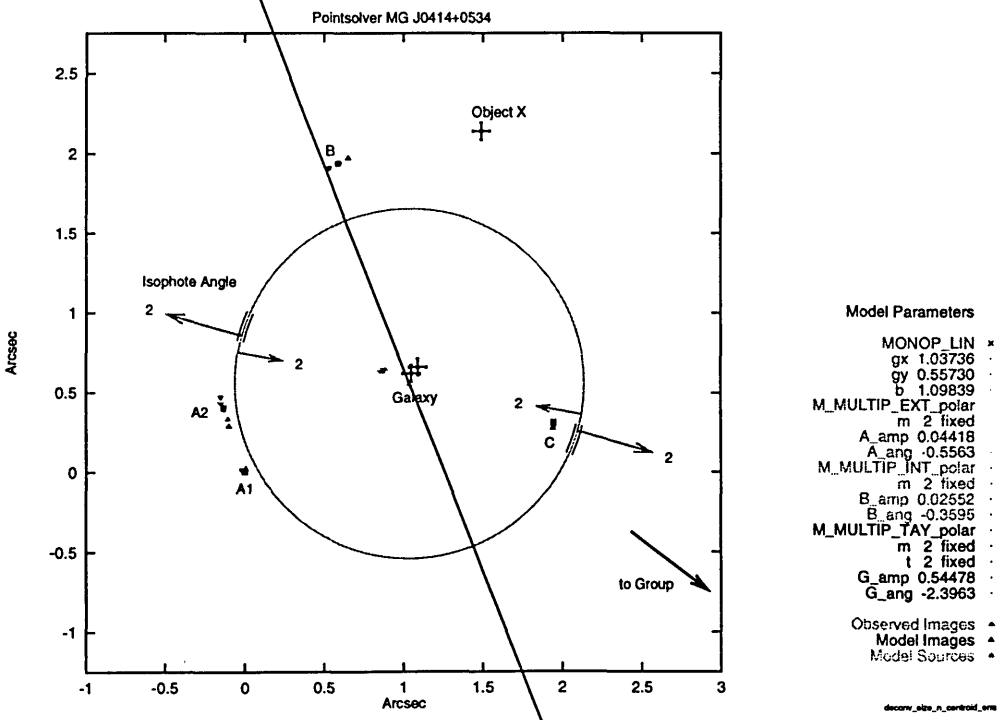

Figure B-5: Model m0t1+m2ext+m2int+mm2t2, fitted to MG J0414+0534.

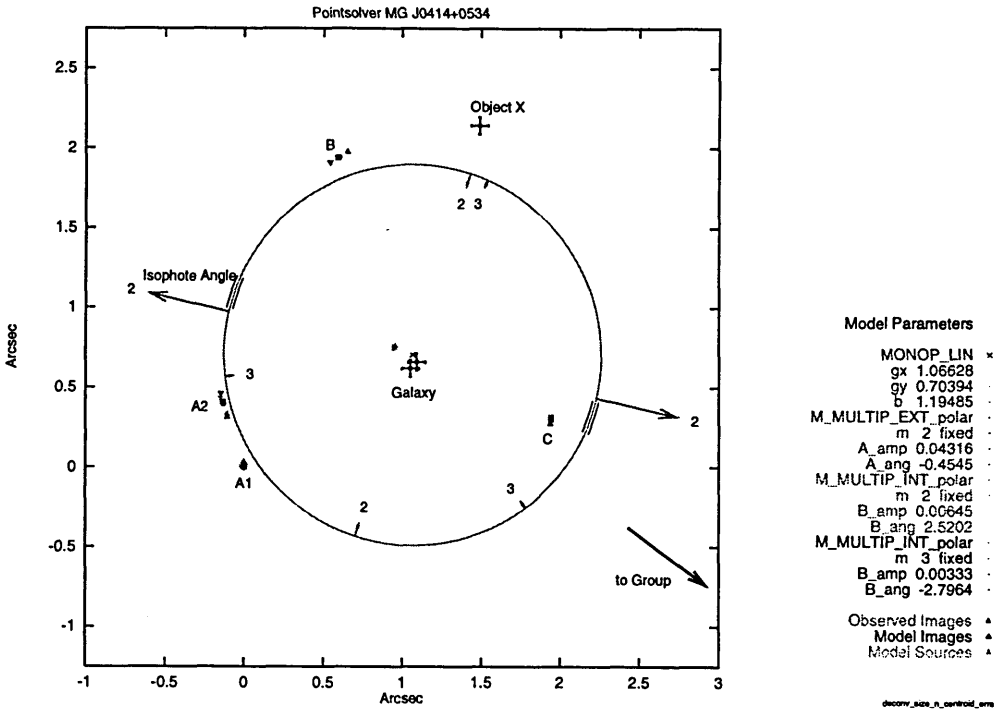

Figure B-6: Model m0t1+m2ext+m2int+m3int, fitted to MG J0414+0534.

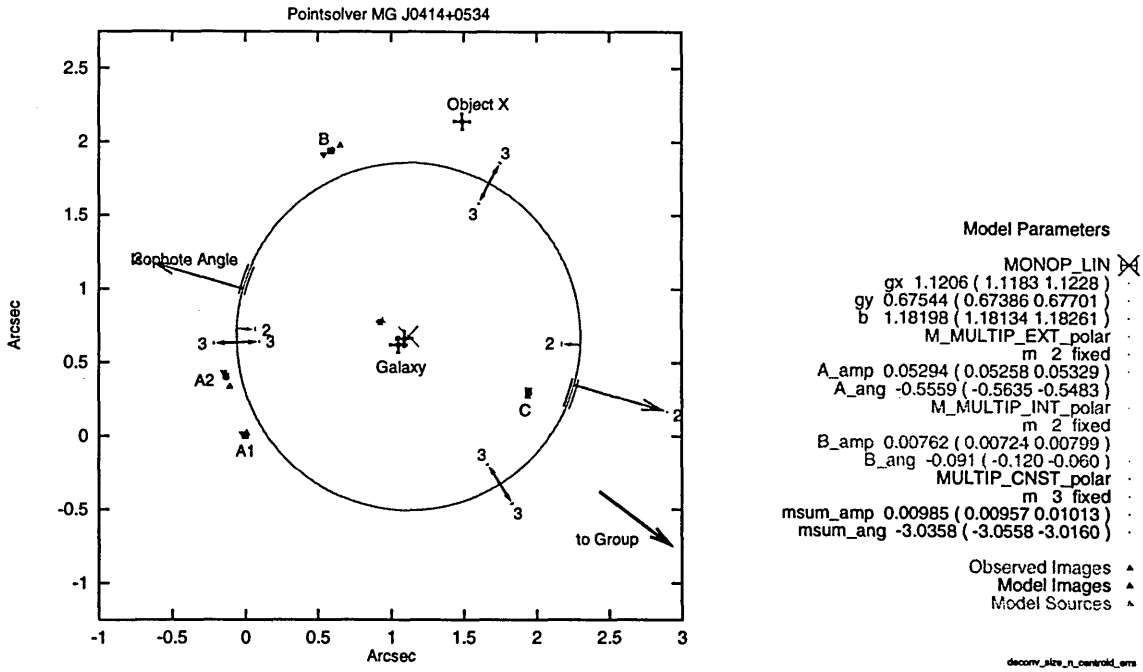

Figure B-7: Model m0t1+m2ext+m2int+m3t0, fitted to MG J0414+0534.

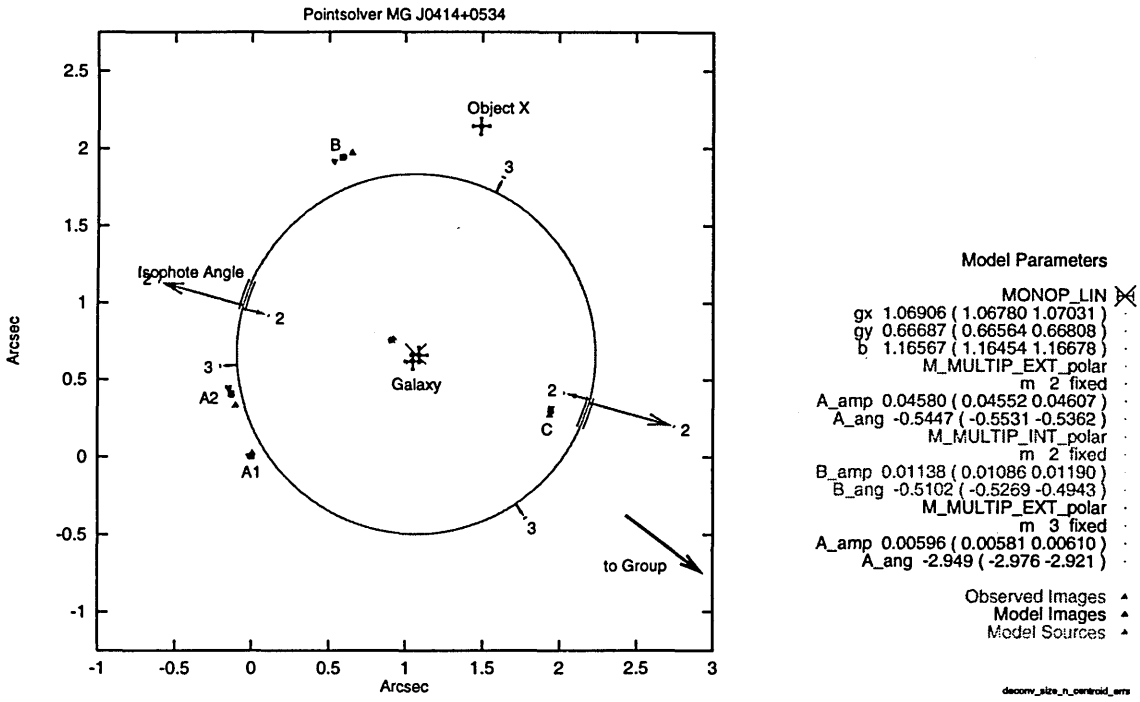

Figure B-8: Model m0t1+m2ext+m2int+m3ext, fitted to MG J0414+0534.

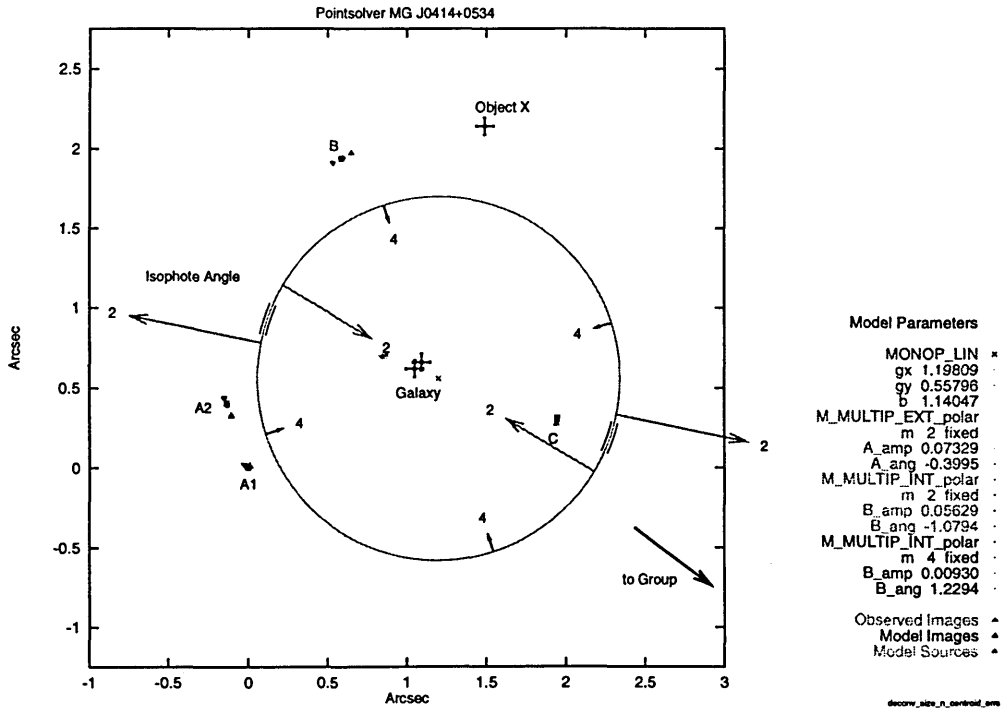

Figure B-9: Model m0t1+m2ext+m2int+m4int, fitted to MG J0414+0534.

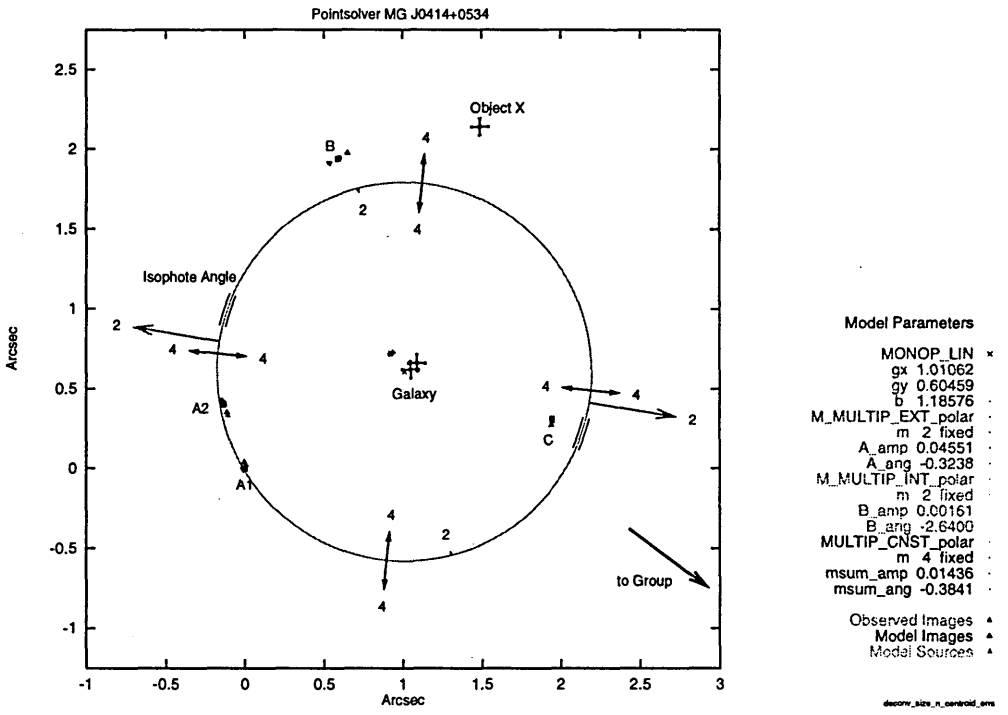

Figure B-10: Model m0t1+m2ext+m2int+m4t0, fitted to MG J0414+0534.

