Studies in Gravitational Lensing

and

Numerical Hydrodynamics

Thesis by

Christopher Sharpe Kochanek

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted May 19, 1989)

iii

Acknowledgements

The traditional acknowledgements in a thesis are dividable into four groups: the people who nurtured the intellectual curiosity required to go to graduate school, the people who guided you through graduate school, the people who maintained your sanity during the process, and the groups that bore the fiscal cost.

My parents, Stanley Kochanek and Patricia Sharpe, deserve most of the credit for the first step, and it is to them that this thesis is dedicated. Stuart Shapiro, Saul Teukolsky, and David Chernoff at Cornell University introduced me to scientific research, and it is due to their influence that I continued in astrophysics as a graduate student.

My research advisor, Roger Blandford, has in the last four years provided a (near) perfect mix of advice, collaboration, and independence in the pursuit of this degree. Roger supplies a steady stream of ideas and encouragement, combined with a careful and critical but non-confrontational approach to the ideas of others. John Hawley and Charles Evans introduced me to numerical hydrodynamics, providing both the technical basis for doing research, and the discipline needed to define, develop, and complete projects involving large scale simulations. Sterl Phinney, Kip Thorne, and Peter Goldreich supplied critical readings of many manuscripts, and a willingness to answer (frequent) inane questions. I have also enjoyed my collaborations with Ramesh Narayan, Charles Lawrence, and John Apostolakis.

Sanity was maintained due to the Conspiracy Lunch cabal, and "the astronomers," including Omer Blaes, Chuck Evans, Sam Finn, John Hawley, Pierre-Yves Longaretti, Sterl Phinney, Jens Villumsen, and Tim de Zeeuw in the former, and John Apostolakis, Paul Coppi, Helen Johnston, Steve Meyers, Julie Moses, Alain Picard, Rich Rand, Neil Reid, Chris Tinney, and Chris Wilson in the latter. The daily feast of random conversations, excursions into the absurd, the mountains, and the Pacific helped to keep morale levels reasonably high.

Pat Lyon and Donna Driscoll provided an effective defense against the local bureaucracy.

Abstract

(1) The gravitational lensing cross sections for multiple imaging by elliptical galaxy potentials is examined. Lenses are found to divide into strong and marginal lenses. Three image systems form either *allied* geometries in which the two brightest images lie on the same side of the lens, or *opposed* geometries in which the two brightest images lie on opposite sides of the lens. Strong lenses are dominated by the three image opposed geometry at low amplifications and the five image geometry at high amplifications. Marginal lenses are dominated by the three image allied geometry.

(2) The cross sections for multiple imaging are integrated over the expected distribution of lenses and sources. The sources are taken to be quasars with a standard number-magnitude relation. Approximately one in one thousand quasars will be multiply imaged. Bright lensed quasars are likely to have five images due to the effects of amplification bias.

(3) Approximately one to ten percent of lens systems will involve more than one lensing galaxy either at the same or at a different redshift. The statistical properties of such "two screen" gravitational lenses are evaluated.

(4) An inversion technique for resolved gravitational lenses is developed and applied to the radio ring image MG1131+0456. The technique works both for intensity and polarization maps. The velocity dispersion, position, ellipticity and position angle of the lens are tightly constrained – typically to within ten percent or two tenths of an arc second.

(5) The propagation of a precessing hydrodynamic jet is studied using finite difference techniques in an axisymmetric system. The implications for the precessing jet in SS433 is examined. It is unlikely that the SS433 jet can be hydrodynamic in nature unless the kinetic luminosity is much lower than that required to form the lobes of the W50 remnant.

(6) The tidal disruption of a star on a parabolic orbit past a supermassive black hole is examined using smooth particle hydrodynamics. The spectrum of specific energies for the debris is in close agreement with analytic expectations. Processes leading to the formation of an accretion disk are discussed.

Table of Contents

Part I: Gravitational Lensing

Part II: Numerical Hydrodynamics

Chapter 1

Introduction

1. Gravitational Lensing

The first gravitational lens, Q0957+561, was discovered in 1979 (Walsh *et al.*, 1979). It consisted of two redshift 1.41 quasar images separated by six arc-seconds on the sky, with a redshift 0.36 cluster in between. Since then, gravitational lensing has undergone an explosive growth in terms of both observed lens candidates and theoretical studies. There are at least eight strong candidates for multiple imaging of point sources, and an equal number of weaker candidates. Two types of phenomena have been found in which extended rather than point sources are lensed. The first type, found in 1986 in the cluster Abell 370, consists of extended arcs (which are typically ten to twenty arc seconds in extent) about the cores of rich, distant clusters of galaxies (Lynds and Petrosian, 1986, Soucail *et al.*, 1987). There are now nearly as many arc lenses as there are point source lenses. The second type is the even more dramatic case of the radio ring images MG1131+0456 and MG1634+1346 (Hewitt *et al.*, 1988, Langston *et al.*, 1988). Searches are also underway for statistical effects due to distortions of singly imaged sources by intervening matter. Detailed reviews of gravitational lensing can be found in Blandford and Kochanek (1987) and Blandford, Kochanek, Kovner, and Narayan (1989).

1.1. Historical Introduction

The basic physics of gravitational lensing was understood in 1919, when the solar eclipse expedition of that year confirmed that the gravitational bending of light by the sun agreed with the predictions of general relativity,

$$\alpha = \frac{4M_\odot}{R_\odot} = 1.75''$$

(1.1)

at the solar limb (in units where $G = 1$ and $c = 1$, which we will use throughout). It appears that no one *seriously* discussed the amplification or multiple imaging of background sources until Einstein discussed gravitational lensing by stars in 1936 (Einstein, 1936). The great leap was made in 1937 by Zwicky when he proposed gravitational lensing by galaxies (Zwicky, 1937). Remember that both the sizes and distances to the "nebulae" (as distant galaxies were then called) were extremely uncertain in 1937, so that it was a rather radical proposal.

In Zwicky's lifetime the subject remained quiescent because no examples of the phenomenon were ever found. Sargent reports that Zwicky had an object that he claimed was a gravitational lens although he never published the data. A fellow Caltech astronomer (Munch) said that he would eat the photographic plate if Zwicky's claim was correct. The promise was made at a conference, and Munch appeared to be so perturbed that the moderator of the session hastily brought it to a close. After Zwicky's death, Sargent examined the plate with the hope of serving it to Munch on a platter – unfortunately, the putative lens appears to have been a plate defect (W.L.W. Sargent, *private communication*). Some theoretical progress was made prior to 1979, by Refsdal and Barnothy and Barnothy in the 1960s and Bourassa, Kantowski, Norton, Press, Gunn and several others in the 1970s. The discovery of Q0957+561 in 1979 did not take (all) theorists by surprise.

1.2. Observations of Lenses

In the first few years after the discovery of Q0957+561, progress was slow. A second lens, Q1115+080, was found in 1980 (Weymann *et al.*, 1980), and a third, Q2016+112,

was found in 1984 (Lawrence *et al.*, 1984). (Schematic diagrams of the Q1115+080 and Q2016+112 systems are shown in Figure 1.) Several other objects were claimed to be lenses (*e.g.*, Q2345+007 by Weedman *et al.*, 1982, and Q1635+267 by Djorgovski *et al.*, 1984) but the evidence for lensing was not convincing. The strong cases have more than two images or direct evidence of a lens, while the weak cases have only two quasar images with no indication of a lens.

An object lesson on the danger of assuming that the weak cases are lensed is given by the binary quasar PKS1145-071 (Djorgovski *et al.*, 1987). This system consists of two optical images separated by 4.2 arc-seconds with an optical flux ratio of 2.5 and a redshift separation consistent with $\Delta z = 0$. Based on this information, the system matches the characteristics of the other "dark matter" lenses such as Q2345+007. In this case, however, the brighter quasar is also a radio source, while the fainter one is not – the lower limit on the radio luminosity ratio is 500. This system cannot be a gravitational lens. Yet only $1 - 10\%$ of QSOs (depending on luminosity) are radio sources, so that a clear distinction between binaries and lenses will be much more difficult in most cases. For each pair such as PKS1145-071 there will be ten or more pairs in which neither object is a radio source.

The first example of lensing an extended object was found in 1986 when Lynds and Petrosian (1986) and Soucail *et al.* (1987) reported the discovery of a faint arc extending for 20 arc seconds about the core of the cluster Abell 370. Initially the lens hypothesis for the arc was considered unlikely because it required the core of the cluster to be far more compact than conventional wisdom allowed. In 1988, however, Soucail *et al.* found that the arc had a different redshift from that of the cluster (0.72 for the arc, and 0.37 for the cluster), which proved that the effect was due to lensing. Several smaller arc structures have been found in A370, which reinforce the lensing hypothesis. Since then many rich, distant clusters have been found to contain arcs (Giraud, 1988, Lavery and Henry, 1988, Lynds and Petrosian, 1989).

The object MG1131+0456 found by Hewitt *et al.* (1988) is a radio core-jet source which is transformed into a ring by an intervening galaxy. There are two images of the core, and the jet is wrapped around the two images of the core to form a nearly complete elliptical ring. Unfortunately both the underlying radio source and the lens are too faint for spectroscopy and the redshifts have not been measured. Optical images do not show a ring, but this may be due to the difficulty in resolving the ring at optical wavelengths because of atmospheric seeing. A second example, MG1634+1346, was found by the same group (Langston *et al.*, 1989) – in this case the source contains an unlensed radio lobe, an unresolved core, and a second radio lobe that is lensed into a ring. Both the quasar and the lens are bright enough for spectroscopy, and the redshifts have been determined to be 1.74 and 0.25 respectively.

A summary of the observed lenses is given in Table 1. The lenses are classified to be point sources, arcs, or rings and the case for each lens is summarized by the options of a Scottish jury: guilty, not proven, and not guilty. For their own protection, cases that have been found not guilty are not listed.

Even in the absence of multiple imaging, gravitational lenses produce a distorted image of the universe by altering the shape or flux of objects behind the lens. These distorted, single images rarely show dramatic evidence of the lensing effect (with the possible exception of some of the arc lenses) and it is only in a statistical sample that the distortions can be measured. Several attempts have been made to detect the presence of *statistical* gravitational lensing since 1979. The first attempt (Tyson *et al.*, 1984)

looked for changes in the shape of galaxies. A region of the sky was split into foreground and background galaxies based on their magnitudes. The foreground galaxies induce a tangential distortion in the isophotes of the background galaxies, the magnitude of which can be used to measure the mass of the lens. The goal of Tyson and his collaborators was to set limits on galactic dark matter halos. The effect is subtle because it involves a change in the slope of the lensing potential, and the results were not conclusive.

More recent statistical studies have centered on examining the density of quasars near galaxies in magnitude limited samples (Webster *et al.*, 1988, Fugmann, 1988). The number of quasars increases steeply with the limiting magnitude of the survey, with an eight-fold increase per magnitude below 19, and a three-fold increase per magnitude above 19. The halo of a galaxy slightly amplifies the quasars behind it, so that the effective depth of the survey is higher near galaxies. If the quasar number-magnitude relation is $N(m)$, and the mean amplification induced by galaxies within some radius is A, then the overdensity of lenses near galaxies is $\sim N(m + 2.5 \log A)/AN(m)$. If the typical induced amplification is $A \sim 2$ (it is probably somewhat smaller) then the overdensity will be about 2.4. The Webster survey, which is the largest, reports a mean overdensity of 4.1, which is much larger than expected.

1.3. Theoretical Considerations

The theoretical formalism that is most useful for understanding the lensing phenomenon is based on Fermat's principle (Schneider, 1984, Blandford and Narayan, 1986). General relativity tells us that the time delay induced by a Newtonian gravitational potential Φ is simply 2Φ in geometric units. This allows the construction of a virtual time delay surface, which gives the time delay associated with virtual each ray passing from the source to the observer. Fermat's principle requires that the path that a light ray follows from the source to the observer be an extremum of the virtual time delay.