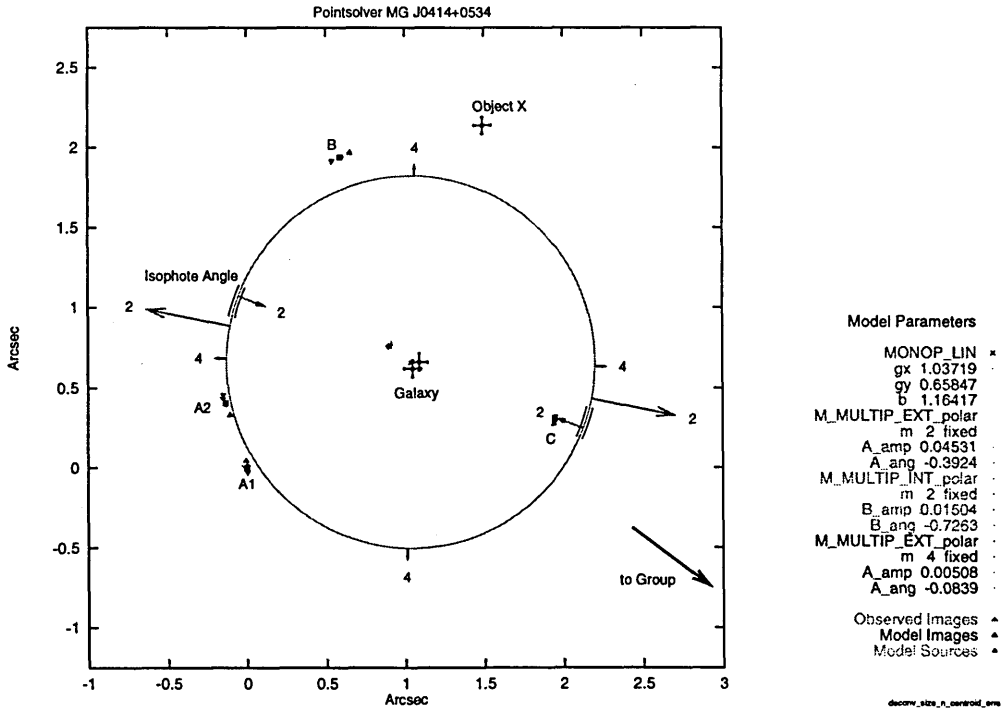

Figure B-11: Model m0t1+m2ext+m2int+m4ext, fitted to MG J0414+0534.

B.3 Five-term models

B.3.1 Models without an $m = 3$ multipole term

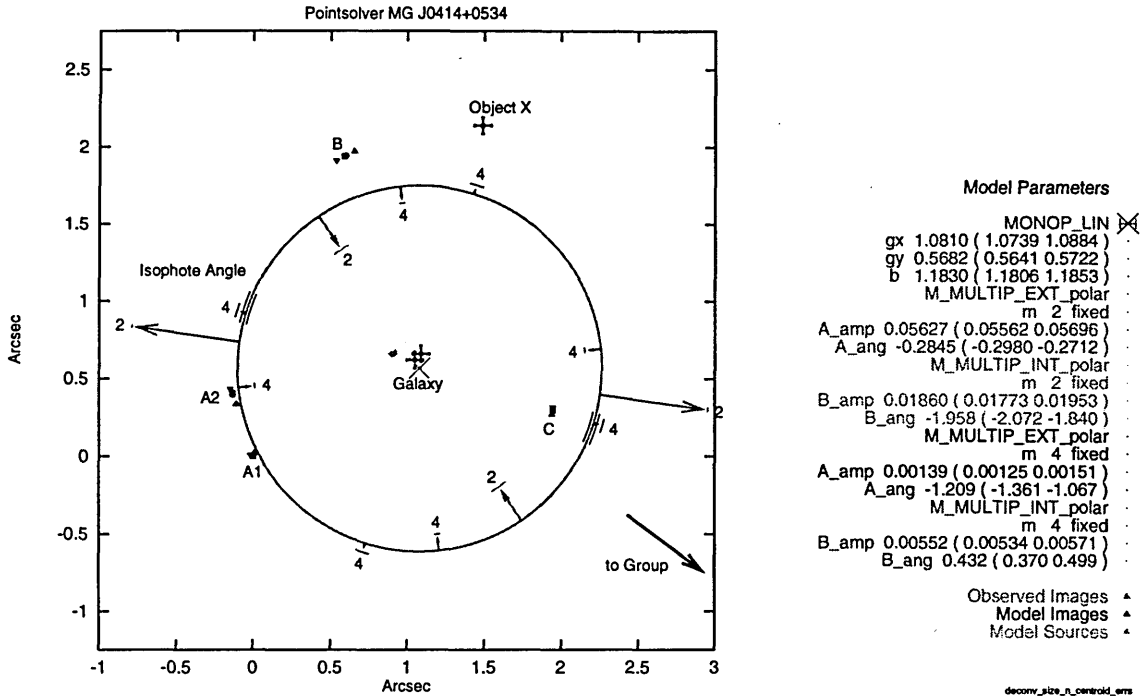

Figure B-12: Model m0t1+m2ext+m2int+m4ext+m4int.

B.3.2 Models with a mixed internal-external $m = 3$ multipole term

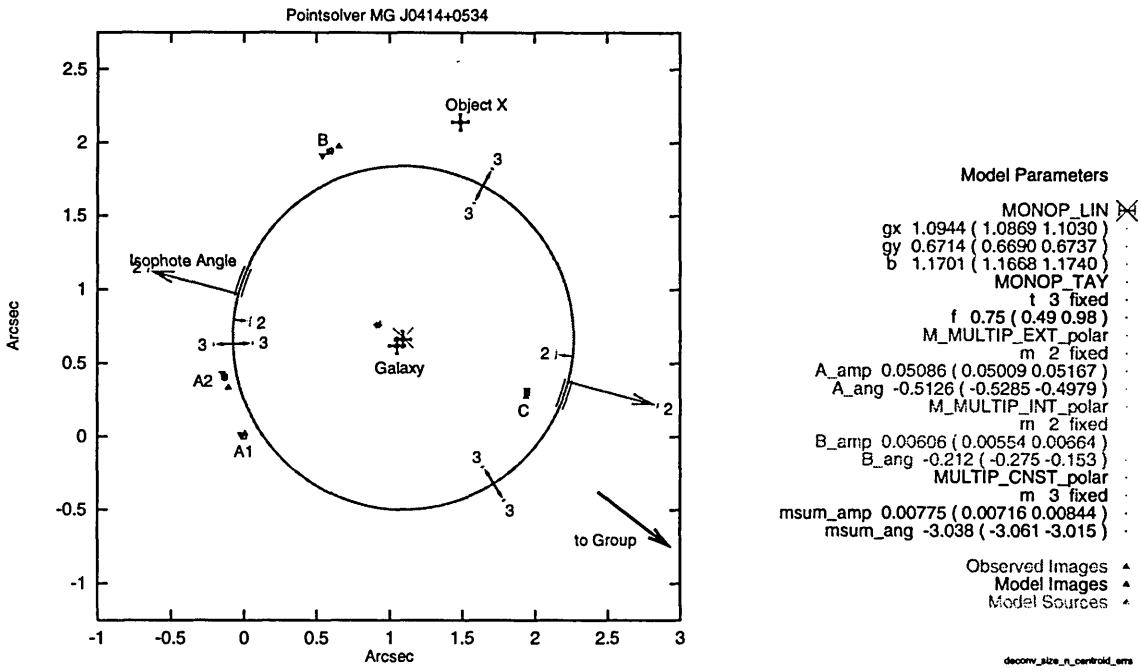

Figure B-13: Model m0t1+m0t3+m2ext+m2int+m3t0, fitted to MG J0414+0534.

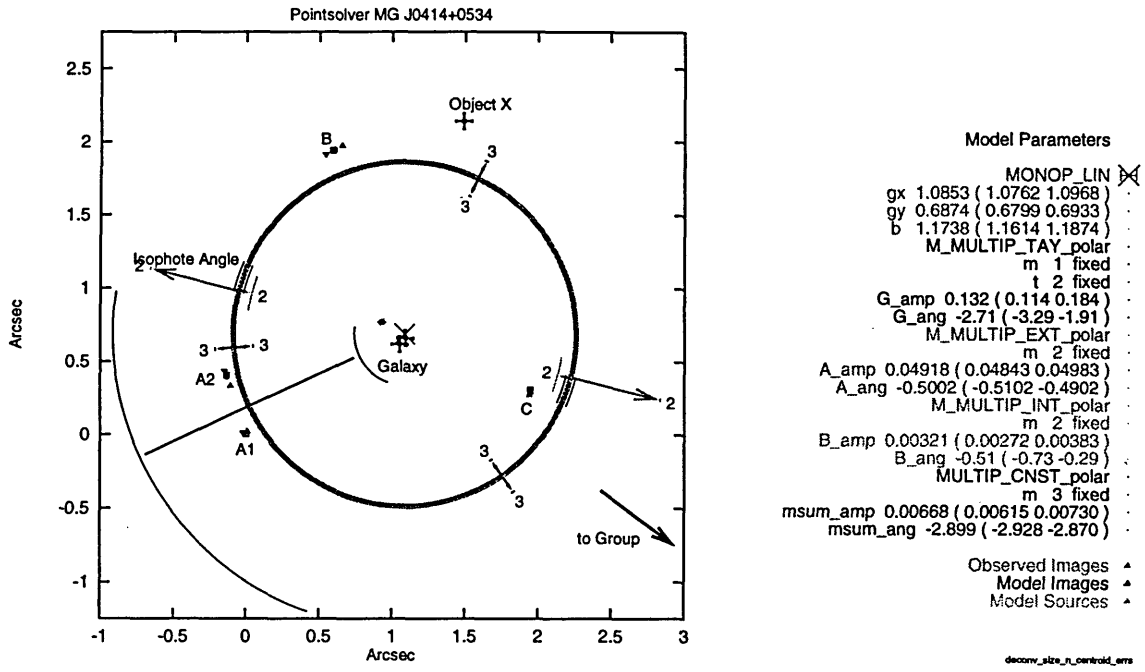

Figure B-14: Model m0t1+mm1t2+m2ext+m2int+m3t0.

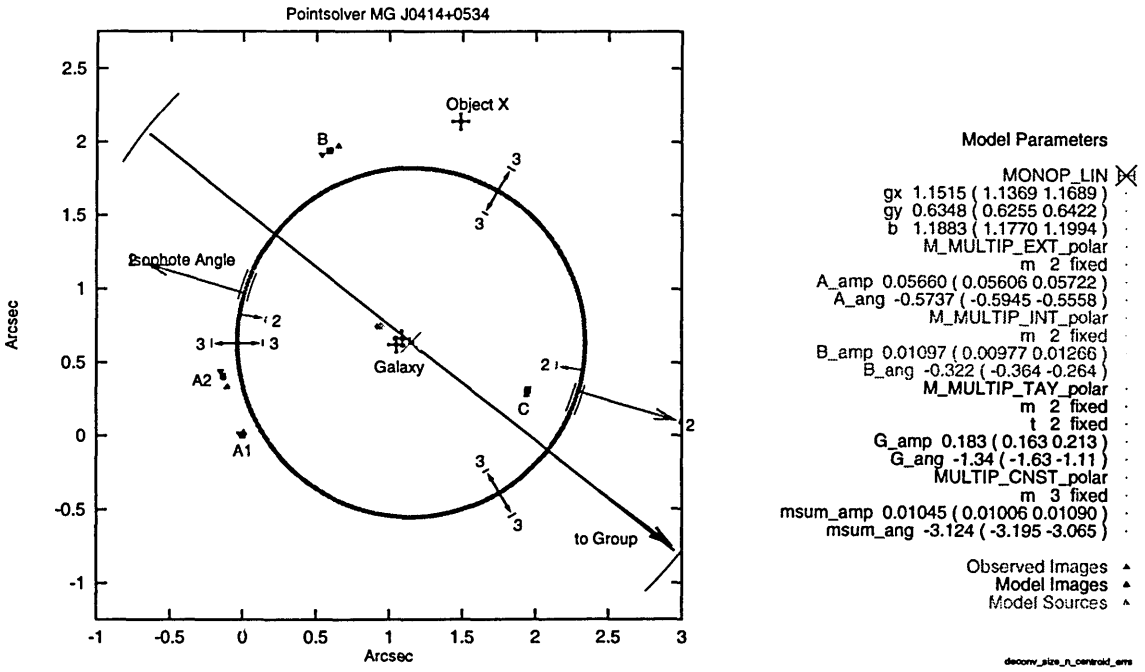

Figure B-15: Model m0t1+m2ext+m2int+mm2t2+m3t0.

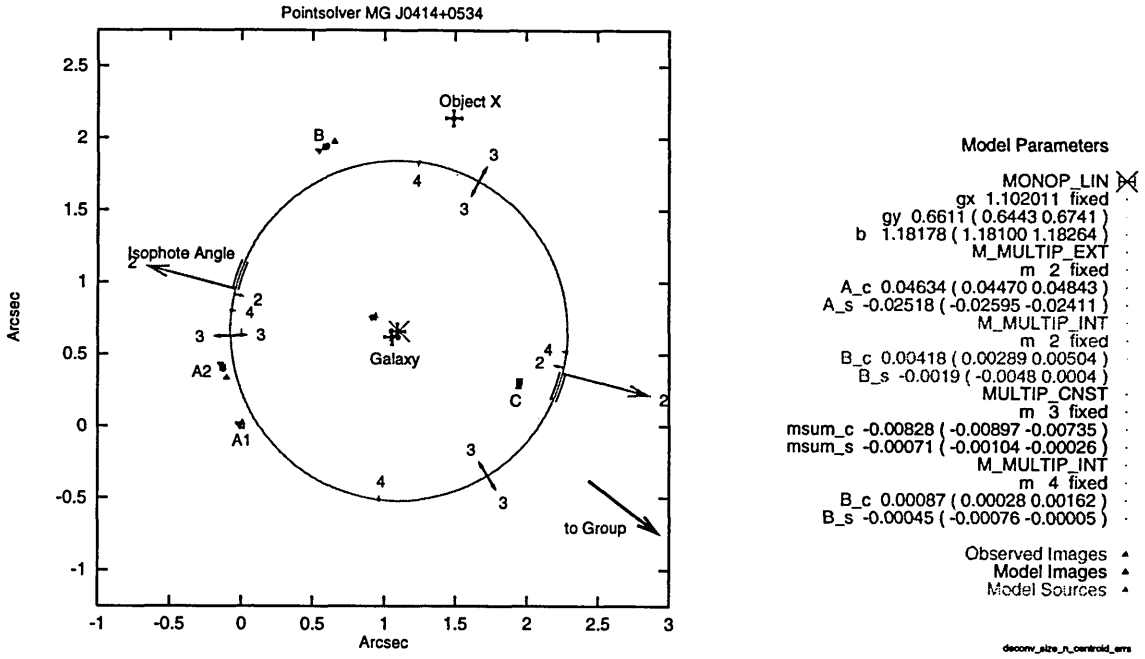

Figure B-16: Model m0t1+m2ext+m2int+m3t0+m4int, fitted to MG J0414+0534.

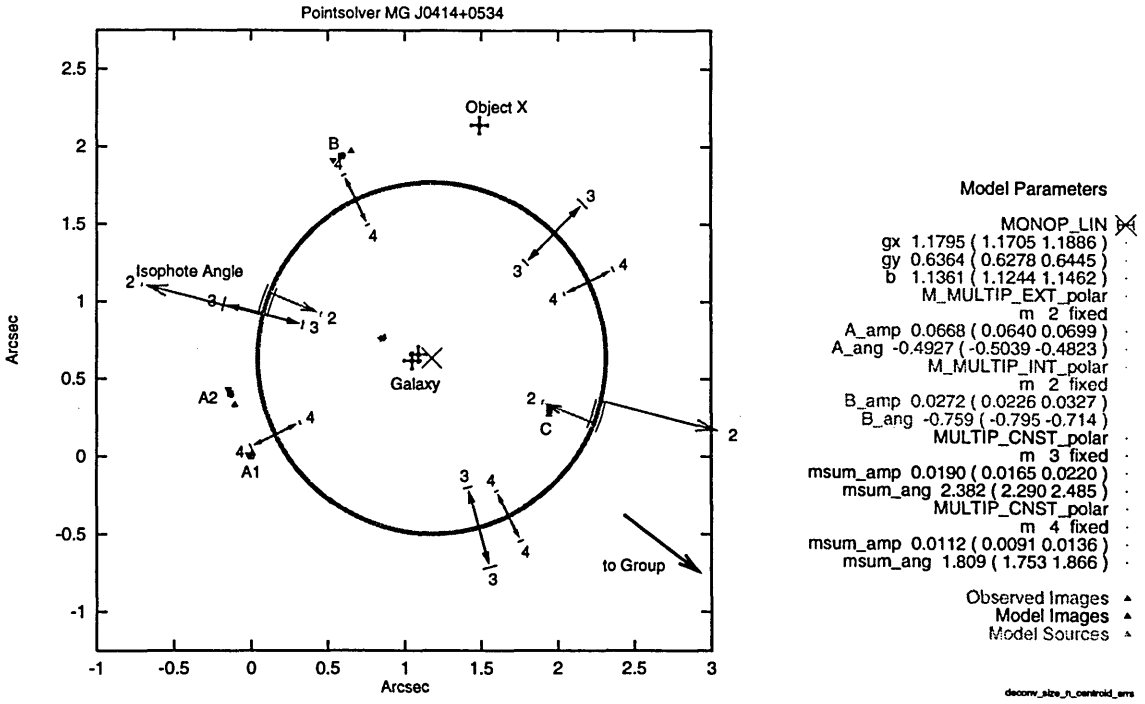

Figure B-17: Model m0t1+m2ext+m2int+m3t0+m4t0.

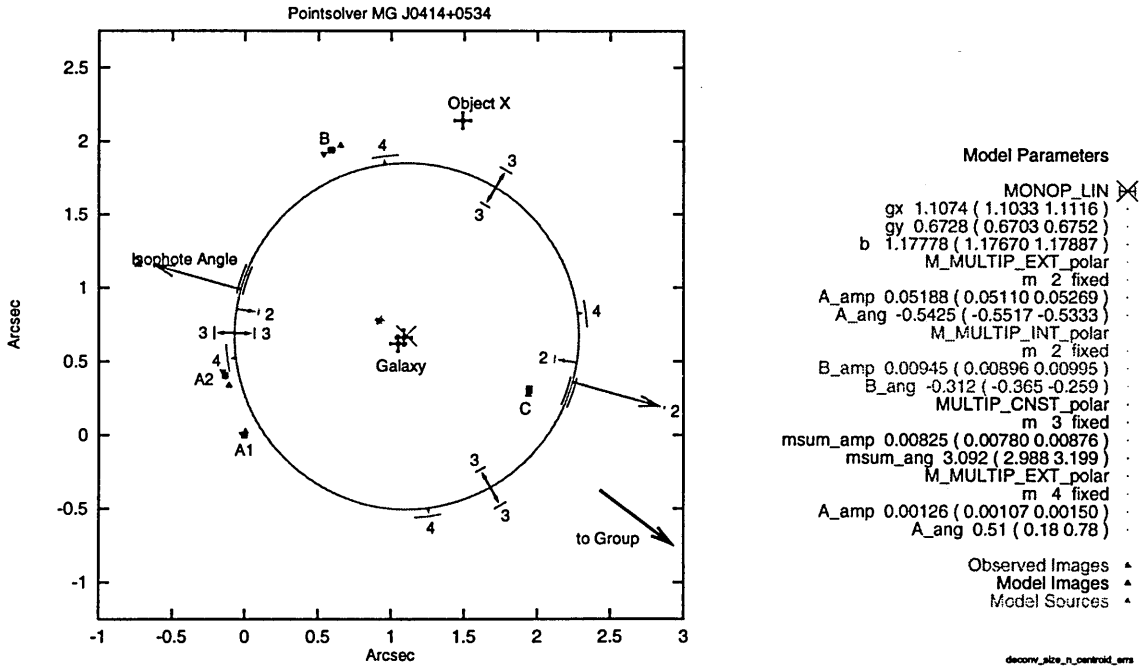

Figure B-18: Model m0t1+m2ext+m2int+m3t0+m4ext.

B.3.3 Models with an external $m = 3$ multipole term

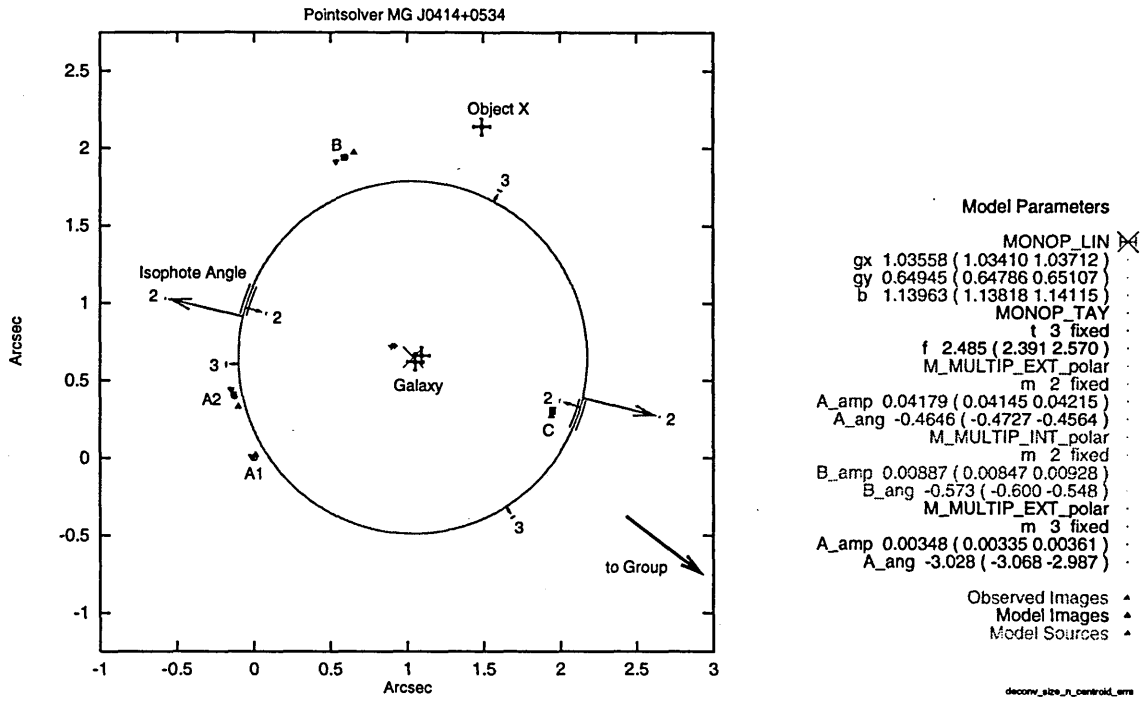

Figure B-19: Model m0t1+m0t3+m2ext+m2int+m3ext, fitted to MG J0414+0534.

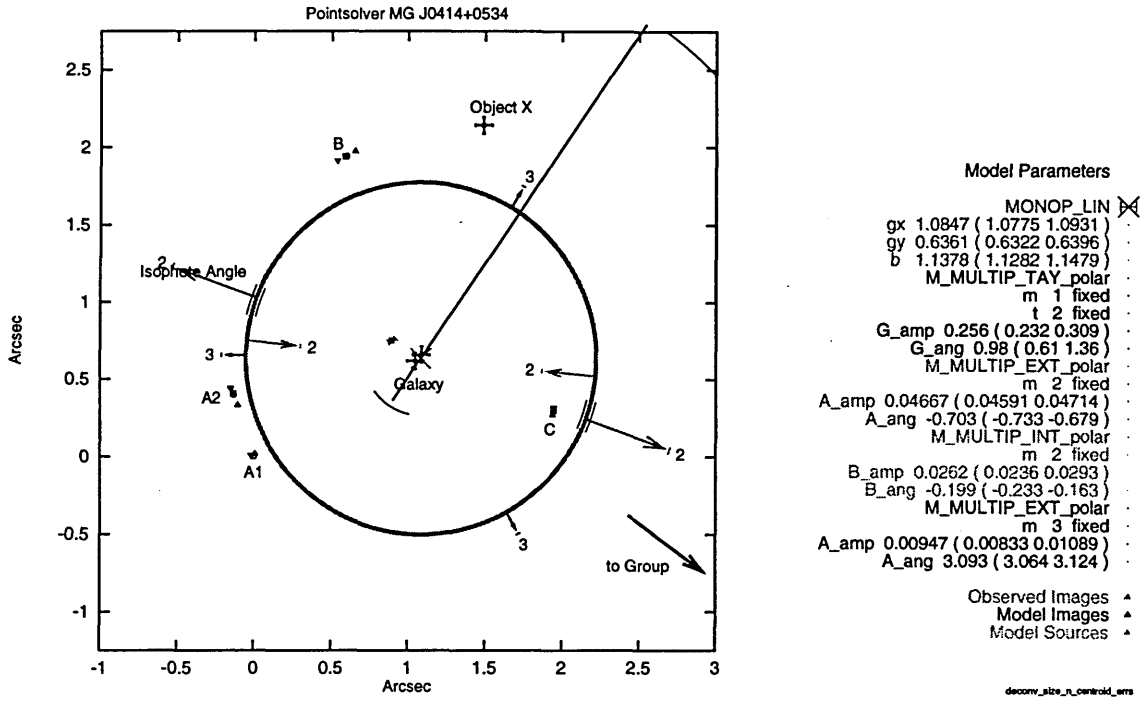

Figure B-20: Model m0t1+mm1t2+m2ext+m2int+m3ext.

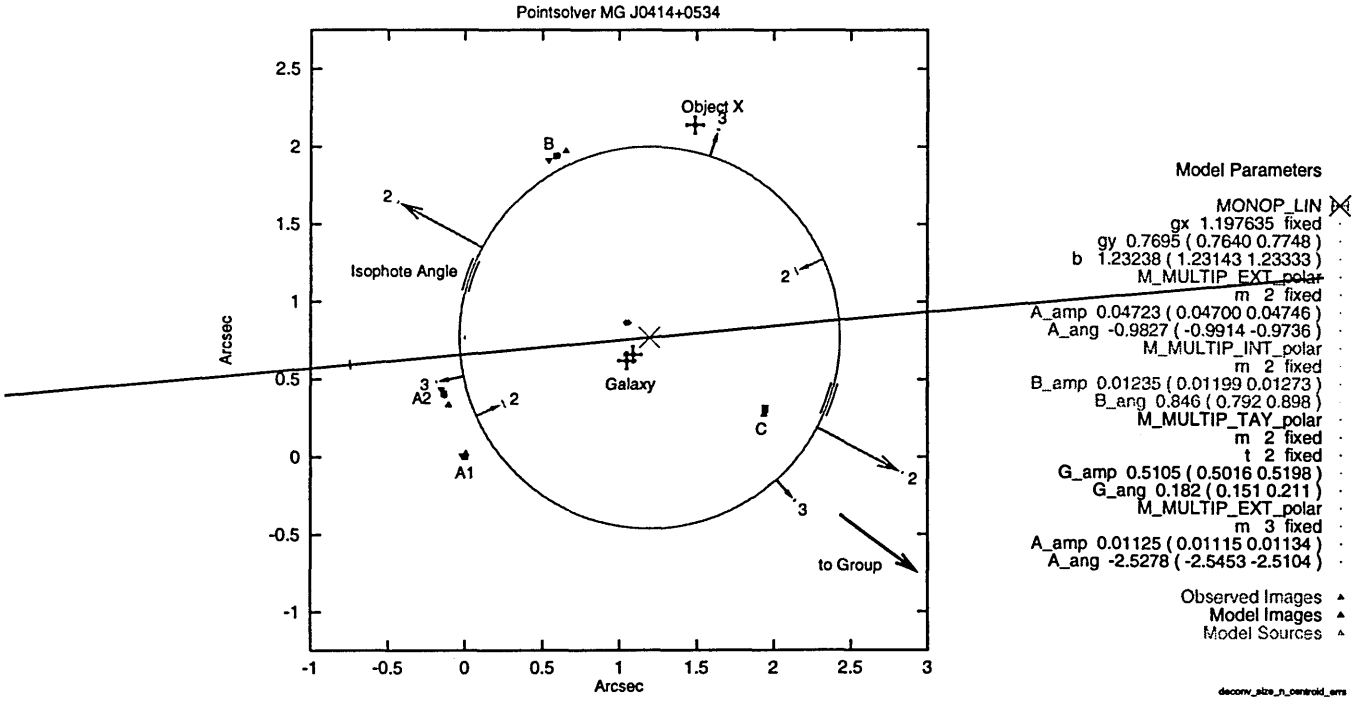

Figure B-21: Model m0t1+m2ext+m2int+mm2t2+m3ext.