The typical source and lens in our systems lie at cosmological distances from the observer, and the separation is characterized by the Hubble radius $c/H_0 = 3000h^{-1}$ Mpc where $h = H_0/100$ km s^{-1}Mpc^{-1} is between 0.5 and 1. Because the typical lens is very small (~ 10 kpc for galaxies, or ~ 500 kpc for clusters) compared to the cosmological separations between the lens and the observer or the source, we can use the thin lens approximation. The first consequence of this approximation is that the structure of lens along the line of sight can be ignored, and the effective lensing potential is the two-dimensional Newtonian potential found by integrating the gravitational field along the line of sight, $\phi^{(2)} = \int \Phi ds$. The second consequence of the thin lens approximation is that the deflection angles are small, so that the geometric components of the time delay due to the differences in path length of virtual rays can be expanded in the small deflection angle. The virtual time delay is

$$\tau = (1 + z_L) \left[\frac{1}{2} \frac{D_{OL}D_{OS}}{D_{LS}} (\vec{x} - \vec{u})^2 - 2\phi^{(2)}(\vec{x}) \right] \qquad (1.2)$$

where $\vec{x}$ and $\vec{u}$ are two-dimensional angular distances on the sky for the position of the virtual image and the source. The quantities D_{OL}, D_{OS}, and D_{LS} are the *angular diameter distances* between the Observer, the Lens, and the Source. An angular diameter distance, D, relates the angular separation of two points on the sky, θ, to their physical separation, l, through the relation $\theta = l/D$. In laboratory optics, D is simply the distance, while in gravitational lensing it depends on the background cosmology of the universe.

The final factor of $(1 + z_L)$ corrects the time delay at the lens for the general relativistic time dilation between the lens and the observer.

Given a source position, $\vec{u}$, the image positions, $\vec{x}$, are found using Fermat's principle to lie at solutions of $\nabla_x \tau = 0$, or

$$\vec{u} = \vec{x} - \frac{D_{LS}}{D_{OS}} \vec{\nabla}_x \phi^{(2)}(\vec{x}) \tag{1.3}$$

The distortion of the image is related to the curvature of the time delay surface through $\delta u_i = \tau_{,ij} \delta x_j$ where $\tau_{,ij}$ is the two-dimensional matrix of second derivatives of the time delay surface. If the time delay surface is very flat in some direction (i.e., an eigenvalue of $\tau_{,ij}$ is nearly zero), the image of a finite sized source will be stretched in that direction. The ratio of the area of a source to the area of its image is $A^{-1} = |\tau_{,ij}|$ where A is the amplification. Image positions on which A diverges are said to lie on critical lines, and the corresponding locus of source positions are termed caustics. On the critical line, one of the eigenvalues of $\tau_{,ij}$ is zero, and images are "infinitely" stretched along the corresponding eigenvector.

The caustics mark the boundaries for transitions between different image multiplicities, and the critical lines mark the points where images are created or destroyed as a source crosses a caustic. Caustics can be classified by the local shape of the time delay surface in terms of "catastrophe" theory. Only two types of "catastrophes" are important in gravitational lensing: the fold and the cusp. Both fold and cusp caustics mark transitions in which the image multiplicity changes by two – it is the character of the process by which the images are created or destroyed that differentiates them. When a source crosses a fold catastrophe from the higher to the lower image multiplicity region, two images merge on the critical line associated with the caustic. The amplification of these two images diverges, while that of all other images in the system remains approximately constant. The images vanish when the source is across the caustic. If the transition is made by crossing a cusp caustic, three images merge instead of two. The amplification of the merging images again diverges, but once the source is across the caustic one of the three images is still present and highly amplified. The statistical properties of lenses are dominated by the properties of the fold caustics (because the folds are curves while the cusps are points in the source plane, so that the region of source space dominated by cusps has a lower "dimension" than that dominated by folds).

1.4. Studies in Gravitational Lensing

Early work on gravitational lenses focused on singular, circular potentials (Turner, Ostriker and Gott, 1983). A singular, monotonically-decreasing, two-dimensional lensing potential can have two types of singularities: a divergence at the origin, or a discontinuity in derivatives at the origin. A divergent potential always generates multiple images, and if the potential is monotonic there will always be two images lying on the line connecting the source and the lens, with one image on either side of the lens. Lenses with discontinuities in the derivative of the potential generate multiple images over only a limited region of the source plane, but the boundary separating the multiply-imaging region from the singly imaging region is not a true caustic because the amplification of the images does not diverge. Sources directly behind the lens generate an image that is a circular ring, termed an Einstein ring, about the center of the lens. At this point, the amplification is infinite in the direction along the ring and the origin is a "point caustic."

In Chapter 2 we discuss the structural effects on lensing behavior due to lifting the two degeneracies of these simple lenses by giving the potential a finite core radius or ellipticity. The addition of a core radius gives the lenses a true outer caustic, which is associated with the inner critical line of the lens. We call this caustic the radial caustic because it corresponds to two radially amplified images merging on the inner critical line. The point directly behind the lens still generates Einstein rings due to the merging of two tangentially extended images on the tangential critical line. Image systems that are dominated by the radial caustic we term *allied images* because the two brightest images are found on the same side of the lens, straddling the radial critical line, and image systems dominated by the tangential caustic we term *opposed images* because the two brightest images lie on opposite sides of the lens. If the core radius is small enough, the cross section is dominated by the opposed image geometry – two images are near the tangential critical line, and the third image is captured and deamplified by the core of the lens. As the core radius is increased, the allied geometry becomes increasingly important until it dominates the cross section. In this case the lens is only marginally able to generate multiple images, and a further increase in the core radius leads to the lens becoming sub-critical and unable to generate multiple images.

The point tangential caustic is structurally unstable to non-axisymmetric perturbations, and the introduction of a small ellipticity to the lensing potential causes it to unfold into an astroid consisting of four cusps joined by folds. The region inside the astroid generates five images, and transitions from five to three images are characterized by images merging on the tangential critical line. For the moderate ellipticities characteristic of galaxies, the elliptical lenses are dominated by the three-image opposed geometry, followed by the five-image geometries and only then the three-image allied geometries. Marginal lenses, however, are still dominated by the three-image allied geometry. In the elliptical lenses the three-image opposed geometry is no longer associated with a caustic – it is really a transition between five images and the three-image allied geometry, which is associated with mergers on the radial caustic. As a result, the high amplification cross sections for the various geometries are dominated by the five-image geometries (except when the lens is marginal and the allied geometry again dominates).

In Chapter 3 we apply the results of Chapter 2 to estimate the number and properties of gravitational lenses in the universe. We use the language of particle physics to describe the interaction of background sources with a lens. A lens has a *cross section* for multiple imaging, which is the area of the source plane in which the lens will generate multiple images of a source. The lens can produce three types of multiple image geometries: the three-image allied, three-image opposed, and five-image geometries. We characterize the probability of an image type by its *branching ratio*, which is the fraction of multiply-imaged sources that have that image type. We integrate the cross sections for imaging over the redshift distribution of galaxies to find an optical depth for gravitational lensing. For typical spiral galaxies with velocity dispersion $\sigma = 177$ km s^{-1}, core size $s = 1/3$ kpc, ellipticity $\epsilon = 0.2$, and local number density $0.02h^3$ Mpc^{-3}, the mean optical depth is $\sim 1.5 \times 10^{-4}$ and the branching ratios for the opposed, allied, and five-image geometries are $0.73 : 0.03 : 0.24$ For typical elliptical galaxies with velocity dispersion $\sigma = 306$ km s^{-1} and local number density $0.0088h^3$ Mpc^{-3}, the mean optical depth is $\sim 1.3 \times 10^{-3}$, and the branching ratios are $0.88 : 0.00 : 0.12$. The optical depth for highly amplified lenses is much lower (7×10^{-6} and 3×10^{-5} respectively), but the branching ratios are dominated by the five-image geometry.

Because the quasar number counts increase very steeply as you look at fainter ob-

jects, the high amplification lensing events can have a disproportionate effect on the lenses found in a magnitude limited sample. At a fixed observed magnitude, m, the contribution from images with amplification A is weighted by the number of sources at magnitude $m + 2.5 \log A$. If the number of quasars increases faster than the cross section decreases, the highly amplified faint quasars contribute more than the intrinsically bright quasars – this effect is called *amplification bias*. If we want to understand the probable properties of observed lenses, we must integrate the optical depths against the quasar distribution including the effects of amplification bias to find the properties of a magnitude limited sample of lensed quasars. Because the quasar number counts become less steep above 19 magnitudes, we expect the effects of amplification bias to be most significant for magnitudes less than 17. This can be seen in the ratio of the expected number of five-image lenses to the expected number of three-image lenses. Since the five-image geometry dominates the high magnification cross section, the characteristic sign of amplification bias is an anomalously large ratio of five to three images. Strong amplification bias is found at the bright end of the distribution where the five-image geometry becomes more probable than the three-image geometries. The bias weakens and then vanishes as the magnitude limit approaches the break in the quasar number counts. The lensed images are typically separated by between one-half and two arc seconds, and one image is strongly deamplified and trapped in the core of the galaxy.

Chapter 3 was inspired by the Q2016+112 and Q2237+030 lens systems. The Q2016+112 system seemed impossible to model if all of the galaxies were at the same redshift, and the Q2237+030 lens system showed signs of a second lens system because of the existence of an absorption line in the quasar spectrum. Subramanian, Rees and Chitre (1987) proposed that the "dark matter" lenses (such as Q2345+007 and Q1635+267) were generated by using several weak, dark lenses at different redshifts to give the large observed image separations. The statistical distribution of galaxies leads us to expect that one to ten percent of lensed objects should have a significant contribution from a second galaxy. The equations for the two-screen system were discussed by Schneider and Borgeest (1986), Blandford and Narayan (1986), Kovner (1987), and Padmanabhan and Subramanian (1988), but they did not undertake a systematic study of their properties. The only study of the interactions of two central potentials was a study of two point masses lying at the same redshift by Schneider and Weiss (1986). We therefore undertook a large scale survey of the properties of the two-screen gravitational lens, in which there are two galaxies lying at different redshifts and separations.

The two-screen lens is also dominated by the properties of the fold caustics in the lens although the details of the cross sections show complicated resonances in the cross section and branching ratios. The most efficient lenses always consist of two lenses at the same redshift with separations that are small compared to the typical radius of the tangential critical line. The cross section shows a minimum for separations of twice the radius of the critical line, and for larger separations it increases to the cross section for two isolated lenses. The minimum occurs when the multiply-imaging regions for each galaxy overlap in the source plane. Associated with the cross section minimum is a much higher branching ratio for having seven images. Similarly as the redshift separation increases, the first lens overfocuses light rays at the redshift of the second lens leading to a drop in the overall cross section. These studies were made on the Caltech Mark III hypercube, a coarse grained parallel computer, to take advantage of the large total memory available on parallel machines. There are a number of interesting problems involved in parallelizing the lensing algorithm, which we discuss in Apostolakis and Kochanek (1989).

Unresolved lenses such as Q1115+080 provide relatively few constraints on models for the lensing potential because the lensing potential is sampled only at the locations of the five images. The radio rings (and potentially the arcs) provide a constraint for every multiply-imaged pixel in the map, and Chapter 4 exploits this property of extended images to invert the first radio ring, MG1131+0456. The key to our method is that a gravitational lens preserves the surface brightness of the source so that a resolved image of a lens has the true surface brightness of the source. (An image can be amplified in the sense that the total flux from the source is larger than what we would receive in the absence of the lens if the area of the image is larger than the area of the source.) If we have a model potential and we invert the image to obtain a model source, the extent to which image pixels of differing surface brightnesses are mapped onto the same source pixel provides a measure of the error in the inversion. We define an error measure based on the r.m.s. mismatch in surface brightnesses, and then vary the parameters of the lens model so as to minimize the error measure. The resulting inversion shows a normal radio source consisting of a compact core and a jet lensed by a typical galaxy. Not only can we invert the shape of the ring correctly, but we can also consistently invert polarization maps of the source. The inversion is highly sensitive to the values of the parameters for the lens – the velocity dispersion is constrained to within five percent, and the position of the lens to within two tenths of an arc second. This technique produces stronger constraints on the mass distribution in a moderate redshift galaxy than any other method.

2. Numerical Hydrodynamics

Many problems in astrophysics reduce to solving the Euler equations of hydrodynamics. Until recently, progress was restricted to problems that could be reduced to ordinary differential equations, or to model problems that admit analytic solutions. Only rarely could multi-dimensional dynamical systems be studied with any detail. The typical astrophysical flow is characterized by moderate-to-high Mach numbers, and extremely high Reynolds numbers, which leads to the expectation that the flows will have complicated shock structures mixed with uncharacterizable turbulent behavior on all scales. The recent availability of supercomputers allows us to begin studying these systems without some of the restrictive assumptions required to make analytic progress. Computers, however, are never a panacea, and they introduce their own restrictions and limitations.

The greatest danger is the tendency to regard the tool of simulation as a *deus ex machina* from which results spring as Athena from the head of Zeus. Simulation is never a replacement for a qualitative, physical understanding of a process, but rather it is a tool for isolating which processes dominate the dynamics in systems where there are a large number of competing effects. The second danger in simulations is to confuse the apparent success of a technique, as manifested by a lack of error messages or the continued cooperation of the computer in evolving the system further, with a successful simulation. This particular aspect of a successful simulation is the least part of the overall effort. The real test of correctness is the provision of conclusive evidence that the technique is appropriate for the problem to which it is applied. This requires a careful series of tests and calibrations of the numerical technique, initially on trivial problems, but culminating with problems as closely related to the system we wish to study as is practical. I will discuss the application of the tools of simulation to two astrophysical problems: the propagation of precessing jets, and the tidal disruption of a star during a close encounter with a supermassive black hole.

2.1. Hollow Conical Jets

The first serious astrophysical jet simulations were carried out by Norman, Smarr, Winkler, and Smith in 1982. In these simulations, axisymmetric jets were injected with varying Mach numbers into a homogeneous medium in pressure equilibrium with the jet. The two parameters in the study are the density contrast between the jet and the medium, $\eta = \rho_j / \rho_a$, and the jet internal Mach number, M_j (Norman and Winkler, 1985, Norman, Winkler, and Smarr, 1983, 1984, Kössel and Müller, 1988, Lind, Payne, Meier, and Blandford, 1988).

The simulations of jets have been restricted to two-dimensional (usually axisymmetric) hydrodynamic or restricted MHD (including only the toroidal field components) models. The hydrodynamic equations are probably valid even though the jet is composed of a hot, rarefied plasma because magnetic fields keep the particles closely coupled. (In the absence of magnetic fields, the mean free path of a particle would typically be larger than the jet's width.) The magnetic fields may be strong enough to be dynamically important, in which case the MHD equations must be used, and studies of jets with strong toroidal magnetic fields indicate significant differences in their dynamics (Lind, Payne, Meier and Blandford, 1988, Clarke, Norman, and Burns, 1986). Detailed comparisons between observed and simulated astrophysical jets are not practical because the only diagnostics we have of the observed jets are the line of sight integrals of extremely complicated radiative emission processes through the jet. Nonetheless, we can obtain physical insight into the dynamical processes through the use of simulation, even if it represents a highly idealized model.

The stability of the jet is largely determined by the Mach number. If we assume a steady flow at velocity v through a channel of width A, then the mass flux $\rho v A$ must be a constant, and Bernoulli's principle requires that the specific energy, $v^2/2 + \gamma P/\rho(\gamma - 1)$, remain constant along flow lines. Under these assumptions, the response of the pressure to a change in the width of the channel is

$$\frac{dP}{P} \propto \frac{1}{1 - M^2} \frac{dA}{A} \tag{2.1}$$

If the channel's area is reduced ($dA < 0$) a subsonic flow experiences a pressure drop in the constriction ($M < 1 \longrightarrow dP \propto dA$) while a supersonic flow experiences a pressure increase in the constriction ($M > 1 \longrightarrow dP \propto -dA$). A pressure drop tries to constrict the channel further, leading to an instability, while a pressure increase resists the constriction and stabilizes the flow. If we move into the rest frame of the jet, the same criterion holds except that the internal Mach number of the jet is replaced by the Mach number of the jet relative to the ambient sound speed. The jet will be stable against pinching modes if the jet is supersonic in both frames, $v > c_j + c_a$ or $M_j > 1 + \eta^{1/2}$. The first unstable mode of the supersonic jet must now be a "kinking" mode rather than a "sausage" mode. As most astrophysical jets are extremely light ($\eta \ll 1$) the condition on the Mach number for stability reduces to $M_j > 1$ (Norman and Winkler, 1985).

Most numerical studies have been restricted to two-dimensional, axisymmetric jets in which the symmetry suppresses all kinking modes. Real jets are, of course, three-dimensional and analytic studies of the growth of perturbations in cylindrical jets indicate the presence of unstable helical modes (Payne and Cohn, 1985, Ferrari, Massaglia, and Trussoni, 1982). Numerical studies of kink modes have been limited to examining two-dimensional slab jets, which are excited with side to side oscillations at the orifice (Hardee and Norman, 1988, Norman and Hardee, 1988). The slab jets are unstable, but they are

somewhat unrealistic because the jet acts like a wall separating two regions of ambient gas. In the cylindrical geometry, the ambient gas can communicate by sending signals around, as well as through, the jet. Moreover, the models for jets that are used to compute the instabilities differ from the observed jets by neglecting to include the cocoon of shocked jet gas that sheaths astrophysical jets and protects them from the influences of the external medium. Three-dimensional simulations will be required to understand the (observed) stability of supersonic jets.

In the rest frame of the interface between the head of the jet and the shocked ambient medium the ram pressures of the jet and the ambient medium must be balanced. If the interface advances at velocity v_w, and the areas of the jet and the working surface are A_j and A_w respectively, then

$$A_j \left[\rho_j(v_j - v_w)^2 + P_j\right] = A_w \left[\rho_a v_w^2 + P_a\right] \tag{2.2}$$

The thermal pressure is of order M_j^{-2} of the ram pressure so that the pressure can be neglected in high Mach number flows. This implies that the Mach number of the working surface (relative to the ambient sound speed) is

$$M_w \simeq M_j \left(\frac{A_j}{A_w}\right)^{1/2} \frac{1}{1 + \epsilon} \qquad \epsilon \equiv \left(\frac{\rho_j A_j}{\rho_a A_w}\right)^{1/2} \tag{2.3}$$

where the radius of the working surface is typically two to three jet radii (Norman and Winkler, 1985, Lind, Payne, Meier, and Blandford, 1988). The jet gas passes through a terminal shock structure typically consisting of a Mach disk on the axis and an annular shock that deflects the gas into the cocoon. Low Mach number or high density jets require no cocoon to contain the spent jet gas because the jet moves almost ballistically through the ambient medium. Jets in this regime are called "naked beam" jets. High Mach number and low density jets generate a large cocoon because they advance very slowly through the ambient gas.

The cocoon is not a quiescent region – it is filled with vortices of varying size and strength. Some of the vortices are supersonic and exhibit radial shocks in their cores. The motions in the cocoon periodically impinge on the surface of the beam, triggering the formation of crossed shocks along the axis of the jet. The perturbations travel in from the surface of the beam at approximately the sound velocity (the more violent perturbations can travel supersonically) so that the characteristic angle of the crossed shocks is the Mach angle.

Most astrophysical jets are thought to resemble the cylindrical jets described above. The largest extragalactic jets have velocities near the speed of light and maintain their structure over hundreds of thousands of light years (see Begelman, Blandford, and Rees, 1984 for a review). There are a few exceptions, the most spectacular one of which is the galactic object SS433 (see Margon 1984 for a review). The SS433 jet precesses with a period of about one half of a year on the surface of a cone with a half angle of twenty degrees. The jet velocity is known to be $0.26c$ from radio observations of the jet near SS433 where clumps of gas emit radio waves that allow the trajectory to be mapped as the jet precesses. The mildly relativistic velocity allows the geometry to be fixed exactly (Hjellming and Johnston, 1982, 1985). Surrounding SS433 is the radio shell and possible supernova remnant W50. W50 is strongly elongated along the precession axis of the SS433 jet: along the major axis W50 is 100 parsecs from SS433, and along the minor axis it is only 50 parsecs away.

The natural assumption is that the elongation of W50 is due to the action of the jet (Begelman *et al.*, 1980, Davidson and McCray, 1980, König1, 1983). The lobes of W50 are considerably narrower than the opening angle of the precession cone: the half angle relative to SS433 is approximately ten degrees compared to twenty degrees for the jet. This implies that the jet must be focussed as it propagates from SS433 to the lobes. The only model for a hydrodynamic focusing mechanism was proposed by Eichler (1983) who assumed that the pressure on the interior of the precession cone is negligible. This is equivalent to assuming that the interior gas cools very efficiently. Unfortunately the jet can only be observed on scales that are small compared to W50 so there is no direct evidence of the jet focusing and then impinging on the lobes. There is only indirect evidence for the presence of the jet in the region between SS433 and W50 from observations of diffuse X-ray emission and the positions of a few optical filaments (Zealey, Dopita, and Malin, 1980, Watson *et al.*, 1983). In Chapter 6 we make a detailed study of the propagation of hollow jets to understand the differences between hollow jets and the filled jets discussed above. We are restricted to two-dimensional, axisymmetric simulations, but this approximation is quite good because the precession time of the jet is short compared for the time scale to the jet to propagate from SS433 to W50 (0.5 years versus ~ 1000 years). Our conclusion is that purely hydrodynamical effects are unable to account for the apparent geometry of the SS433-W50 system. One must either invoke additional physics such as non-adiabaticity (through radiative cooling – which is not a likely solution) or non-hydrodynamic focusing (using magnetic fields) or one must drop the assumption of a present day interaction between the jet and W50.

2.2. *Tidal Disruption*

Active galactic nuclei (AGN) are the most luminous objects in the universe, with luminosities from 10^{42} to 10^{46} ergs s^{-1}. The higher luminosity is the equivalent of converting several solar masses of matter into radiation every year. AGNs are thought to be driven by accretion onto a supermassive black hole lying at the center of the galaxy. Material enters an accretion disk about the black hole, and viscous forces slowly release the potential energy of the disk gas leading to the radiative emission. The luminosity is limited by the Eddington luminosity of the black hole, which is the luminosity at which the radiation pressure on an infalling electron exceeds the gravitational attraction. For a black hole of mass $M_h = 10^6 M_\odot = M_6$ solar masses, $L_E \simeq 10^{44} M_6$ ergs s^{-1}, and an accretion rate of $\simeq 0.002 M_6 \epsilon^{-1}$ solar masses per year is required for efficiency $\epsilon < 1$ in converting mass to radiated energy. This implies that the highest luminosity AGNs require central black holes of $\sim 10^9 M_\odot$ radiating near the Eddington limit, with accretion rates of $2\epsilon^{-1}$ solar masses per year. Lower luminosity AGNs require much smaller black holes and accretion rates (see Begelman, Blandford, and Rees, 1984, or Begelman, 1985 for reviews of AGN physics).

The accretion disk must be continuously fed if the AGN is to maintain its luminosity and a variety of possible mechanisms have been proposed, including global instabilities, stellar winds, stellar collisions, and tidal disruption. We want to study the dynamics of the tidal disruption mechanism in detail. Disruption can supply fuel only for low luminosity AGNs because the stars pass through the event horizons of the black hole before disrupting if the black hole is more massive than $10^9 M_\odot = 10^3 M_6$, and because the density of stars required to maintain a disruption rate of one star each year is so high that other processes such as collisions are more important (Phinney, 1989, or Rees, 1988 for reviews). The black hole disrupts all stars on nearly radial orbits in a dynamical time, so that the "loss

cone" of stars with angular momenta low enough to be captured is rapidly emptied. The loss cone is refilled on the diffusion time scale, which is controlled either by the relaxation time scale of the star cluster (Young et al., 1977) or by the asymmetries in the potential (Norman and Silk, 1983). In either case, the time between disruptions for a M_6 black hole such as the one which may lie at the galactic center is approximately one thousand years. In Chapter 7 we consider the detailed dynamics of the tidal disruption mechanism for the most common encounter using the technique of smooth particle hydrodynamics.