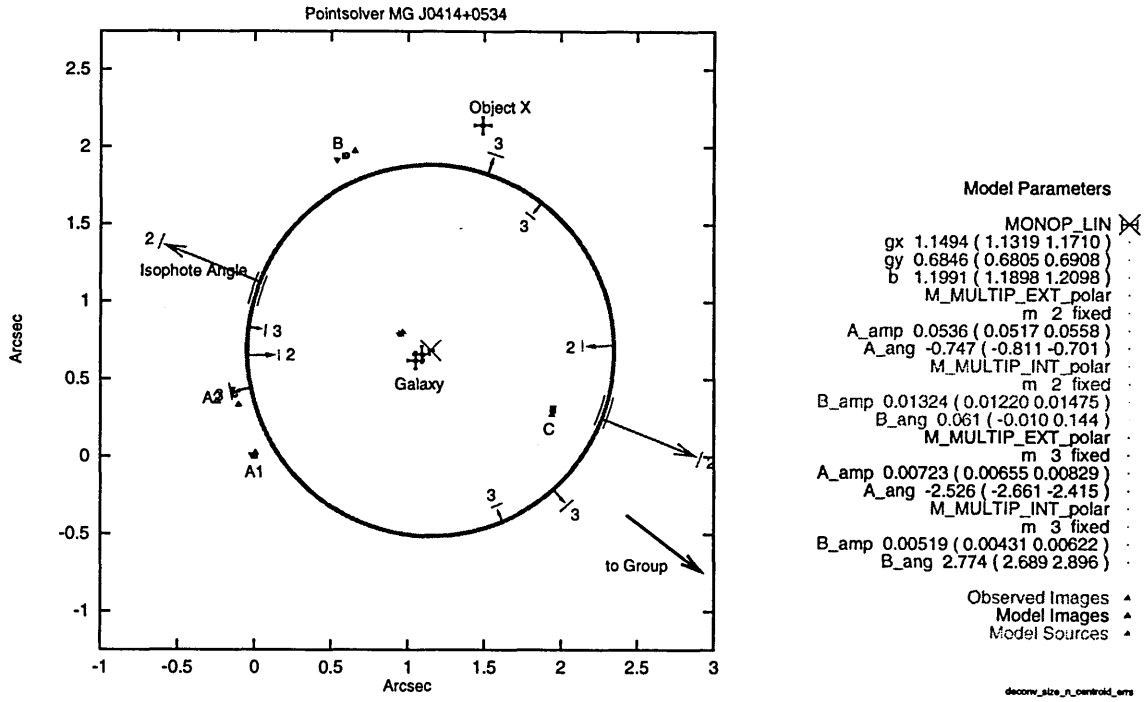

Figure B-22: Model m0t1+m2ext+m2int+m3ext+m3int.

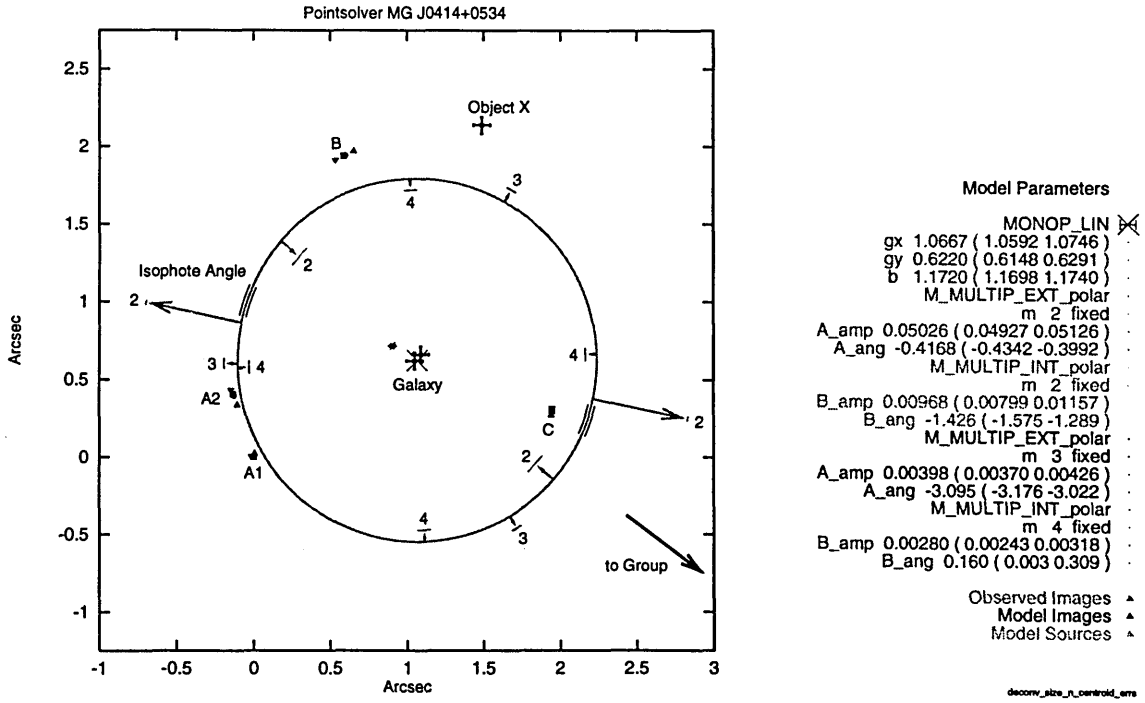

Figure B-23: Model m0t1+m2ext+m2int+m3ext+m4int.

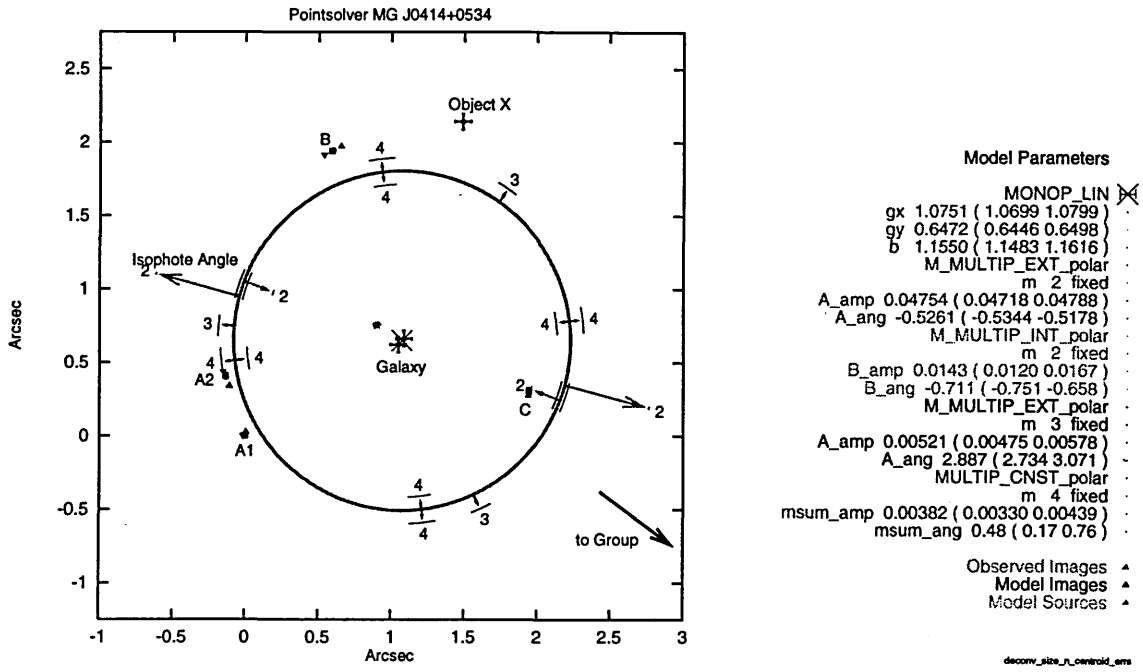

Figure B-24: Model m0t1+m2ext+m2int+m3ext+m4t0.

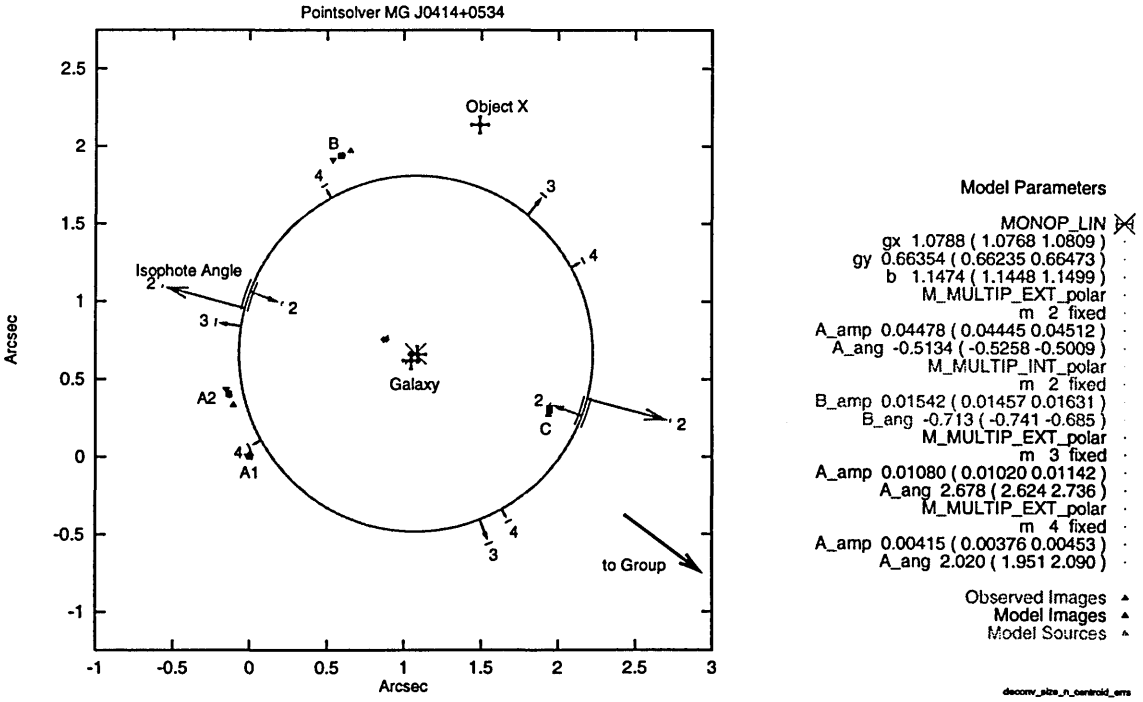

Figure B-25: Model m0t1+m2ext+m2int+m3ext+m4ext.

B.4 Six-term models

B.4.1 Models without an $m = 3$ multipole term

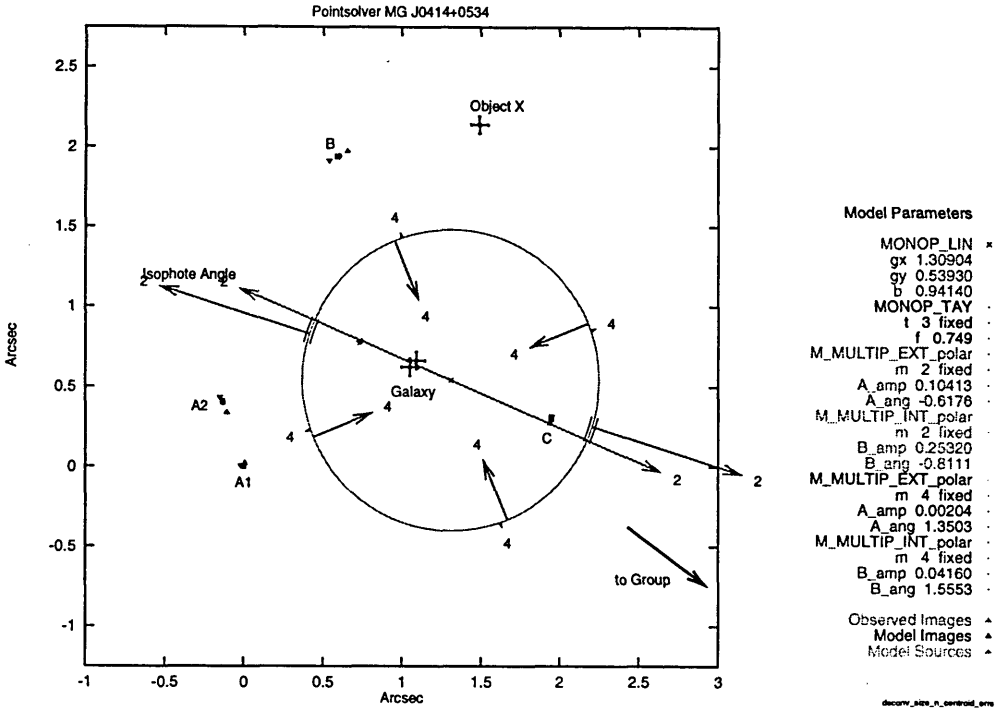

Figure B-26: Model m0t1+m0t3+m2ext+m2int+m4ext+m4int.