A star passing a black hole will be tidally disrupted if the surface gravity of the star is less than the tidal gravity of the black hole. If the radius of the orbital pericenter is R_p the star will disrupt when the ratio,

$$\eta \equiv \left(\frac{R_p^3}{GM_h} \frac{GM_*}{R_*^3} \right)^{1/2} \tag{2.4}$$

satisfies $\eta \lesssim 1$ (Press and Teukolsky, 1977). For a star like the sun passing a $10^6 M_\odot$ black hole this occurs when $R_p \lesssim 100 R_\odot$, which is much larger than the Schwarzschild radius of the black hole ($\sim 4 R_\odot$). The disruption orbits are effectively parabolic, so that when the star disrupts, half of the mass is bound to the black hole and eventually captured, while the other half is unbound and ejected from the system. The key realization about this process was made by Lacy, Townes, and Hollenbach in 1982 when they pointed out that the characteristic energy of the debris is not the binding energy of the star, ϵ_b, but the potential energy difference across the star at pericenter, $\Delta\epsilon$. The spread in the specific binding energies of the debris is $\Delta\epsilon \sim (GM_h/R_p)(R_*/R_p) \gg \epsilon_b \sim (GM_h/R_*)$. This corresponds to velocities of several thousand kilometers per second for the ejecta from a solar type star. The captured debris begins to return after one month and continues to return at a super-Eddington rate for approximately two years after the disruption of a solar type star on an $\eta = 1$ orbit.

The binding energy of the debris is very small compared to the binding energy of a circular orbit with the same angular momentum, so that the time scale for the orbits to circularize into a disk near the black hole is limited by the time scale to radiate the binding energy. If the gas radiates at fraction f of the Eddington luminosity, the time scale for the gas to cool and settle after the orbits are circularized is $\sim (GM_h M_*/R_p)/f L_E \sim 3 f^{-1}$ years. Hence the disruption of a single star supports accretion onto a $10^6 M_\odot$ black hole for anywhere from a few to one hundred years depending on the details of the post-disruption evolution (Rees, 1988). This means that AGN with black hole masses near M_6 will be in a quiescent state most of the time because of the long intervals between disruption events punctuated by Eddington level emission flares while the material from a disruption is accreted onto the black hole. Perhaps our own galaxy contains such a black hole, but it is in its quiescent state.

2.3. References

Apostolakis, J., and Kochanek, C.S., 1988, in *The Third Conference on Hypercube Concurrent Computers and Applications, Vol. 2*, ed. G.C. Fox, (New York: Assoc. for Computing Machinery) p. 963.

Barnothy, J.M. and Barnothy, M.F., 1968, *Science,* **162**, 348.

Barnothy, J.M., and Barnothy, M.F., 1972, *Astrophys. J.,* **174**, 477.

Begelman, M.C., 1985, in *Astrophysics of Active Galaxies and Quasi-Stellar Objects*, ed. J.S. Miller, (Univ. Science Books: Mill Valley), p. 411.

Begelman, M.C., Blandford, R.D., and Rees, M.J., 1984, *Rev. Mod. Phys.,* **56**, 255.

Begelman, M.C., Sarazin, C.L., Hatchett, S.P., McKee, C.F., and Arons, J., 1980, *Astrophys. J.,* **238**, 722.

Blandford, R.D., Kochanek, C.S., Kovner, I., Narayan, R., 1988, *Science, in press.*

Blandford, R.D., and Kochanek, C.S., 1987, in *Dark Matter in the Universe*, ed. J. Bahcall, T. Piran, and S. Weinberg, (World Scientific: Singapore), p. 133.

Blandford, R.D., and Narayan, R., 1986, *Astrophys. J.,* **310**, 568.

Bourassa, R.R., Kantowski, R., and Norton, T.D., 1973, *Astrophys. J.,* **185**, 747.

Bourassa, R.R., and Kantowski, R., 1975, *Astrophys. J.,* **195**, 13.

Clarke, D.A., Norman, M.L., and Burns, J.O., 1986, *Astrophys. J. (Letters),* **311**, L63.

Davidson, K., and McCray, R., 1980, *Astrophys. J.,* **241**, 1082.

Djorgovski, S., and Spinrad, H., 1984, *Astrophys. J. (Letters),* **282**, L1.

Djorgovski, S., Perley, R., Meylan, G., and McCarthy, P., 1987, *Astrophys. J. (Letters),* **321**, L1.

Eichler, D., 1983, *Astrophys. J.,* **272**, 48.

Einstein, A., 1936, *Science,* **84**, 506.

Ferrari, A., Massaglia, S., and Trussoni, E., 1982, *Mon. Not. R. astr. Soc.,* **198**, 1065.

Fugmann, W., 1988, *MPI preprint.*

Giraud, E., 1988, *Astrophys. J. (Letters),* **334**, L69.

Hardee, P.E., and Norman, M.L., 1988, *Astrophys. J.,* **334**, 70.

Hewitt, J.N., Turner, E.L., Lawrence, C.R., Schneider, D.P., Gunn, J.E., Bennett, C.L, Burke, B.F., Mahoney, J.H., Langston, G.I., Schmidt, M., Oke, J.B., and Hoessel, J.G., 1987, *Astrophys. J.,* **321**, 706.

Hewitt, J.N., Turner, E.L., Schneider, D.P., Burke, B.F., Langston, G.I., and Lawrence, C.R., 1988, *Nature,* **333**, 537.

Hjellming, R., and Johnston, K., 1982, in *Extragalactic Radio Sources*, IAU Symposium 97, ed. D. Heeschen, and C.M. Wade, (Dordrecht: Reidel), p. 197.

Hjellming, R., and Johnston, K., 1985, in *Radio Stars*, ed. R. Hjellming and D.M. Gibson, (Dordrecht: Reidel), p. 309.

Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., and Perley, R., 1985, *Astr. J.,* **90**, 691.

Königl, A., 1983, *Mon. Not. R. astr. Soc.,* **205**, 471.

Kössl, D., and Müller, E., 1988, *Astr. Astrophys.,* **206**, 204.

Kovner, I., 1987, *Astrophys. J.,* **316**, 52.

Lacy, J.H., Townes, C.H., and Hollenbach, D.J., 1982, *Astrophys. J.,* **262**, 120.

Langston, G.I., Schneider, D.P., Conner, S., Carilli, C.L., Lehar, J., Burke, B.F., Turner, E.L., Gunn, J.E., Hewitt, J.N., and Schmidt, M., 1989, *preprint.*

Lavery, R.J., and Henry, J.P., 1988, *Astrophys. J. (Letters),* **329**, L21.

Lawrence, C.R., Schneider, D.P., Schmidt, M., Bennet, C.L, Hewitt, J.N., Burke, B.F., Turner, E.L., and Gunn, J.E., 1984, *Science,* **223**, 46.

Lind, K.L., Payne, D.G., Meier, D.L, and Blandford, R.D., 1988, *Caltech preprint.*

Lynds, R., and Petrosian, V., 1986, *Bull. AAS,* **18**, 1014.

Lynds, R., and Petrosian, V., 1988, *Astrophys. J.,* **336**, 1.

Magain, P., Surdej, J., Swings, J.P., Borgeest, U., Kayser, R., Kuhr, H., Refsdal, S., and Remy, M., 1988, *Nature,* **334**, 325.

Margon, B., 1984, *Ann. Rev. Astron. Astrophys.,* **22**, 507.

Meylan, G., and Djorgovski, S., 1988, *Astrophys. J. (Letters),* **338**, L1.

Norman, C.A., and Silk, J., 1983, *Astrophys. J.,* **266**, 502.

Norman, M.L., Winkler, K.-H., and Smarr, L.L., 1983, in *Physics of Energy Transport in Extragalactic Radio Sources,* ed. A. Bridle and J. Eilek (Green Bank: NRAO), p. 150.

Norman, M.L., Smarr, L.L., Winkler, K.-H., and Smith, M.D., 1982, *Astr. Astrophys.,* **113**, 285.

Norman, M.L., Winkler, K.-H., and Smarr, L.L., 1983, in *Astrophysical Jets,* ed. A. Ferrari and A.G. Pacholczyk, (Dordrecht: Reidel), p. 227.

Norman, M.L., and Winkler, K.-H., 1985, *Los Alamos Science,* **12**, 38.

Norman, M.L., and Hardee, P.E., 1988, *Astrophys. J.,* **334**, 80.

Padmanabhan, T., and Subramanian, K., 1988, *Mon. Not. R. astr. Soc.,* **236**, 265.

Payne, D.G., and Cohn, H., 1985, *Astrophys. J.,* **291**, 655.

Phinney, E.S., 1989, in *The Galactic Center,* IAU Symposium 136, ed. M. Morris, (Dordrecht: Kluwer Academic).

Press, W.H., and Gunn, J.E., 1973, *Astrophys. J.,* **185**, 397.

Press, W.H., and Teukolsky, S.A., 1977, *Astrophys. J.,* **213**, 183.

Rees, M.J., 1988, *Nature,* **333**, 523.

Refsdal, S., 1964, *Mon. Not. R. astr. Soc.,* **128**, 295.

Refsdal, S., 1964, *Mon. Not. R. astr. Soc.,* **128**, 307.

Schneider, P., 1984, *Astr. Astrophys.,* **140**, 119.

Schneider, P., and Weiss, A., 1987, *Astr. Astrophys.,* **164**, 237.

Soucail, G., Mellier, Y., Fort, G., Mathez, G., and Cailloux, M., 1988, *Astr. Astrophys.,* **191**, L19.

Soucail, G., Fort, B., Mellier, Y., and Picat, J.P., 1987, *Astr. Astrophys.,* **172**, L14.

Subramanian, K., Rees, M.J., and Chitre, S.M., 1987, *Mon. Not. R. astr. Soc.,* **224**, 283.

Surdej, J., Magain, P., Swings, J.P., Borgeest, U., Courvoisier, T.J.L., Kayser, R., Keller-man, K.I., Kuhr, H., and Refsdal, R., 1987, *Nature,* **329**, 22.

Turner, E.L., Ostriker, J.P., and Gott, J.R., 1984, *Astrophys. J.,* **284**, 1.

Tyson, J.A., Valdes, F., Jarvis, J.F., and Mills, A.P., 1984, *Astrophys. J. (Letters),* **281**, L59.

Walsh, T., Carswell, R.F., and Weymann, R.J., 1979, *Nature,* **279**, 381.

Watson, M.G., Willingale, R., Grindlay, J.E., and Seward, F.D., 1983, *Astrophys. J.,* **273**, 688.

Webster, R.L., Hewett, P.C., Harding, M.E. and Wegner, G.A., 1988, *Nature,* **336**, 358.

Weedman, D.W., Weymann, R.J., Green, R.F., and Heckmann, T.M., 1982, *Astrophys. J. (Letters),* **255**, L5.

Weymann, R.J., Latham, D., Angel, J.R.P., Green, R.F., Turnshek, D.A., Turnshek, D.E., and Tyson, J.A., 1980, *Nature,* **285**, 641.

Young, P.J., Shields, G.A., and Wheeler, J.C., 1977, *Astrophys. J.,* **212**, 367.

Zealey, W.J., Dopita, M.A., and Malin, D.F., 1980, *Mon. Not. R. astr. Soc.,* **192**, 731.

Zwicky, F., 1937, *PhysRevLett,* **51**, 290.

Zwicky, F., 1937, *PhysRevLett,* **52**, 679.

Zwicky, F., 1937, *Helvetica Physica Acta,* **6**, 1933.

TABLE 1
Lens Candidates

| QSO	z_Q	z_L	m_Q	$\theta('')$ arc-seconds	Number of Images	How	Type	Verdict	Who
0957+561	1.41	0.36	17	6	2	R,O	point + jet	guilty	Walsh et al., 1979
1115+080	1.72	?	16	2	4	O	point		Weymann et al., 1980
2016+112	3.27	0.8, ?	22	3	3	R,O	point		Lawrence et al., 1984
2237+031	1.7	0.04, 0.6?	17	1	4	O	point		Huchra et al., 1985
0142-100	2.72	~0.5	17	2	2	O	point		Surdej et al., 1987
3C324	1.2	0.84	20	2	2	O	point		Le Fèvre et al., 1987
1413+117	2.55	?	17	1	4	O	point		Magain et al., 1988
0414+456	?	?		2	4	R,O	point		Hewitt et al., 1988
1131+0456	?	?	22?	2	2	R,O	ring	guilty	Hewitt et al., 1988
1634+1346	1.74	0.25	21	2		–	R	ring	Langston et al., 1989
A370	0.72	0.37	22	20	–	O	arc	guilty	Lynds et al., 1986
A963	?	0.21	22	18	–	O	arc		Lavery et al., 1988
CL0500-24	?	0.32		14	–	O	arc		Giraud, 1988
CL2244-02	?	0.33	20	15	–	O	arc		Lynds et al., 1989
2345+007	2.2	?	19	7	2	O	point	not proven	Weedman et al., 1982
1635+267	2.0	?	19	4	2	O	point		Djorgovski et al., 1984
0023+171	1.0	?	23	5	2	R	point		Hewitt et al., 1987
UM425	1.5	~0.6	16	6	4	O	point		Meylan et al., 1989
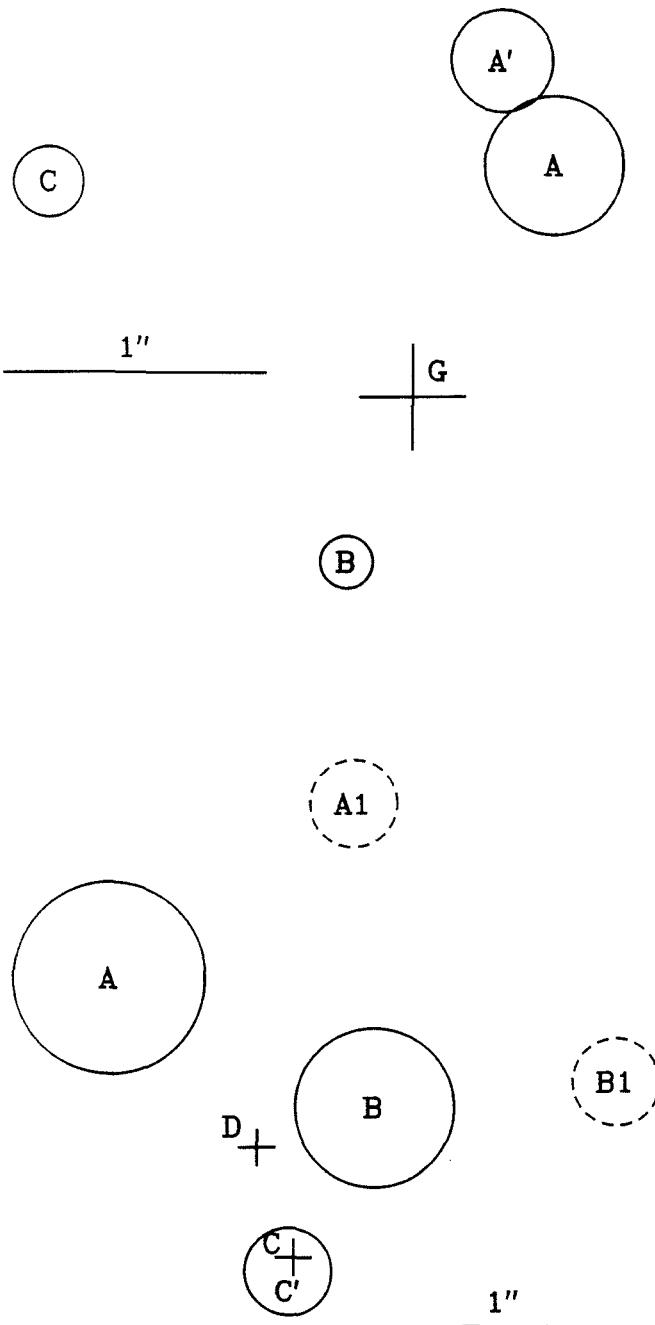
Figure 1. Schematic diagram (top) of the gravitational lens Q1115+080 showing the four images (A,A′,B,C) and the position of the lensing galaxy. The area of the circles surrounding the images indicates their relative fluxes. Images A and A′ are probably merging on a critical line. Schematic diagram (bottom) of the gravitational lens Q2016+112 showing the three images (A,B,C), two emission regions at the same redshift as the images (A1,B1), and the two lensing galaxies (D,C′). The area of the circles surrounding the images indicates their relative fluxes.

Chapter 2

Gravitational Imaging by Isolated

Elliptical Potential Wells

I: Cross Sections

Roger D. Blandford,

and

Christopher S. Kochanek

Theoretical Astrophysics 130-33,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

Abstract

The imaging properties of isolated gravitational lenses are investigated numerically using a model two-dimensional potential parametrized by its hardness, core radius and ellipticity. Cross sections for creating image arrangements with specific characteristics are presented and analyzed. Strong circular lenses can produce bright opposed images on opposite sides of the potential center with an asymptotic cross section for producing a pair of images magnified by more than a factor $\bar{M}_{12}$ of $\simeq A_+ / \bar{M}_{12}^2$ where A_+ is the area of the of the outermost critical circle on the sky. When the ellipticity of the potential exceeds $1/\bar{M}_{12}$ five image configurations become possible and predominate for maximum amplifications $\gtrsim 5$. In this case, highly amplified images are generally found close together; it is not possible to account for widely separated images being highly amplified. The faintest image can be amplified by a factor $\sim f^{-2}$, where f is the central surface density in units of the critical surface density for multiple imaging. Marginal gravitational lenses generally create three amplified images. The time delay associated with bright images separated by r_{12} is proportional to $r_{12}^2 / \bar{M}_{12}$.

1. Introduction

In spite of intensive observational and theoretical effort, not one of the eight claimed examples of multiple imaging of a quasar by a gravitational lens is properly understood. In some cases, (e.g. Q0957+561, Gorenstein *et al.*) one or more galaxies are seen in the field of the quasar and although they must contribute to the light deflection, they cannot account for the observed arrangement of images alone. In other cases, (e.g. Q1635+267, Djorgovski and Spinrad) there is no evidence for any intervening galaxies and indeed doubt has been cast upon the reality of some of these claims (e.g. Shaver and Cristiani, 1986, Phinney and Blandford, 1986). It is clear that studies of multiply imaged quasars, just like those of the dynamics of galaxies and clusters of galaxies, require the presence of sub-luminous material. Observations and analyses of multiple images then take on a special importance as they provide one of the few diagnostics we have of dark matter.

In principle, the images contain quite a lot of information about the gravitational potential well formed by the dark matter. When two images lie on opposite sides of the potential well, their angular separation is a measure of the velocity dispersion of the particles that form the potential well (e.g. Young *et al.*, 1980). If the images are resolved, their relative parities and shears may fix the topology of the arrival time surface and, indirectly, the rough shape of the potential well (e.g. Blandford and Narayan, 1986, henceforth BN). Finally, the curious absence of the odd images from most of the examples we have so far may imply the presence of a singular potential, or at least one that possesses a small core radius. It is a tantalizing prospect that an astronomer can use beams of photons to probe a condensation of dark matter in much the same way that a nuclear physicist uses beams of electrons to study the structure of an atomic nucleus.

In one of the more detailed studies of gravitational imaging, Turner, Ostriker, and Gott, 1983, (henceforth TOG) used a singular isothermal sphere and a point mass to describe the distribution of matter in the lens. However, as several authors have pointed out, this may not be representative of actual lenses in three distinct ways. Firstly, the addition of even a mild degree of asymmetry can have a profound effect on the images (e.g. Nityananda and Ostriker, 1984). In particular, when the potential is sufficiently elliptical in shape, bright images will mostly be found close together straddling a critical line and on the same side of the potential well, rather than on opposite sides as is the case for a strictly circular potential (e.g BN). Secondly, a singular isothermal sphere potential will only produce two images and is quite unrepresentative of a flat bottomed potential well that will usually create a fairly bright odd image. Finally, the fairly "soft" potential well associated with an isothermal sphere, behaves rather differently from the "hard" potential generated by a more concentrated mass distribution; in particular, the former potential is more likely to create five, as opposed to three, images. Quantitative calculations of the rate of occurrence of multiple images, predicated on overly simple assumptions about the form of the intervening matter, can be quite misleading.

Earlier work on the imaging properties of elliptical potentials has either been confined to specific examples tailored to a particular quasar image pair (e.g. Narasimha *et al.*, 1982,1984ab), or constrained by the limitations of analytical calculations to small perturbations of circular potentials (e.g. Subramanian and Cowling, 1986, or Kovner, 1986). Nevertheless, some qualitative understanding of the problem has been gained and several theorems (albeit of limited applicability) have been proven.

In this paper, we describe a numerical study of the imaging properties of a three parameter family of isolated elliptical potential wells. Specifically, we compute the magnification distributions, the relative probabilities of observing three and five images and

their arrangement on the sky relative to the potential well, and the differences in the light travel times. We believe that our models span the range of qualitative behavior to be expected from a single, centrally concentrated mass distribution and although there is some model dependence in our quantitative results, the approximate descriptions of these distributions, which we derive, are of quite general applicability.

Unfortunately, the present work also has some limitations, especially when it comes to analyzing observations in detail. Our models assume that the potential is smooth and we ignore the possibility of micro-lensing of a sufficiently compact source by stars (e.g. Paczyński 1986ab, Schneider 1986abc). We also only consider lenses confined to a single plane; multiple lenses, located at different distances along the line of sight from the observer to the source (e.g. Subramanian and Chitre, 1985), admit an even richer variety of image configurations. Finally, by restricting our investigation to purely elliptical potentials we exclude lenses comprising several galaxies, probably lying in the same group or cluster, and superposed on the sky. Such relatively rare, large amplitude fluctuations in the cosmic density distribution could be responsible for the observed preference for large image splittings. In fact, even had we analyzed a more complete family of potentials, and included these additional effects, we would still not have understood what types of image arrangements should be found in observed systems. Selection effects bias the discovery of gravitational lenses. In particular, amplification bias, the tendency to over represent intrinsically faint but highly magnified quasars, must also be included.

In this paper we only compute cross sections; in an accompanying paper we use these cross sections to attempt to calculate the expected distribution of quasar images if the lenses are described by our model potential. In §2, we describe the model potentials that we use and in §3, we review the geometric optics of a gravitational lens. Cross sections are defined in §4 and described in §5 for strong lenses of varying ellipticity, core radius, and hardness. In §6, a similar study is made of marginal lenses. These potentials are normalized by ensuring that they have the same radius within which the average surface density equals the "critical density" for producing multiple images at the distance of the lens. We then present our results in the form of rough "rules of thumb" that express the variation of image arrangements with potential shape. The distribution of another important observable, the arrival time differences for the different images, is discussed in §7. Our results are collected in the concluding section. The appendix deals with analytical calculations for circularly symmetric potential wells that we use to check the accuracy of our numerical computations and to help understand the results.

2. Model Elliptical Potentials

We use a three parameter class of two-dimensional Newtonian potentials that characterize the properties of isolated galactic and dark matter potentials.

$$\psi(x,y) = \frac{A}{2\alpha}\left[\left(1 + (1-\epsilon)\left(\frac{x}{s}\right)^2 + (1+\epsilon)\left(\frac{y}{s}\right)^2\right)^\alpha - 1\right]. \qquad (2.1)$$

The parameter α determines the *softness* of the potential, spanning the range from an asymptotically isothermal form, which we henceforth refer to as isothermal, ($\alpha = 1/2$) to a Plummer model ($\alpha \to 0$, $\psi \to (A/2)\ln(1 + (1+\epsilon)x^2/s^2(1-\epsilon)y^2/s^2)$. A measures the *depth* of the potential, and s the *radius of the core*. We consider values of the *ellipticity* ϵ between 0 and 0.2. For values of ϵ beyond a critical value $\epsilon_c = \alpha/(1-\alpha)$, these potentials

develop negative surface densities at radii greater than

$$\left(\frac{r}{s}\right)^2 = \left[(1+\epsilon)(1-\alpha)(\epsilon - \epsilon_c)\right]^{-1} \qquad (2.2)$$

beginning along the y-axis for positive ϵ. This behavior will not affect the statistical lensing properties of the potential, that are essentially governed by average quantities which do not behave unphysically. Moreover, as we confine our attention to small $\epsilon \leq 0.2$, the negative surface density region begins outside the region of interest for lensing for almost all of the cases studied. For example, the $\alpha = 0$ case has negative surface density regions for all $\epsilon \neq 0$, yet the lensing properties are nearly the same as for nearby models with $\alpha \sim 0.1$, which do not suffer from the problem. In the limit of small core radius and ellipticity, the ellipticity of the associated surface density contours is a factor of $(1+\alpha)/\alpha \geq 3$ larger than the ellipticity of the surface potential, ϵ. In the limit of large core radius, the surface density ellipticity is $3\epsilon/2$ independent of α.