B.4.2 Models with a mixed internal-external $m = 3$ multipole term

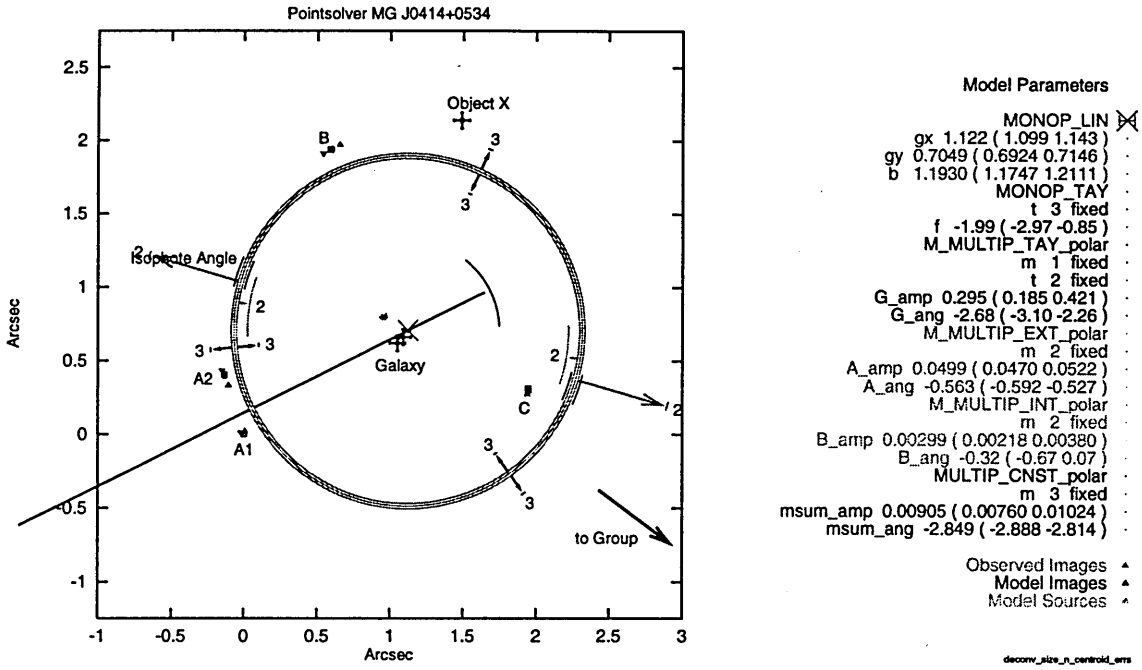

Figure B-27: Model m0t1+m0t3+mm1t2+m2ext+m2int+m3t0.

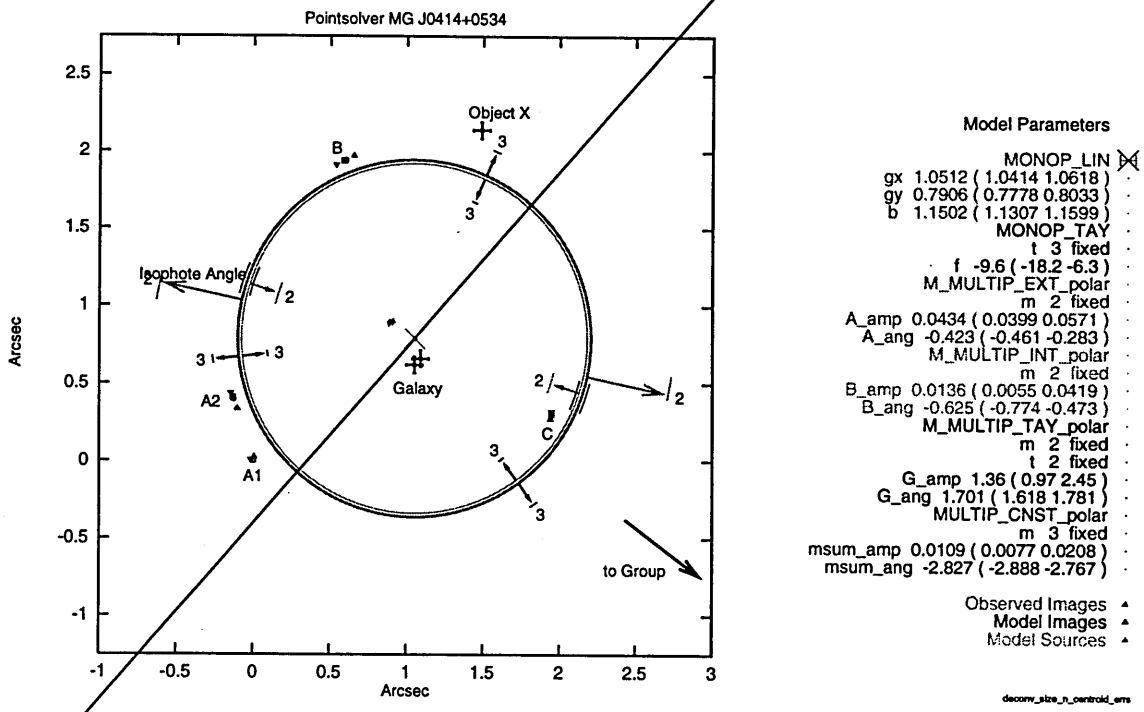

Figure B-28: Model m0t1+m0t3+m2ext+m2int+mm2t2+m3t0.

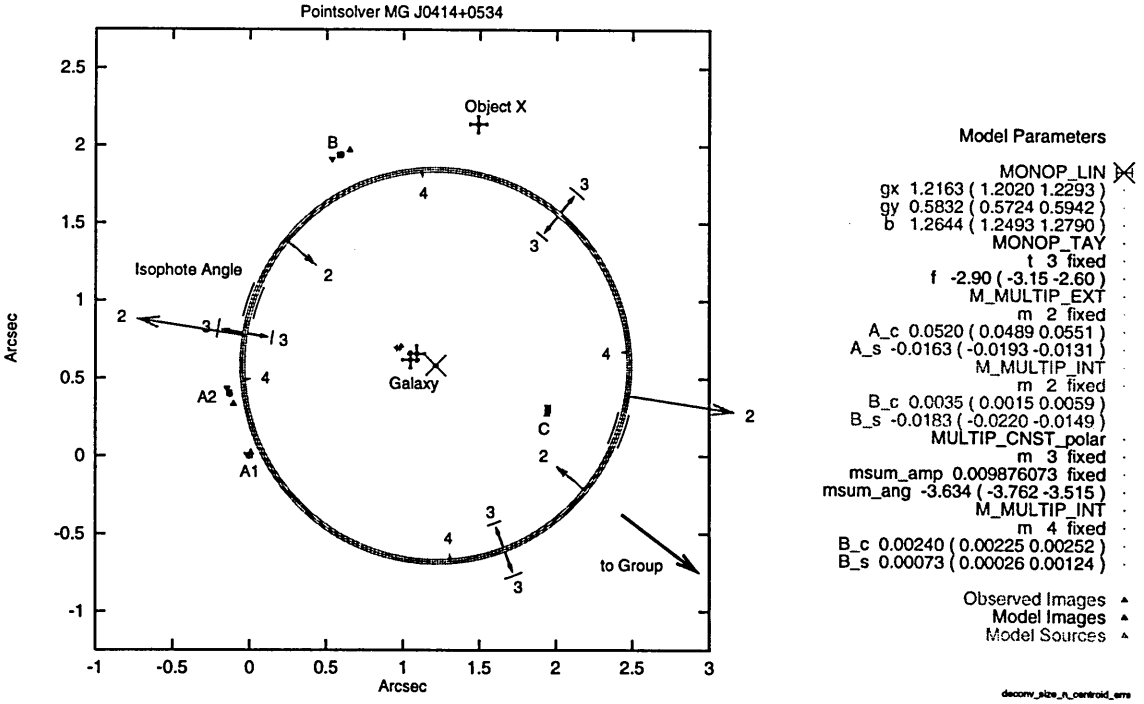

Figure B-29: Model m0t1+m0t3+m2ext+m2int+m3t0+m4int.

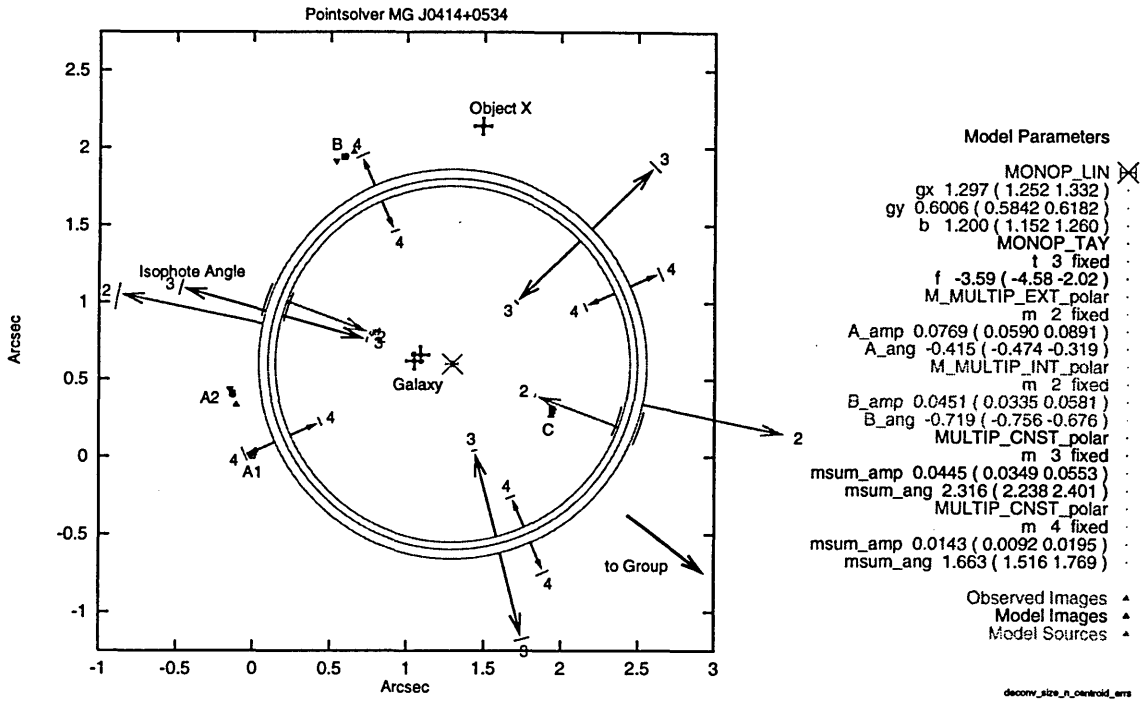

Figure B-30: Model m0t1+m0t3+m2ext+m2int+m3t0+m4t0.

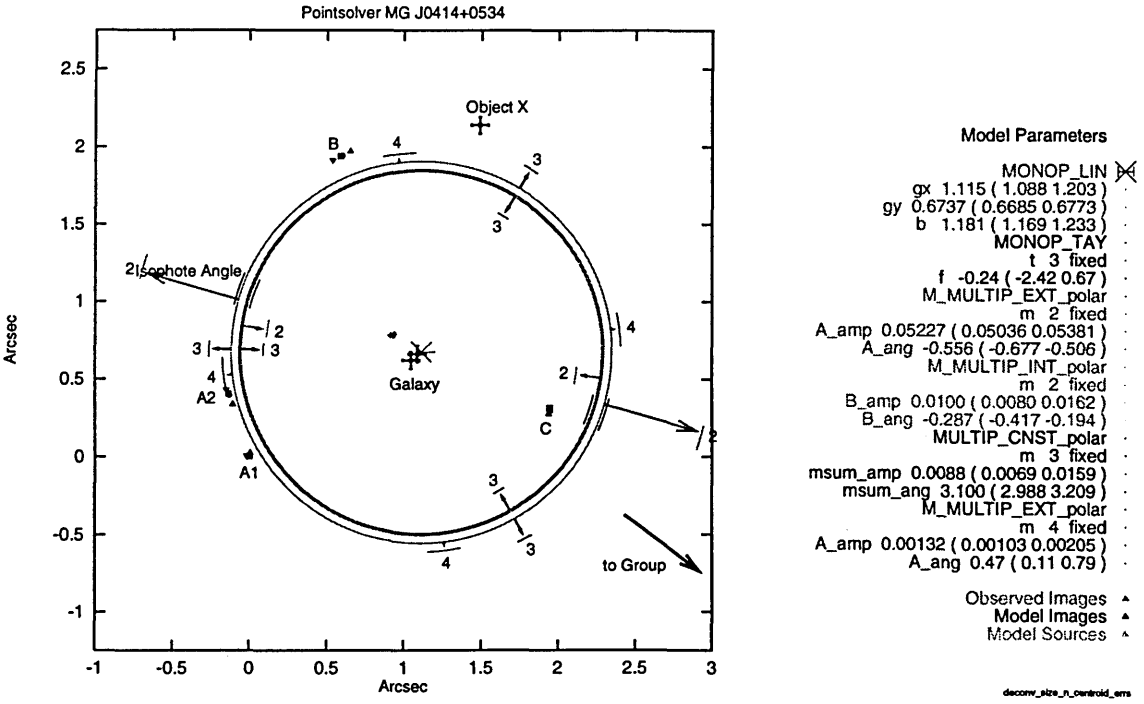

Figure B-31: Model m0t1+m0t3+m2ext+m2int+m3t0+m4ext.

B.4.3 Models with an external $m = 3$ multipole term

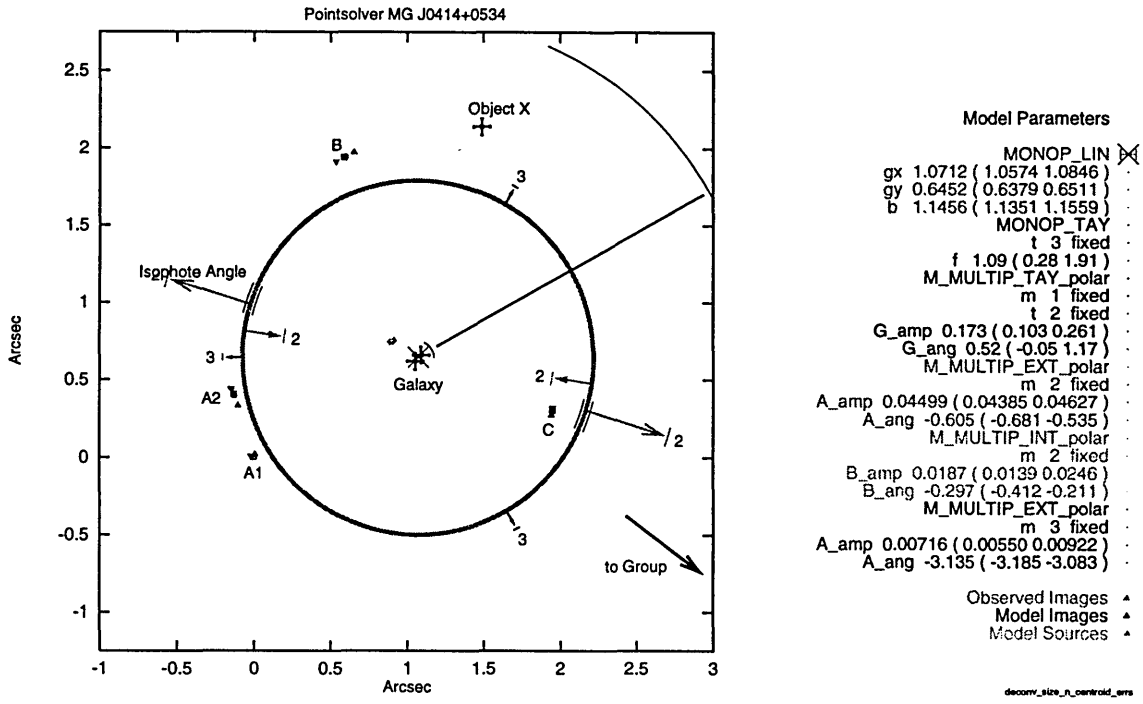

Figure B-32: Model m0t1+m0t3+mm1t2+m2ext+m2int+m3ext.

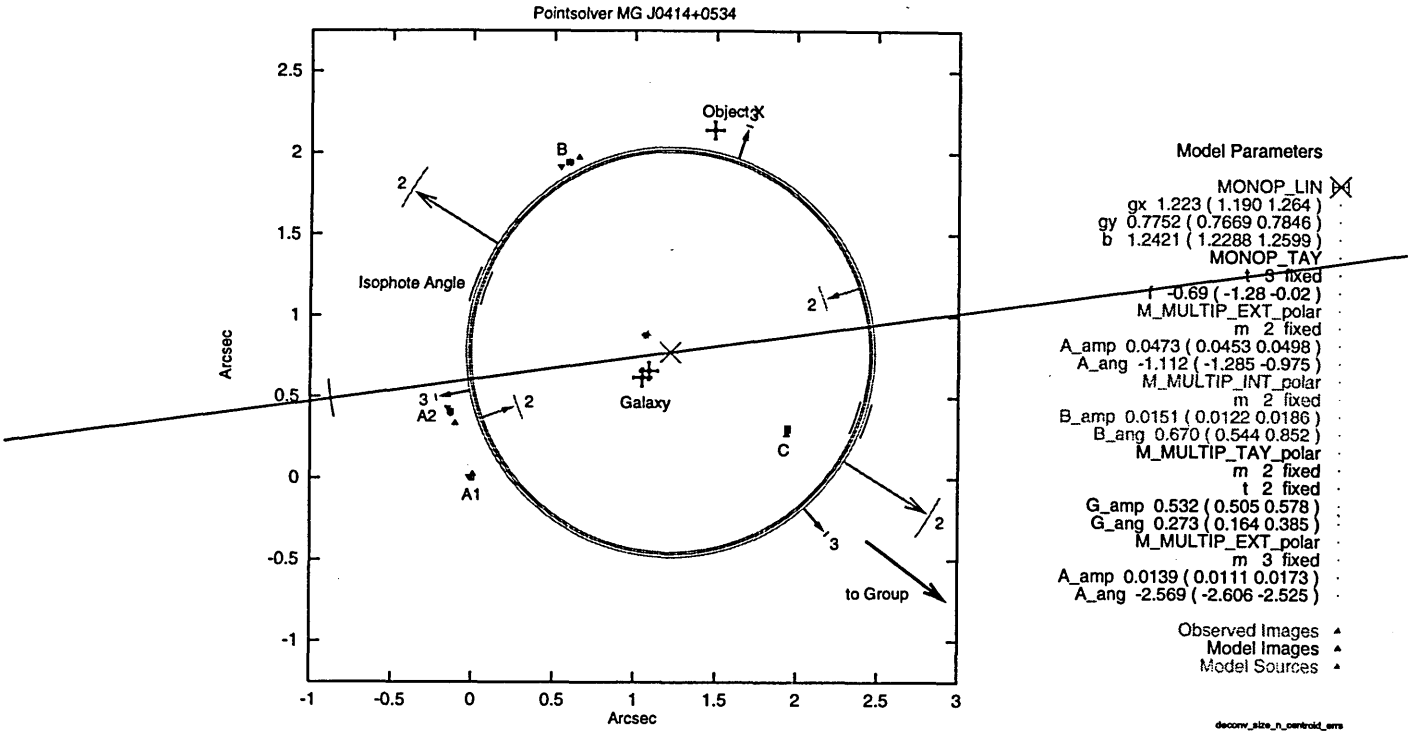

Figure B-33: Model m0t1+m0t3+m2ext+m2int+mm2t2+m3ext.

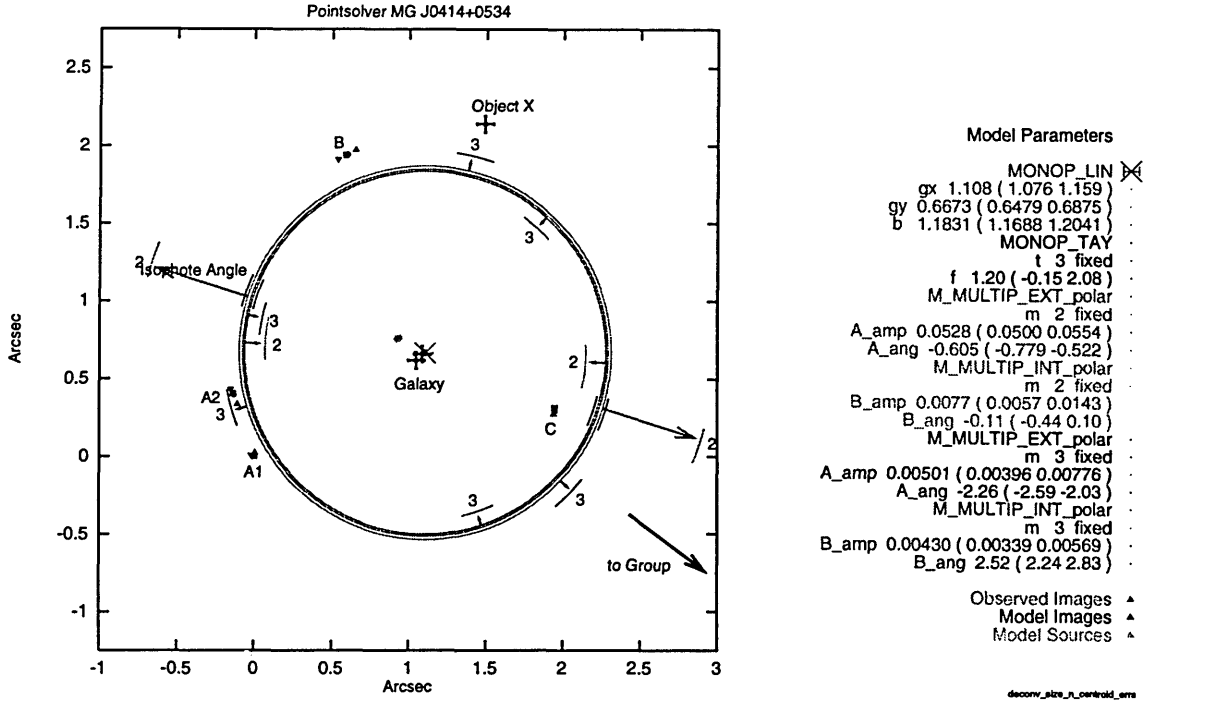

Figure B-34: Model m0t1+m0t3+m2ext+m2int+m3ext+m3int.

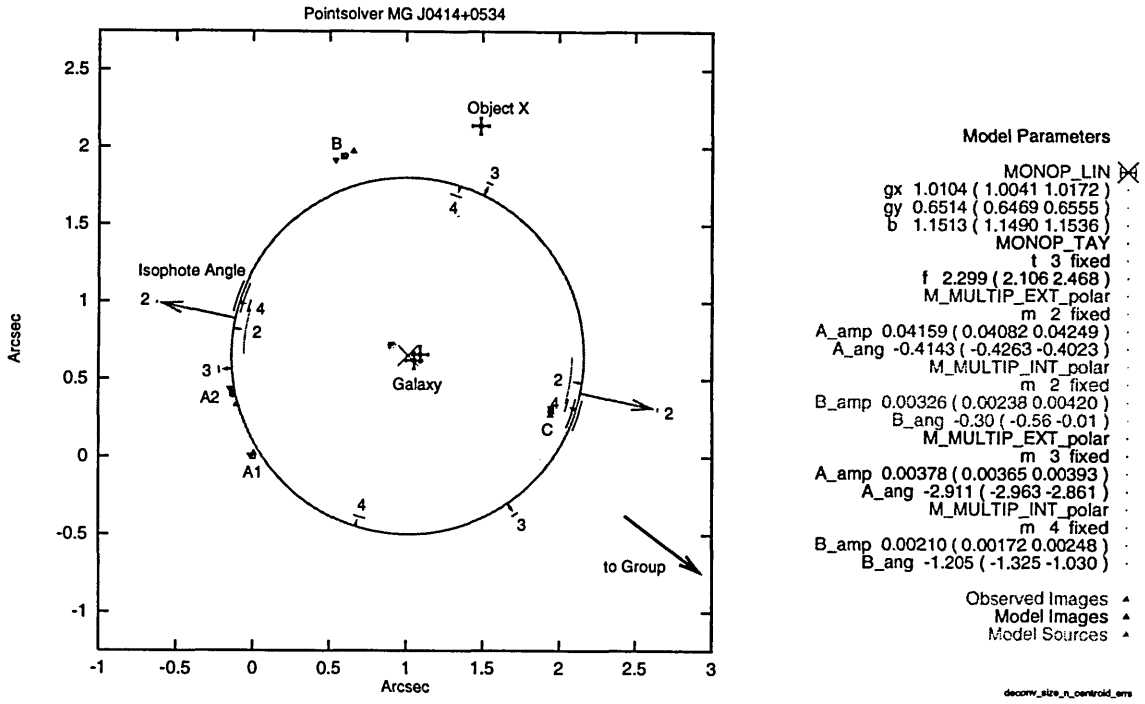

Figure B-35: Model m0t1+m0t3+m2ext+m2int+m3ext+m4int.

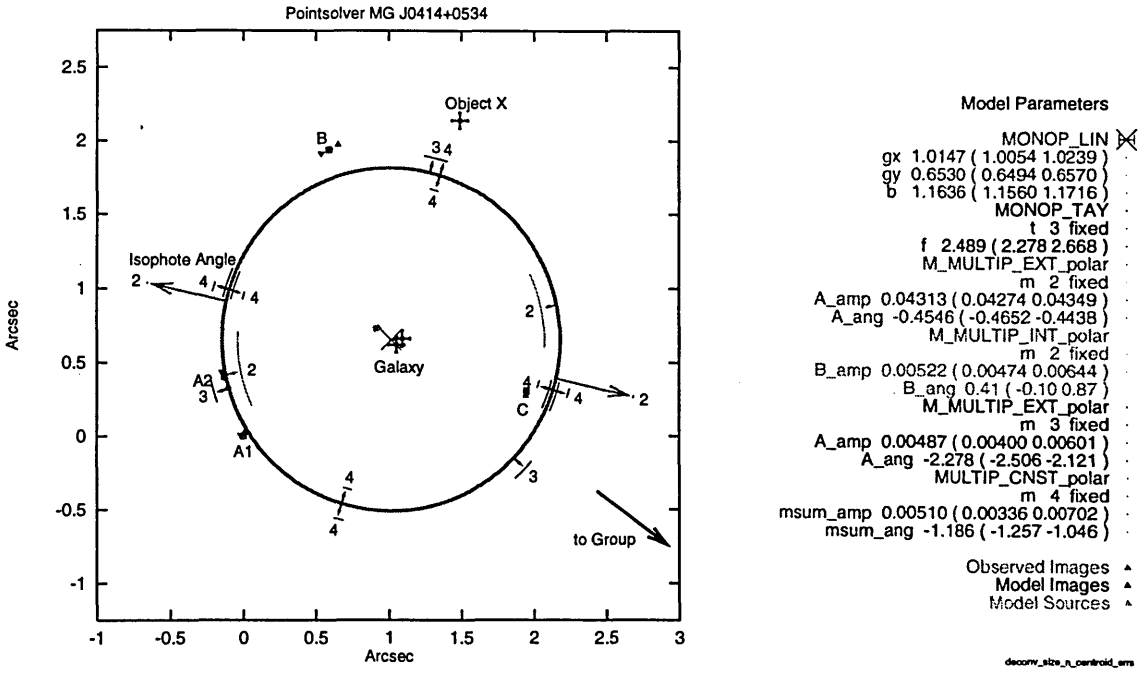

Figure B-36: Model m0t1+m0t3+m2ext+m2int+m3ext+m4t0.