Real potentials of elliptical mass distributions become more circular with increasing radius. This introduces at least one extra parameter. As most of the multiple imaging events are associated with light rays having small impact parameters, we have found that using a single ellipticity is adequate and avoids additional complication. A further parameter that could be introduced in a more complete study would describe the twisting of equipotential surfaces associated with triaxial three-dimensional potentials which are not viewed along their principal axis.

The lensing properties of the galaxy are closely connected to the *average* surface density $\langle \Sigma \rangle$ of the lens within a given radius. For circular models,

$$\langle \Sigma \rangle(r) = \Sigma(0) \left[1 + \left(\frac{r}{s}\right)^2\right]^{\alpha - 1} \qquad (2.3)$$

where $\Sigma(0) = A/2\pi G s^2$.

3. Imaging By Gravitational Potentials

Consider a ray from the source S to the observer O passing through lens L with impact parameter $\mathbf{r}$ relative to the lens center. Let the same ray have impact parameter $\mathbf{u}$ in the absence of the lens. The general lens equation (eg. Bourassa and Kantowski 1975) becomes

$$\mathbf{u} = \mathbf{r} - \nabla_{\mathbf{r}} \phi(\mathbf{r}) \qquad (3.1)$$

where $\phi = 2 D_{OL} D_{LS} \psi / D_{OS} c^2$, and D_{OL}, D_{LS}, and D_{OS} are angular diameter distances. The images produced by a source with coordinates $\mathbf{u} = (u_x, u_y)$ are found at solutions $\mathbf{r} = (x, y)$ of equation (3.1). The image magnification relative to free propagation in the absence of the lens, denoted by M, is the inverse of the determinant of the transformation $\mathbf{u} \to \mathbf{r}$, and is given by

$$M^{-1} = (1 - \partial_x \partial_x \phi)(1 - \partial_y \partial_y \phi) - (\partial_x \partial_y \phi)^2. \qquad (3.2)$$

The time delay t (and its normalized form τ) are given by

$$t = \frac{D_{OS}\tau}{c} = \frac{(1 + z_L) D_{OS}}{D_{OL} D_{LS} c} \left[\frac{1}{2}(\mathbf{r} - \mathbf{u})^2 - \phi(\mathbf{r})\right] \qquad (3.3)$$

where all quantities are evaluated at the positions of the images.

If the lens is strong enough to produce multiple images of sources with redshift z_S, there will be *caustic lines* in the *source plane* that separate regions associated with different numbers of images. Sources located on these caustics will create infinitely amplified images on *critical lines* in the *lens plane*. For a circularly symmetric potential there will be two critical lines both of which are circles. The smaller circle, with radius r_-, is the locus of pairs of radially merging images produced by sources lying on the radial caustic. The larger circle, at radius r_+, coincides with the ring image produced by a source on the optic axis. This second caustic degenerates to a point in the source plane. If the source is displaced slightly off the axis, two bright images are created on opposite sides of the lens center, one just inside and the other just outside the critical radius. The critical radii satisfy the equations

$$K\left[1 + r_+^2/s^2\right]^{\alpha-1} = 1 \tag{3.4}$$

$$K\left[1 + r_-^2/s^2\right]^{\alpha-2}\left[1 + (2\alpha - 1)r_-^2/s^2\right] = 1 \tag{3.5}$$

where

$$K = \frac{2A D_{OL} D_{LS}}{D_{OS} c^2 s^2}. \tag{3.6}$$

Previous discussions of gravitational lensing (*eg* TOG) have emphasised the importance of the critical surface density, $\Sigma_c = c^2 D_{OS}/4\pi G D_{OL} D_{LS}$, which is the density of a uniform sheet of matter that is just able to focus rays at the observer. Application of Gauss' theorem tells us that the mean density within the radius r_+ is just Σ_c (*cf* equation (2.3)). We choose this radius as a fiducial radius measuring the linear size of the potential. This definition is necessary because when comparing cross sections for potentials with different values of α and s we must normalize the potentials. When we compare different lenses, we shall measure cross sections in units of r_+^2, and distances on the sky in units of r_+. The core size s will be replaced by its renormalized value $\beta = s/r_+$. Similarly, we introduce the normalized time delay

$$\hat{t} = D_{OL} D_{LS} r_+^2 (1 + z_L)^{-1} D_{OS}^{-1} ct. \tag{3.7}$$

The depth of the potential well A, used in equation (2.1), can be expressed in terms of β by

$$A = \frac{D_{OS} c^2 r_+^2}{2 D_{OL} D_{LS}} \beta^2 (1 + \beta^{-2})^{1-\alpha} \tag{3.8}$$

so that

$$\phi = \frac{\beta^2 r_+^2}{2\alpha}(1 + \beta^{-2})^{1-\alpha}\left[\left(1 + (1 - \epsilon)x^2 r_+^{-2}\beta^{-2} + (1 + \epsilon)y^2 r_+^{-2}\beta^{-2}\right)^\alpha - 1\right] \tag{3.9}$$

We emphasize that our choice of normalization is necessarily arbitrary. A different normalization may be preferable for some applications, and an elementary rescaling must be performed. Because there is no simply defined mean surface density for an elliptical potential, we will continue to use the value of r_+ from equation (3.4). This is justified because it is still the mean radius of the outermost critical line to first order in ϵ and we are only interested in small ϵ. Multiple images will be found with equation (3.9) provided

that β is real. Models with $\beta \gtrsim 1$ are only marginally dense enough to accomplish this, however, as their central surface density is just larger than the critical surface density by a factor $\sim (1-\alpha)\beta^{-2}$. Galaxies at $z_L \sim 0.5$ have $\beta \lesssim 0.1$ and make strong lenses for sources at $z_S \gtrsim 1$. However, for clusters of galaxies, multiple imaging is at best marginal, $\beta \gtrsim 1$ (eg Narayan, Blandford, and Nityananda, 1986, Kovner, 1986). As discussed in §2, Plummer models with finite ellipticity and small core radii develop negative surface densities. The condition that the surface density remain positive within the critical contour is $\beta > \sqrt{\epsilon(1+\epsilon)}$. Although a few of our models violate this constraint, we emphasize that none of our conclusions are in fact sensitive to the surface density distribution.

4. Cross Sections

Although some analytical results have been given for simple, circularly symmetric potentials, it is necessary to compute most cross sections numerically. We use a 'backwards ray tracing' algorithm to follow rays from the observer, through the lens plane, to the source. The rays corresponding to a system of multiple images in the lens plane are traced back to their origin in the source plane. In practice, the lens and source planes are divided into N by N grids where N is 200, 500, or even 1000, depending on the resolution required. The lensing equation (3.1) is then evaluated by finite differencing, and is used to project triangles (half of a grid box) from the lens to the source plane. (Triangles are used because their projection is always convex–any other object may have a concave projection near caustic lines.) All source plane grid points inside the projected triangle are associated with an image in the lens plane at one of the vertices of the triangle. Multiple images are made when the projections of triangles from different areas of the lens plane overlap on the source plane. The amplification is computed as the average amplification over one grid box by taking the ratios of the areas of the triangles and its projection. Although the method necessarily satisfies the odd image number theorem (eg. Burke, 1981), some care must be taken to avoid the generation of spurious pairs of images.

Once the map of grid points in the source plane onto the image plane is known, all other quantities can be found by direct calculation. We have evaluated the bright image separation, the distance from the faintest image to the origin, the amplification of the faintest image, the amplification ratios, the time delay, the merger angle of the two bright images relative to the origin, and the angle between the brightest and faintest images. Each lens calculation consists of the image positions for each of the N^2 positions on the source plane grid. If the lens model is characterized by total cross section σ we can compute the cross section for a particular image property provided it has total cross section $\gtrsim 10\sigma N^{-2}$ and image separations $\gtrsim \sqrt{\sigma} N^{-1}$.

The cross section for some property (eg. creation of five images), is also estimated using

$$\sigma = \int d^2 u = \int \frac{d^2 r}{M} \qquad (4.1)$$

The first integral is carried out over all sources that produce images of the specified type and the second integral must be integrated over only one image per source. The cross section for large amplification is too small to be computed accurately on a grid of this size and we use an asymptotic formula obtained by locating all the critical lines in the lens plane and then expanding about this line. If we write $H = M^{-1}$, then, by definition, $H = 0$ on the critical line and the cross section for amplification of two images, both amplified by more than $|M|$, is given asymptotically by a line integral around all the

critical lines

$$\sigma(\bar{M}_{12} > M) = \frac{1}{2M^2} \int_{H=0} \frac{ds}{|\nabla H|} \tag{4.2}$$

(Blandford and Narayan 1986). These asymptotic formulae match well onto the numerical calculations. A similar problem arises when the potential well is deep and one image is located near its center and is consequently highly deamplified. We will use equation (5.6) below in this case.

5. Magnification and Arrangement of Images: Strong Lenses

It is simplest to discuss the image configurations in two limiting cases. In this section, we assume that the central surface density is substantially greater than critical ($\beta \leq 1$). In the following section we assume $\beta \geq 1$.

5.1. The Singular Isothermal Sphere

The simplest model of a gravitational lens, explored in some depth by TOG is a singular isothermal sphere. In our notation this is described by a potential with $\alpha = 0.5$ and $\epsilon = 0$ in the limit $\beta \to 0$. This case is in many respects atypical of generic transparent potentials. It only creates two images and these lie on opposite sides of the potential. The caustic associated with tangential mergers and the critical curve associated with radial mergers both degenerate to a point at the origin. Bright images are created when the source lies close to the optic axis; when it lies on the axis, a structurally unstable ring image is created.

When the source lies behind the critical circle of radius r_+ with projected radius $u \leq r_+$ the two images are located at radii $r_{1,2} = r_+ \pm u$ and the (unsigned) amplifications are $M_{1,2} = r_+/u \pm 1$. The mean amplification is $\bar{M}_{12} = r_+/u$ and the cross section for producing two images with mean amplification $\bar{M}_{12} > M$ is then $\sigma(\bar{M}_{12} > M) = \pi r_+^2/M^2$ The second image disappears when the source crosses the radial caustic at $u = r_+$ and at this point, $M_1 = 2$, $M_2 = 0$ so that $\bar{M}_{12} = 1$ and the total 2 image cross section is πr_+^2.

5.2. Influence of a Hard Potential

An isothermal potential with a constant velocity dispersion (appropriate to the outer parts of galaxies) is the softest type of potential conventionally considered for self-gravitating bodies (but see Blandford, Phinney, and Narayan, 1986). We now consider the effect of *hardening* the potential by allowing the exponent α to decrease from $\alpha = 1/2$ to $\alpha = 0$. For the moment we keep the potential singular ($\beta \to 0$).

The bending angle is now $r_+^{1-2\alpha} r^{2\alpha-1}$ and diverges at the origin unless $\alpha = 1/2$. Image amplifications are related to their positions through

$$M = \left[1 + (1 - 2\alpha)\left(\frac{r}{r_+}\right)^{2\alpha-2} \right]^{-1} \left[1 - \left(\frac{r}{r_+}\right)^{2\alpha-2} \right]^{-1} \tag{5.1}$$

As long as the potential remains singular, all sources are doubly imaged but the second image becomes increasingly faint with increasing impact parameter,

$$M \sim \frac{1}{(1 - 2\alpha)} \left(\frac{u}{r_+}\right)^{4(1-\alpha)/(1-2\alpha)} \qquad \alpha \neq 1/2 \tag{5.2}$$

Formally this implies an infinite total cross section for the generation of multiple images. This is very misleading because the ratio of the faint to bright image amplification is rapidly converging to zero, and the "brightest" image amplification is rapidly converging to one as the impact parameter increases beyond r_+.

$$\frac{M_2}{M_1} = 1 - \frac{1 + 2(1 - \alpha)(1 - 2\alpha)}{4(1 - \alpha)^2} \frac{2}{\bar{M}_{12}} \qquad \bar{M}_{12} \gg 1 \tag{5.3}$$

The physically interesting cross section is not the total cross section, but the cross section for the average amplification of the two images, $\bar{M}_{12}$, to be greater than some value M,

$$\sigma(\bar{M}_{12} > M) \simeq \frac{\pi r_+^2}{4(1 - \alpha)^2} \frac{1}{M^2}, \qquad M \gg 1. \tag{5.4}$$

The physical multiple image cross section, in the sense of what it is possible to observe, (as compared to the formal but divergent total cross section) is equation (5.3) evaluated at a fiducial amplification M. From this point of view, the hardness of the potential has a much weaker effect on the total cross section (5.3) because the bright image cross section for $\bar{M}_{12} > 1$ decreases only by a factor of $1/4(1 - \alpha)^2 \geq 1/4$.

5.3. Implications of a Finite Core Size

Adding a finite core radius to the potential ensures that the total multiple image cross section is finite and introduces a third image located near the core. This also creates a second circular caustic and an associated critical circle on which pairs of images produced by sources well away from the optic axis can merge radially and be infinitely amplified. We will refer to two bright images that lie on the same side of the potential as *allied* images, and two bright images that lie on opposite sides of the potential as *opposed* images. Allied images merging parallel to the radial direction will be call *radially merging*, and those merging perpendicular to the radial direction will be called *tangentially merging*. We introduce the distinction now, although circular potentials can generate only radially merging allied images. (See Figure 1a,b,c).

Although the total cross section is now finite for all values of α, it still misrepresents the physical cross section. The total cross section $\sigma \sim \beta^{-2+4\alpha} r_+^2$ (see the Appendix, equation (A.25)), while (5.3) describes the opposed image cross section, $\sigma(\bar{M}_{12} > M)$. Together these give an estimate of the *minimum* average amplification of the *brightest* two images. For the $\alpha = 1/2$ isothermal case, it is $\bar{M}_{12} \gtrsim 1$, but for the $\alpha = 0$ Plummer model most of the cross section is for creating at least one highly deamplified image (see Figures 2a-d).

We can also give analytic expressions for the asymptotic cross sections for bright allied images associated with the radial caustic when the core radius is small, (i.e., when $\beta << 1$). In all cases, the critical circle C_- is located at $r_- \simeq s$ with the image magnification

$$M \sim \beta^{5-4\alpha} r_+/r_{12} \qquad \text{if } \alpha \neq 1/2 \tag{5.5}$$

where r_{12} is the bright image separation (see Figure 3). The integral cross section for producing two allied images with M_{12} in excess of M is

$$\sigma(\bar{M}_{12} > M) = \frac{\pi r_+^2}{M^2} \begin{cases} 4\beta^6 & \alpha = 0 \\ 3^{5/2}\beta^5 & \alpha = 1/4 \\ \beta^2/3 & \alpha = 1/2 \end{cases} \tag{5.6}$$

Inspecting these cross sections, we conclude that if the core radius is small ($\beta \ll 1$), then the cross sections for bright allied images (all of which merge radially), are ignorably small compared with those for opposed images as computed in §5.2. This difference is most marked for hard potentials because they can deflect rays through a much larger angle. The third opposed image is located within the core, ($r \lesssim s$), there is no shear, and the amplification is given by

$$M_3 \sim \beta^{4(1-\alpha)} = \left(\frac{\Sigma(0)}{\Sigma_{\text{crit}}}\right)^2, \qquad \beta \ll 1 \tag{5.7}$$

Here, we see an important difference between hard and soft potentials; the former create fainter third images for a given core size because they have larger central surface densities (see Figure 4). These asymptotic formulae, which are discussed further in the Appendix, are only valid for $\beta \lesssim 0.3$. They have been used to test the numerical method.

5.4. Elliptical Potentials

The character of the images changes again when the circular symmetry of the potential is broken. If we start with a circularly symmetric, multiply-imaging potential and increase the ellipticity ϵ from zero, a small four cusped caustic will be created around the optic axis. When the source lies within this caustic, five images (two saddle points and two minima lying in a ring around a central maximum in the arrival time surface) will be produced.

It is of interest to ask how elliptic the potential need be in practice for there to be a qualitative change in the image properties. We can provide a partial answer to this question by considering a circular potential with $\beta < 1$. When the impact parameter **u** of the source is small, two opposed images will be created close to the outer critical circle of radius r_+. A simple calculation reveals that the normalized arrival time difference between these two images is $\Delta \hat{t}_{12} = 1/(1-\alpha)M$ (*cf.* Appendix). Now, if we allow the potential to become elliptic, an azimuthal perturbation $\Delta\phi \sim \epsilon r_+^2$ will be imposed over the arrival time surface near the outer critical circle. When $\Delta \hat{t}_{12} \sim \Delta\phi$, the opposed images will be displaced, and two new images created. The condition for validity of results based on circular potentials is roughly that the bright image amplification $\bar{M}_{12}$ not exceed $1/(1-\alpha)\epsilon$. Numerical calculations verify this rule.

Surrounding the inner caustic will be a region in which only three images are created. This three image region will itself be bounded by an elliptical second caustic associated with radially merging images as discussed in §5.3. Sources that lie outside this second caustic will only create one image. As we increase the ellipticity, the size of the inner, cusped caustic will increase and it will eventually expand beyond the other caustic. Hence cusps can be associated with three images as well as five images.

The two brightest images will always straddle one of the two critical lines in the lens plane. The inner (outer) critical curve C_- (C_+) associated with the outer (inner) caustic in the source plane is associated with the merger of the brightest image pair in the five-image (three-image) configuration. The faintest image I_5 (I_3) is located near the core of the potential, inside of C_-, with amplification given roughly by (5.6). In the five image case, the middle images, I_3 and I_4 are located between the critical curves, and outside C_+ respectively. For allied image mergers (closely straddling a critical line), we can subclassify the mergers by whether they are merging radially or tangentially. We expect most allied three image cases to merge radially as they are associated with the

limaçon topology time delay surface. Allied five image cases will merge tangentially in most cases because they are associated with the lemniscate time delay contours formed inside the limaçon contour (*cf* BN).

5.5. *Results of Numerical Computations*

In order to understand the types of images to be expected most frequently from strong isolated elliptical potentials, we have computed cross sections and tabulated image properties for 27 models with all combinations of $\alpha = 0, 0.25, 0.5; \beta = 0.03, 0.1, 0.3; \epsilon = 0, 0.1, 0.2$. These models cover the volume of parameter space for which the analytical formulae presented in the Appendix in the limit $\beta \ll 1$ are essentially correct for the circular potentials.

In Table 1, we display cross sections and *branching ratios* for model isothermal potential wells. Total multiple imaging cross sections and cross sections for $\bar{M}_{12} > 10$ are given. The branching ratios measure the fraction of the total cross section associated with opposed image configurations, allied image configurations and five image configurations. Table 2 repeats this information for Plummer potential models. In Figures 2a-d, we display the integral cross sections for average magnification greater than some value M for four circularly symmetric potentials and find that opposed images predominate. The asymptotic cross sections agree well with both the estimate derived by integrating around the critical curves and the analytical expressions described in the Appendix. In particular the inverse square law for the integral cross section (equation (4.5)), is satisfied. These circular cross sections can be contrasted with the elliptical cross sections displayed in Figures 5a-d, which show that if we require that two of the images are amplified by more than ~ 3, five images are more likely to be produced than three images.

We have also computed the amplifications of the faintest images for a range of models and these results are given in Table 3. As the eccentricity is increased and the inner (3 to 5) caustic expands away from the origin, the behavior of the faint image changes. As with the circular potentials, the position of minimum amplification is located at the center of the potential (for strong potentials), but it is no longer associated with images of sources lying very near caustic lines. For low amplification, the faintest image I_5 approaches the origin and a limiting magnification of M_{min}. As the source approaches the three to five caustic, I_5 moves away from the origin and brightens (see Figure 8).

The properties of opposed images are similar to those found for circular potentials. However, the minimum distance of the faint image I_3 from the origin occurs at the three to five transition, where it asymptotes to the position and amplification of the faintest of the five images just before the merger of I_1 and I_2 occurs. The faint allied image will only be affected by the deformations of the C_- critical line, which is the image of the outer elliptical caustic and is therefore not as severely deformed by the introduction of ellipticity. (This is not true if you make the ellipticity very large, and the cusped caustic interpenetrates or surrounds the elliptical caustic.) The positions and amplifications of the radially merging allied images are not strongly affected by ellipticity. Tangentially merging allied images, however, only arise in the presence of ellipticity.

The ratio of the amplification of the second brightest image M_2 to the amplification of the brightest image M_1 for the three image configurations has the general form of (5.3): a function of the form $1 - a\bar{M}_{12}^{-1}$ where $a \sim 1$ for $\beta \ll 1$. This is borne out by the numerical results even in the elliptical case for images that merge radially and hence depend on radial gradients of the amplification that are only weakly affected by the ellipticity. As a general rule, at $\bar{M}_{12} \sim 1$ the ratio is $\sim .3$, and at amplifications ~ 10 it is $\sim .8$. The

five-image case behaves differently because they merge tangentially and depend on the tangential gradients of the amplification. In general these should be much weaker than the radial gradients (by factors of order ϵ) so we expect that the ratio should be much closer to one even at low amplifications. This is born out by the simulations: at $\bar{M}_{12} \sim 1$, $M_2/M_1 \sim .8$ and it then converges to 1 with the $\bar{M}_{12}^{-1}$ behavior predicted by the circular results.

6. Magnification and Arrangement of Images: Marginal Lenses

In this section we investigate lenses that are only just strong enough to create multiple images ($\beta \geq 1$).

6.1. Qualitative Features

In contrast to the strong lenses discussed in §5, marginal lenses are quite insensitive to α as we are only interested in the approximately quadratic potential variation near the core. Again, let us fix our ideas by considering circular potentials. The total cross section is given by (see Appendix)

$$\sigma = \frac{4\pi r_+^2 (1 - \alpha)^2}{27} \beta^{-4}. \tag{6.1}$$

The ratio of the total cross section for isothermal and Plummer models varies by a factor of only four (cf a β^2 dependence for $\beta \ll 1$). The cross section for producing three images with mean amplification greater than M is

$$\sigma(\bar{M}_{12} > M) = \frac{\pi}{4(1 - \alpha)^2} \frac{\beta^4}{M^2} \tag{6.2}$$

in both the opposed and allied cases associated with the outer and inner critical lines respectively. The magnification of the third image is given by

$$M_3 \simeq \frac{\beta^4}{(1 - \alpha)^2} \tag{6.3}$$

in both cases. The opposed and allied images are equally probable, and the third image is strongly magnified, again in contrast to $\beta \ll 1$.

Now let us introduce some ellipticity, which permits five images to be created. The five-image cross section initially increases with ϵ (at fixed α and β). However, when

$$\epsilon \gtrsim 1 - (1 + \beta^{-2})^{\alpha - 1} \tag{6.4}$$

the lens is only able to focus the source along one direction, and only three-image geometries are possible. (This is the lemniscate case of BN.) In this case, the brightest images are exclusively allied and radially merging.