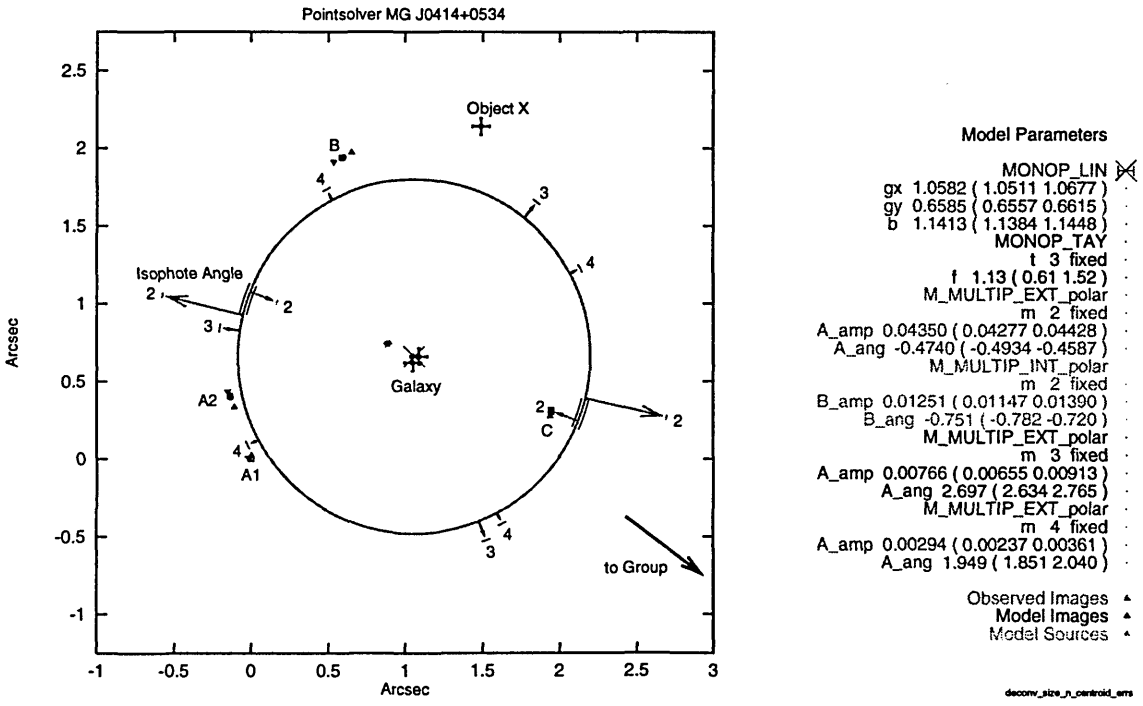

Figure B-37: Model m0t1+m0t3+m2ext+m2int+m3ext+m4ext.

B.5 Time delays tabulated for all models

Model	$\Delta\tau_{A1aA2a}$	$\Delta\tau_{A1aBa}$	$\Delta\tau_{A1aCa}$	$\Delta\tau_{BaCa}$
SIEP	2.01 $\times 10^{-14}$	-2.358 $\times 10^{-12}$	4.525 $\times 10^{-11}$	4.760 $\times 10^{-11}$
SIS+XS (m0t1+m2ext)	2.15 $\times 10^{-14}$	-1.627 $\times 10^{-12}$	2.925 $\times 10^{-11}$	3.088 $\times 10^{-11}$
PM+XS	0.00 $\times 10^{-14}$	0.532 $\times 10^{-12}$	0.482 $\times 10^{-11}$	0.429 $\times 10^{-11}$
m0t1+m2ext+m2int	2.16 $\times 10^{-14}$	-1.422 $\times 10^{-12}$	0.541 $\times 10^{-11}$	0.683 $\times 10^{-11}$
+m3int	3.15 $\times 10^{-14}$	-1.271 $\times 10^{-12}$	0.605 $\times 10^{-11}$	0.732 $\times 10^{-11}$
+mm2t2	4.44 $\times 10^{-14}$	-2.004 $\times 10^{-12}$	0.897 $\times 10^{-11}$	1.097 $\times 10^{-11}$
+m0t3	2.69 $\times 10^{-14}$	-1.413 $\times 10^{-12}$	0.587 $\times 10^{-11}$	0.729 $\times 10^{-11}$
+mm1t2	3.31 $\times 10^{-14}$	-1.487 $\times 10^{-12}$	0.604 $\times 10^{-11}$	0.752 $\times 10^{-11}$
+m4ext	8.10 $\times 10^{-14}$	-3.250 $\times 10^{-12}$	0.760 $\times 10^{-11}$	1.085 $\times 10^{-11}$
+m4t0	13.56 $\times 10^{-14}$	-4.608 $\times 10^{-12}$	0.492 $\times 10^{-11}$	0.953 $\times 10^{-11}$
+m4int	5.59 $\times 10^{-14}$	-1.900 $\times 10^{-12}$	1.846 $\times 10^{-11}$	2.036 $\times 10^{-11}$
+m3ext	$5.19^{+0.09}_{-0.09}$ $\times 10^{-14}$	$-2.326^{+0.038}_{-0.039}$ $\times 10^{-12}$	$0.866^{+0.009}_{-0.009}$ $\times 10^{-11}$	$1.099^{+0.012}_{-0.012}$ $\times 10^{-11}$
+m3t0	$5.88^{+0.09}_{-0.09}$ $\times 10^{-14}$	$-2.481^{+0.037}_{-0.037}$ $\times 10^{-12}$	$0.996^{+0.012}_{-0.012}$ $\times 10^{-11}$	$1.244^{+0.016}_{-0.016}$ $\times 10^{-11}$

Table B.1: "Dimensionless time delays" for two- to four-term models fitted to MG J0414+0534, using the "deconv-size-n-centroid-errs" position error estimate. The time delays are given for the brightest component "a". Image B leads and image C lags.

Model	$\Delta\tau_{A1aA2a}$	$\Delta\tau_{A1aBa}$	$\Delta\tau_{A1aCa}$	$\Delta\tau_{BaCa}$
+m4ext+m4int	$6.99^{+0.18}_{-0.18}$	$-2.154^{+0.059}_{-0.067}$	$0.942^{+0.047}_{-0.043}$	$1.158^{+0.044}_{-0.041}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+m0t3	$5.35^{+0.19}_{-0.17}$	$-2.265^{+0.069}_{-0.077}$	$0.908^{+0.031}_{-0.028}$	$1.134^{+0.038}_{-0.034}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+m4int	$5.88^{+0.28}_{-0.23}$	$-2.489^{+0.037}_{-0.027}$	$0.940^{+0.031}_{-0.025}$	$1.188^{+0.028}_{-0.024}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+m4t0	$6.37^{+0.28}_{-0.27}$	$-1.714^{+0.091}_{-0.097}$	$1.617^{+0.122}_{-0.111}$	$1.788^{+0.113}_{-0.102}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+m4ext	$6.60^{+0.14}_{-0.13}$	$-2.616^{+0.047}_{-0.047}$	$0.987^{+0.032}_{-0.030}$	$1.249^{+0.031}_{-0.029}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+mm2t2	$6.19^{+0.16}_{-0.13}$	$-2.610^{+0.066}_{-0.077}$	$1.133^{+0.020}_{-0.017}$	$1.394^{+0.017}_{-0.016}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3t0+mm1t2	$4.85^{+0.12}_{-0.12}$	$-2.061^{+0.059}_{-0.068}$	$0.838^{+0.027}_{-0.023}$	$1.044^{+0.034}_{-0.029}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m4t0	$7.64^{+0.31}_{-0.32}$	$-2.642^{+0.137}_{-0.132}$	$0.942^{+0.050}_{-0.051}$	$1.206^{+0.039}_{-0.040}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m4int	$6.26^{+0.20}_{-0.22}$	$-2.330^{+0.032}_{-0.028}$	$0.880^{+0.048}_{-0.045}$	$1.113^{+0.046}_{-0.043}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m0t3	$4.23^{+0.07}_{-0.07}$	$-1.912^{+0.027}_{-0.028}$	$0.722^{+0.009}_{-0.008}$	$0.913^{+0.011}_{-0.010}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+mm1t2	$7.15^{+0.48}_{-0.40}$	$-3.027^{+0.185}_{-0.221}$	$1.036^{+0.051}_{-0.046}$	$1.339^{+0.073}_{-0.064}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m3int	$5.89^{+0.12}_{-0.11}$	$-2.486^{+0.034}_{-0.028}$	$1.069^{+0.060}_{-0.051}$	$1.318^{+0.061}_{-0.052}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+mm2t2	$5.36^{+0.03}_{-0.04}$	$-1.985^{+0.019}_{-0.020}$	$0.811^{+0.015}_{-0.015}$	$1.010^{+0.017}_{-0.017}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$
+m3ext+m4ext	$5.15^{+0.19}_{-0.17}$	$-1.802^{+0.034}_{-0.036}$	$1.044^{+0.018}_{-0.017}$	$1.225^{+0.017}_{-0.017}$
	$\times 10^{-14}$	$\times 10^{-12}$	$\times 10^{-11}$	$\times 10^{-11}$

Table B.2: "Dimensionless time delays" for five-term models fitted to MG J0414+0534, using the "deconv-size-n-centroid-errs" position error estimate. The time

Model	$\Delta\tau_{A1aA2a}$	$\Delta\tau_{A1aBa}$	$\Delta\tau_{A1aCa}$	$\Delta\tau_{BaCa}$
+m4ext+m4int+m0t3	6.41×10^{-14}	-1.549×10^{-12}	3.011×10^{-11}	3.166×10^{-11}
+m3t0+m4int+m0t3	$6.91^{+0.20}_{-0.23} \times 10^{-14}$	$-2.734^{+0.058}_{-0.048} \times 10^{-12}$	$1.339^{+0.119}_{-0.111} \times 10^{-11}$	$1.613^{+0.123}_{-0.113} \times 10^{-11}$
+m3t0+m4t0+m0t3	$10.11^{+1.19}_{-1.32} \times 10^{-14}$	$-3.462^{+0.761}_{-0.333} \times 10^{-12}$	$2.477^{+0.348}_{-0.331} \times 10^{-11}$	$2.823^{+0.378}_{-0.322} \times 10^{-11}$
+m3t0+m4ext+m0t3	$6.80^{+1.72}_{-0.72} \times 10^{-14}$	$-2.694^{+0.273}_{-0.703} \times 10^{-12}$	$1.010^{+0.249}_{-0.092} \times 10^{-11}$	$1.280^{+0.318}_{-0.117} \times 10^{-11}$
+m3t0+mm2t2+m0t3	$7.09^{+5.31}_{-1.74} \times 10^{-14}$	$-2.912^{+0.664}_{-2.092} \times 10^{-12}$	$0.864^{+0.548}_{-0.164} \times 10^{-11}$	$1.155^{+0.757}_{-0.230} \times 10^{-11}$
+m3t0+mm1t2+m0t3	$5.25^{+0.17}_{-0.20} \times 10^{-14}$	$-2.202^{+0.075}_{-0.044} \times 10^{-12}$	$0.898^{+0.016}_{-0.031} \times 10^{-11}$	$1.118^{+0.021}_{-0.038} \times 10^{-11}$
+m3ext+mm1t2+m0t3	$6.02^{+0.92}_{-0.79} \times 10^{-14}$	$-2.604^{+0.304}_{-0.360} \times 10^{-12}$	$0.928^{+0.093}_{-0.085} \times 10^{-11}$	$1.188^{+0.129}_{-0.116} \times 10^{-11}$
+m3ext+m4t0+m0t3	$4.62^{+0.29}_{-0.28} \times 10^{-14}$	$-2.910^{+0.323}_{-0.380} \times 10^{-12}$	$0.556^{+0.070}_{-0.076} \times 10^{-11}$	$0.847^{+0.040}_{-0.038} \times 10^{-11}$
+m3ext+m3int+m0t3	$4.93^{+1.11}_{-0.51} \times 10^{-14}$	$-2.133^{+0.161}_{-0.410} \times 10^{-12}$	$0.949^{+0.148}_{-0.093} \times 10^{-11}$	$1.162^{+0.186}_{-0.108} \times 10^{-11}$
+m3ext+mm2t2+m0t3	$5.88^{+0.47}_{-0.51} \times 10^{-14}$	$-2.183^{+0.194}_{-0.174} \times 10^{-12}$	$0.857^{+0.063}_{-0.063} \times 10^{-11}$	$1.075^{+0.077}_{-0.079} \times 10^{-11}$
+m3ext+m4int+m0t3	$4.00^{+0.30}_{-0.25} \times 10^{-14}$	$-2.254^{+0.056}_{-0.059} \times 10^{-12}$	$0.606^{+0.034}_{-0.031} \times 10^{-11}$	$0.832^{+0.030}_{-0.026} \times 10^{-11}$
+m3ext+m4ext+m0t3	$4.70^{+0.22}_{-0.17} \times 10^{-14}$	$-1.734^{+0.029}_{-0.029} \times 10^{-12}$	$0.913^{+0.061}_{-0.046} \times 10^{-11}$	$1.086^{+0.062}_{-0.045} \times 10^{-11}$

Table B.3: "Dimensionless time delays" for six-term models fitted to MG J0414+0534, using the "deconv-size-n-centroid-errs" position error estimate. The time delays are given for the brightest component "a". Image B leads and image C lags.

Model	$\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa}$	$\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa}$
SIEP	0.052	0.053
SIS+XS (m0t1+m2ext)	0.056	0.056
PM+XS	−0.110	−0.110
m0t1+m2ext+m2int	0.263	0.268
+m3int	0.210	0.217
+mm2t2	0.223	0.230
+m0t3	0.241	0.246
+mm1t2	0.246	0.253
+m4ext	0.428	0.443
+m4t0	0.936	0.991
+m4int	0.103	0.106
+m3ext	$0.268^{+0.003}_{-0.003}$	$0.276^{+0.003}_{-0.003}$
+m3t0	$0.249^{+0.002}_{-0.002}$	$0.256^{+0.002}_{-0.002}$

Table B.4: Time delay ratios for two- to four-term models fitted to MG J0414+0534.

Model	$\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa}$	$\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa}$
+m4ext+m4int	$0.229^{+0.014}_{-0.014}$	$0.238^{+0.015}_{-0.014}$
+m3t0+m0t3	$0.249^{+0.002}_{-0.002}$	$0.257^{+0.002}_{-0.002}$
+m3t0+m4int	$0.265^{+0.010}_{-0.012}$	$0.273^{+0.010}_{-0.012}$
+m3t0+m4t0	$0.106^{+0.014}_{-0.013}$	$0.110^{+0.014}_{-0.013}$
+m3t0+m4ext	$0.265^{+0.010}_{-0.011}$	$0.274^{+0.011}_{-0.011}$
+m3t0+mm2t2	$0.230^{+0.007}_{-0.008}$	$0.237^{+0.007}_{-0.008}$
+m3t0+mm1t2	$0.246^{+0.002}_{-0.002}$	$0.253^{+0.002}_{-0.002}$
+m3ext+m4t0	$0.281^{+0.031}_{-0.027}$	$0.291^{+0.031}_{-0.028}$
+m3ext+m4int	$0.265^{+0.017}_{-0.016}$	$0.274^{+0.017}_{-0.016}$
+m3ext+m0t3	$0.265^{+0.003}_{-0.003}$	$0.272^{+0.003}_{-0.003}$
+m3ext+mm1t2	$0.292^{+0.007}_{-0.005}$	$0.301^{+0.007}_{-0.005}$
+m3ext+m3int	$0.233^{+0.011}_{-0.011}$	$0.239^{+0.011}_{-0.012}$
+m3ext+mm2t2	$0.245^{+0.004}_{-0.004}$	$0.253^{+0.004}_{-0.004}$
+m3ext+m4ext	$0.173^{+0.006}_{-0.006}$	$0.178^{+0.006}_{-0.006}$

Table B.5: Time delay ratios for five-term models fitted to MG J0414+0534.

Model	$\Delta\tau_{BaA1a}/\Delta\tau_{A1aCa}$	$\Delta\tau_{BaA2a}/\Delta\tau_{A2aCa}$
+m4ext+m4int+m0t3	0.051	0.054
+m3t0+m4int+m0t3	$0.204^{+0.017}_{-0.015}$	$0.210^{+0.017}_{-0.015}$
+m3t0+m4t0+m0t3	$0.140^{+0.026}_{-0.024}$	$0.144^{+0.027}_{-0.024}$
+m3t0+m4ext+m0t3	$0.267^{+0.015}_{-0.012}$	$0.275^{+0.015}_{-0.013}$
+m3t0+mm2t2+m0t3	$0.337^{+0.020}_{-0.017}$	$0.348^{+0.021}_{-0.018}$
+m3t0+mm1t2+m0t3	$0.245^{+0.001}_{-0.001}$	$0.253^{+0.001}_{-0.001}$
+m3ext+mm1t2+m0t3	$0.281^{+0.010}_{-0.008}$	$0.289^{+0.010}_{-0.008}$
+m3ext+m4t0+m0t3	$0.523^{+0.160}_{-0.110}$	$0.536^{+0.163}_{-0.112}$
+m3ext+m3int+m0t3	$0.225^{+0.013}_{-0.008}$	$0.231^{+0.014}_{-0.008}$
+m3ext+mm2t2+m0t3	$0.255^{+0.016}_{-0.014}$	$0.263^{+0.016}_{-0.015}$
+m3ext+m4int+m0t3	$0.372^{+0.030}_{-0.027}$	$0.381^{+0.030}_{-0.027}$
+m3ext+m4ext+m0t3	$0.190^{+0.012}_{-0.012}$	$0.196^{+0.012}_{-0.012}$

Table B.6: Time delay ratios for six-term models fitted to MG J0414+0534.

Bibliography

Angonin-Willaime, M. C., Vanderriest, C., Hammer, F., & Magain, P., 1994, *Astronomy and Astrophysics* **281**, 388

Bar-Kana, R., 1996, *Astrophysical Journal* **468**, 17+

Barkana, R., 1997, *Astrophysical Journal* **489**, 21+

Binney, J. & Tremaine, S., 1987, *Galactic Dynamics*, Princeton University Press

Blandford, R. D. & Kochanek, C. S., 1987, *Astrophysical Journal* **321**, 658

Breimer, T. G. & Sanders, R. H., 1992, *Monthly Notices of the RAS* **257**, 97

Bridle, A. H., 1989, in R. A. Perley, F. R. Schwab, & A. H. Bridle (eds.), *Synthesis Imaging in Radio Astronomy: A Collection of Lectures from the Third NRAO Synthesis Imaging Summer School*, No. 6 in Astronomical Society of the Pacific Conference Series, pp 443–476, Astronomical Society of the Pacific

Carroll, S. M., Press, W. H., & Turner, E. L., 1992, *Annual Review of Astronomy and Astrophysics* **30**, 499

Clark, B. G., 1980, *Astronomy and Astrophysics* **89**, 377+

Clark, B. G., 1989, in R. A. Perley, F. R. Schwab, & A. H. Bridle (eds.), *Synthesis Imaging in Radio Astronomy: A Collection of Lectures from the Third NRAO Synthesis Imaging Summer School*, No. 6 in Astronomical Society of the Pacific Conference Series, pp 1–10, Astronomical Society of the Pacific

Condon, J. J., 1997, *Publications of the ASP* **109**, 166

Cornwell, T. & Braun, R., 1989, in R. A. Perley, F. R. Schwab, & A. H. Bridle (eds.), *Synthesis Imaging in Radio Astronomy: A Collection of Lectures from the Third NRAO Synthesis Imaging Summer School*, No. 6 in Astronomical Society of the Pacific Conference Series, pp 167–183, Astronomical Society of the Pacific

Ellithorpe, J., 1995, *Ph.D. thesis*, Massachusetts Institute of Technology

Ellithorpe, J. D., Kochanek, C. S., & Hewitt, J. N., 1996, *Astrophysical Journal* **464**, 556+

Falco, E. E., 1993, in J. Surdej, D. Fraipont-Caro, E. Gosset, S. Refsdal, & M. Remy (eds.), *Gravitational Lenses in the Universe*, Proceedings of the 31st Liège International Astrophysical Colloquium, Université de Liège

Falco, E. E., Gorenstein, M. V., & Shapiro, I. I., 1985, *Astrophysical Journal, Letters* **289**, L1

Falco, E. E., Lehar, J., & Shapiro, I. I., 1997, *Astronomical Journal* **113**, 540

Franx, M., Illingworth, G. D., Kelson, D. D., van Dokkum, P. G., & Tran, K.-V., 1997, *A Pair of Lensed Galaxies at z=4.92 in the Field of CL1358+62*, astro-ph/9704090.

Garrett, M. A., Calder, R. J., Porcas, R. W., King, L. J., Walsh, D., & Wilkinson, P. N., 1994, *Monthly Notices of the RAS* **270**, 457+

Garrett, M. A., Patnaik, A. R., Muxlow, T. W. B., Wilkinson, P. N., & Walsh, D., 1992, in R. J. Davis & R. S. Booth (eds.), *Sub-arcsecond Radio Astronomy*, Cambridge University Press

Goldstein, H., 1980, *Classical Mechanics*, Addison-Wesley Publishing Company, second edition

Gorenstein, M. V., Falco, E. E., & Shapiro, I. I., 1988, *Astrophysical Journal* **327**, 693

Grogin, N. A. & Narayan, R., 1996a, *Astrophysical Journal* **464**, 92

Grogin, N. A. & Narayan, R., 1996b, *Astrophysical Journal* **473**, 570

Haarsma, D. B., 1997, *Ph.D. thesis*, Massachusetts Institute of Technology

Haarsma, D. B., Hewitt, J. N., Lehar, J., & Burke, B. F., 1997, *Astrophysical Journal* **479**, 102+

Hewitt, J. N., Turner, E. L., Lawrence, C. R., Schneider, D. P., & Brody, J. P., 1992, *Astronomical Journal* **104**, 968

Hogbom, J. A., 1974, *Astronomy and Astrophysics Supplement Series* **15**, 417

Jackson, J. D., 1975, *Classical Electrodynamics*, John Wiley & Sons, second edition

Kaiser, N. & Squires, G., 1993, *Astrophysical Journal* **404**, 441

Katz, C. A. & Hewitt, J. N., 1993, *Astrophysical Journal, Letters* **409**, L9

Katz, C. A., Moore, C. B., & Hewitt, J. N., 1997, *Astrophysical Journal* **475**, 512+

Keeton, C. R. & Kochanek, C. S., 1997, *Astrophysical Journal* **487**, 42+

Kochanek, C. S., 1991, *Astrophysical Journal* **373**, 354

Kochanek, C. S. & Narayan, R., 1992, *Astrophysical Journal* **401**, 461

Kogan, L., 1995, *Effect of Digitizers Errors on the Cross and Auto Correlation Response of an FX Correlator*, VLBA Scientific Memo 9, National Radio Astronomy Observatory

Kogan, L., 1996, *Errors in Two Dimensional Gaussian Fits*, AIPS Memo 92, National Radio Astronomy Observatory

Kundic, T., Turner, E. L., Colley, W. N., Gott, J. Richard, I., Rhoads, J. E., Wang, Y., Bergeron, L. E., Gloria, K. A., Long, D. C., Malhotra, S., & Wambsganss, J., 1997, *Astrophysical Journal* **482**, 75+

Lawrence, C. R., Cohen, J. G., & Oke, J. B., 1995a, *Astronomical Journal* **110**, 2583+

Lawrence, C. R., Elston, R., Januzzi, B. T., & Turner, E. L., 1995b, *Astronomical Journal* **110**, 2570+

Misner, C. W., Thorne, K. S., & Wheeler, J. A., 1973, *Gravitation*, W. H. Freeman and Company, New York

Moore, C. B., 1996, *Ph.D. thesis*, Massachusetts Institute of Technology

Myers, S. T., Baker, J. E., Readhead, A. C. S., Leitch, E. M., & Herbig, T., 1997, *Astrophysical Journal* **485**, 1+

Narayan, R., 1991, *Astrophysical Journal, Letters* **378**, L5

National Radio Astronomy Observatory, 1995, *AIPS Cookbook*, National Radio Astronomy Observatory, Edgemont Road. Charlottesville, VA 22903-2475, 15-jul-1995 edition, chapter 9: Reducing VLBI data in AIPS

Paczynski, B. & Gorski, K., 1981, *Astrophysical Journal, Letters* **248**, L101

Patnaik, A. R., Browne, I. W. A., King, L. J., Muxlow, T. W. B., Walsh, D., & Wilkinson, P. N., 1993, *Monthly Notices of the RAS* **261**, 435

Patnaik, A. R. & Porcas, R. W., 1996, in C. S. Kochanek & J. N. Hewitt (eds.), *Astrophysical Applications of Gravitational Lensing*, IAU Symposium No. 173, Kluwer Academic Publishers

Pearson, T. J. & Readhead, A. C. S., 1984, *Annual Review of Astronomy and Astrophysics* **22**, 97

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P., 1992, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, second edition

Riess, A. G., Press, W. H., & Kirshner, R. P., 1996, *Astrophysical Journal* **473**, 88+

Schechter, P. L., Bailyn, C. D., Barr, R., Barvainis, R., Becker, C. M., Bernstein, G. M., Blakeslee, J. P., Bus, S. J., Dressler, A., Falco, E. E., Fesen, R. A., Fischer, P., Gebhardt, K., Harmer, D., Hewitt, J. N., Hjorth, J., Hurt, T., Jaunsen, A. O., Mateo, M., Mehlert, D., Richstone, D. O., Sparke, L. S., Thorstensen, J. R., Tonry, J. L., Wegner, G., Willmarth, D. W., & Worthey, G., 1997, *Astrophysical Journal, Letters* **475**, L85

Schechter, P. L. & Moore, C. B., 1993, *Astronomical Journal* **105**, 1

Schmidt, B. P., Kirshner, R. P., Eastman, R. G., Phillips, M. M., Suntzeff, N. B., Hamuy, M., Maza, J., & Aviles, R., 1994, *Astrophysical Journal* **432**, 42

Schneider, P., Ehlers, J., & Falco, E. E., 1992, *Gravitational Lenses*, Springer-Verlag

Schwab, F. R., 1984, *Astronomical Journal* **89**, 1076

Soucail, G., Mellier, Y., Fort, B., Mathez, G., & Cailloux, M., 1988, *Astronomy and Astrophysics* **191**, L19

Thompson, A. R., 1989, in R. A. Perley, F. R. Schwab, & A. H. Bridle (eds.), *Synthesis Imaging in Radio Astronomy: A Collection of Lectures from the Third NRAO Synthesis Imaging Summer School*, No. 6 in Astronomical Society of the Pacific Conference Series, pp 11–37, Astronomical Society of the Pacific

Thompson, A. R., Moran, J. M., & George W. Swenson, J., 1994, *Interferometry and Synthesis in Radio Astronomy*, Krieger Publishing Company, Malabar, Florida

Tonry, J. L., Blakeslee, J. P., Ajhar, E. A., & Dressler, A., 1997, *Astrophysical Journal* **475**, 399+

van Ommen, T., Jones, D. L., Preston, R. A., & Jauncey, D. L., 1995, *Astrophysical Journal* **444**, 561

Vanderriest, C., Angonin-Willaime, M. C., & Rigau, F., 1996, in C. S. Kochanek & J. N. Hewitt (eds.), *Astrophysical Applications of Gravitational Lensing*, IAU Symposium No. 173, Kluwer Academic Publishers

Walker, R. C., 1995, in J. A. Zensus, P. J. Diamond, & P. J. Napier (eds.), *Very Long Baseline Interferometry and the VLBA*, No. 82 in Astronomical Society of the Pacific Conference Series, pp 247–266, Astronomical Society of the Pacific

Witt, H. J., Mao, S., & Schechter, P. L., 1995, *Astrophysical Journal* **443**, 18

Wrobel, J. M., 1996, *VLBA Observational Status Summary*, National Radio Astronomy Observatory, Current version is obtainable from ftp.aoc.nrao.edu:pub/obssum.vlba.ps

Zensus, J. A., Diamond, P. J., & Napier, P. J. (eds.), 1995, *Very Long Baseline Interferometry and the VLBA*, No. 82 in Astronomical Society of the Pacific Conference Series, Astronomical Society of the Pacific