6.2. *Numerical Results*

We have computed cross sections and tabulated image profiles for twelve marginal isothermal and Plummer potentials. A separate investigation verifies that our results are not sensitive to α as asserted in §6.1. We adopt ellipticities $\epsilon = 0$, 0.1, and 0.2 and dimensionless core radii $\beta = 1$, and 3. Cross sections for the larger value of β are well approximated by the asymptotic formulae developed in §6.1 and the Appendix. The results are tabulated in Tables 1 and 2 and displayed in Figure 9. We find that in a circular potential with $\beta = 3$, allied and opposed bright images are equally common (in agreement with equation (6.2)). For the cases $\epsilon = 0.1$ and 0.2 and $\beta = 3$, *only* bright, radially merging, allied images are created, (which is expected from equation (6.4)). The total cross section is substantially greater than for the purely circular potential. The five-image cross sections are completely suppressed in the numerical results for which (6.4) is satisfied.

7. Time Delays

For circular potentials, the normalized time delay between the two brightest images can be calculated by expanding about the appropriate caustic (see the Appendix). For opposed images,

$$\hat{t}_{12} \simeq \frac{r_{12}^2}{r_+^2 \bar{M}_{12}} \frac{1 + \beta^2}{4(1 - \alpha)}, \tag{7.1}$$

while for allied images,

$$\hat{t}_{12} \simeq \frac{r_{12}^2}{2 r_+^2 \bar{M}_{12}} \begin{cases} \begin{cases} \dfrac{1}{3} \beta^{2/3} & \alpha = 1/2 \\[2ex] \dfrac{1}{3} \left(\dfrac{1 - 2\alpha}{2(1 - \alpha)} \right)^{\alpha - 1} \beta^{2(1-\alpha)} & \alpha \neq 1/2 \end{cases} & \beta \ll 1 \\[4ex] \dfrac{\beta^2}{2(1 - \alpha)} & \beta \gg 1 \end{cases} \tag{7.2}$$

As can be seen in Figure 10, where a normalized histogram of $dN/d\hat{t}_{12}$ is plotted, the distribution is very sharply peaked about the asymptotic value, with a tail trailing off towards larger values. While it is surprising that the distribution is so compact for the opposed images, (after all, (7.1) and (7.2) are derived for large amplification events in circular lenses) it will still contribute an error of roughly 10 % to any time delay measurement in addition to errors in determining β, α, and $\bar{M}_{12}$ observationally. The five-image time delay distribution is very broad because of the variations in the tangential derivatives of the magnification along the critical line.

8. Conclusions

In this investigation, we have computed the cross sections for different types of multiple imaging events involving a single elliptical lens described by three parameters that measure its hardness, core radius, and ellipticity. We have found that numerical evaluation can give accurate calculations for image magnifications in the range $0.005 \lesssim M \lesssim 300$. The measured cross sections converge to the asymptotic values calculated at higher magnification and, in the case of circular potentials for which full analytical expressions are available (see the Appendix), have errors consistent with those expected from the finite

grid spacing. Cross sections can only be estimated reliably as long as several discrete grid points are associated with the image arrangement whose probability is being measured. However, as gravitational lensing is itself an intrinsically rare phenomenon, we have no interest in highly improbable lensing configurations and so we believe that our computations are sufficiently accurate for application to quasars. Our choice of potential is necessarily arbitrary because we are a long way from being able to infer the mass distribution in any observed lensing event. So, it is only the general features of our calculations that are likely to be useful. Fortunately these can be summarised in a few "rules of thumb."

For isolated, centrally-concentrated potentials we find that:

1. The cross section for multiple imaging in which the two brightest images have a mean magnification in excess of $\bar{M}_{12}$ is roughly $\sigma(\bar{M}_{12} > M) \sim A_+(1 + \beta^4)/4(1 - \alpha)^2 M^2$ for all β, where A_+ measures the area on the sky of the outer critical line, and α ranges from 1/2 (Isothermal) to 0 (Plummer) as the potential hardens. This formula appears to be good to about 20% for $\bar{M}_{12} \gtrsim 10$. The total cross section is only a useful quantity for very soft potentials for which it is roughly A_+.

2. Although smooth, strong, circular potentials can only create three images, five-image configurations are common if the symmetry is broken and the potential is slightly elliptic as we generally expect to be the case. When $\bar{M}_{12} \lesssim (1 - \alpha)^{-1}\epsilon^{-1}$, three images of which the brightest two are opposed are nearly always created. When $\bar{M}_{12} \gtrsim (1 - \alpha)^{-1}\epsilon^{-1}$, five images (with the brightest pair usually merging tangentially) are generally produced. For strong lenses with $\bar{M}_{12} \gtrsim 10$ the five-image cross section is roughly five times the three-image cross section.

3. If we restrict attention to the three-image cases, opposed images greatly outnumber allied images, except when the potential is only marginally capable of producing multiple images (i.e., $\beta \gtrsim 1$). The allied images that do exist are displaced radially from each other, and straddle the inner critical line.

4. Amplification bias due to highly magnified but intrinsically faint quasars is easily detected for elliptical potentials. Because the five-image and three-image allied cross sections dominate high magnification events, any large biasing effects will result in anomalously high numbers of these events relative to the three-image opposed lenses, which dominate the low amplification cross section. Similarly, the bright image splittings for allied events are much smaller than for opposed events, so that strongly biased samples will have anomalously low bright image separations.

5. The magnification of the faintest image is controlled by the size of the core. When the core is small, it is $\sim \beta^{4(1-\alpha)} \sim (\Sigma(0)/\Sigma_{\mathrm{crit}})^2$, and will be unobservably faint for $\beta \lesssim 0.1$. When β is large the amplification is $(1 - \alpha)^{-2}\beta^4$ so that the faint image is nearly as bright as the two brightest images. In five-image cases, the third and fourth images are usually at least 5 times fainter than the two brightest images at $\bar{M}_{12} \gtrsim 10$.

6. The magnification of the second brightest image M_2 relative to the magnification of the brightest image M_1 is a sensitive measure of $\bar{M}_{12}$ for three-image opposed and radially allied configurations. This is true for a range of magnifications in both the strong and marginal limits (for the strong lenses, $1 \lesssim \bar{M}_{12} \lesssim 10$ and for the marginal lenses $10 \lesssim \bar{M}_{12} \lesssim 100$). In each case the ratio M_2/M_1, has the value $\simeq 0.2$ at the lower amplification, and $\simeq 0.9$ at the upper amplification. The approximate form of the decay with increasing amplification is $1 - a\bar{M}_{12}^{-2}$ with $a \sim 1$. Tangentially allied configurations, such as almost all five-image cases, do not have this property: for $\bar{M}_{12} \simeq 1$, $M_2/M_1 \simeq 0.8$ for strong lenses. The results, however, are strongly dependent on the model in use.

7. The time delay between the two bright images can be approximated by

$$t_{12} \simeq \frac{r_{12}^2}{\bar{M}_{12}} \frac{D_{OS}(1 + z_L)}{c D_{OL} D_{LS}}.$$

It is doubtful that the Hubble constant will ever be measured reliably by these means.

These results have immediate, though not necessarily original, implications for modelling observed gravitational lens candidates:

The double quasar $0957 + 561$ is believed to be imaged by a giant elliptical galaxy and an associated cluster. The potential is almost certainly elliptical, which implies that as only two images are seen, their magnification is unlikely to be very large ($\bar{M}_{12} \lesssim 10$). The third image is presumably located in the core of the galaxy and can be rendered invisible if the core radius is $\lesssim 1$ kpc. (See Gorenstein *et al.*, 1984, Greenfield *et al.*, 1980, Narasimha *et al.*, 1984a, Young *et al.*, 1980, 1981b.)

In the quasar $1115 + 080$, four images have been observed. These are most probably produced by an elliptical potential. The separation of the A images is roughly one-half of the radius of the potential and the amplification ratio is reported to be $\simeq 0.9 \pm 0.2$ (Foy *et al.*). This implies that the mean magnification of the A images is ~ 20 and that of B,C is ~ 4. The missing fifth image can be deamplified by a modest galactic core size. (See Foy *et al.*, 1985, Shaklan *et al.*, 1986, Young *et al.*, 1981a)

The observed image configuration of $2016 + 112$ has not been reproduced in any single screen models. This is not surprising because there appear to be two galaxies present. Nevertheless, our attempts to model the system with two superposed potential wells also failed. We suspect that the two galaxies are at quite different redshifts so that the single lens approximation is invalid. (See Schneider, D. *et al.*, 1985,1986.)

The quasar $2237 + 0305$ is observed to give two and possibly three images behind a $z_L = 0.04$ Zwicky galaxy. This is likely to be a marginally imaging case, and it is possible that we are observing a lemniscate arrival time surface. (See Huchra *et al.*, 1985.)

The importance of the calculations that we present in this paper is that they allow one to estimate the probability of a particular lens model being realized in practice. In the accompanying paper, we use these cross sections in integrations over source redshifts and luminosity distributions together with lens redshift in order to compute the expected distribution of multiple image arrangements from a potential well of specified form.

Acknowledgments

Appendix: Analytic Results for Circular Potentials

For circular potentials the lens equation simplifies to the polynomial expression

$$\pm \hat{u} = \hat{r} - \hat{r}\left[\beta^{-2} + 1\right]^{1-\alpha}\left[1 + \hat{r}^2\beta^{-2}\right]^{\alpha-1} = f(\hat{r}) \qquad (A.1)$$

$$\rightarrow \hat{r} - \hat{r}^{(2\alpha-1)} \qquad \beta \ll 1 \qquad (A.2)$$

with $\hat{r}$ the ray position, and $\hat{u}$ the source position in the lens plane relative to the center of the potential in units of r_+. All images are on the line connecting the source and the origin, with the exception of the degenerate result for $\hat{u} = 0$ when there is a ring image. The $+$ branch gives the single image on the same side of the potential as the source, and the $-$ branch gives either two or no solutions on the opposite side of the potential. We will use the following notation: $f_{\hat{r}}$ and $f_{\hat{r}\hat{r}}$ denote first and second radial derivatives of the function f defined in (A.1) and $f_{\hat{r}}^+$, $f_{\hat{r}}^-$, and $f_{\hat{r}}^0$ denote the derivative evaluated at $\hat{r}_+$, $\hat{r}_-$ and 0 respectively. Similarly for the higher order derivatives. The amplification for these potentials is given by

$$M^{-1} = \frac{f f_{\hat{r}}}{\hat{r}} \qquad (A.3)$$

$$= 1 - 2\left[\beta^{-2} + 1\right]^{1-\alpha}\left[1 + \hat{r}^2\beta^{-2}\right]^{\alpha-2}\left[1 + \alpha\hat{r}^2\beta^{-2}\right]$$

$$+ \left[\beta^{-2} + 1\right]^{2(1-\alpha)}\left[1 + \hat{r}^2\beta^{-2}\right]^{2(\alpha-2)}\left[1 + 2\alpha\hat{r}^2\beta^{-2} + (2\alpha - 1)\hat{r}^4\beta^{-4}\right]. (A.4)$$

For strong ($\beta \ll 1$) potentials, this reduces to

$$M^{-1} \rightarrow \left[1 - \hat{r}^{2(1-\alpha)}\right]\left[1 + (1 - 2\alpha)\hat{r}^{2(1-\alpha)}\right]. \qquad (A.5)$$

Opposed Images

The tangential, or opposed image, caustic degenerates to a point on the optic axis of the lens ($u \rightarrow 0$). For sources near this caustic, the two bright images have a roughly constant separation of

$$\hat{r}_{12} \simeq 2 \qquad (A.6)$$

to $O(1/M^2)$. The faint third image is still located near the origin. By performing a Taylor expansion about $\hat{r} = 1$ and $\hat{u} = 0$, we obtain

$$\hat{r}_{FO} \simeq \frac{1}{M}\frac{1}{f_r^+ f_r^0} \qquad (A.7)$$

$$\simeq \frac{1}{M}\frac{1+\beta^2}{2(1-\alpha)}\frac{1}{\left[\beta^{-2} + 1\right]^{1-\alpha} - 1} \qquad (A.8)$$

$$\rightarrow \begin{cases} \dfrac{1}{M}\dfrac{\beta^{2(1-\alpha)}}{2(1-\alpha)} & \beta \ll 1 \\[4mm] \dfrac{1}{M}\dfrac{\beta^4}{2(1-\alpha)^2} & \beta \gg 1. \end{cases} \qquad (A.9)$$

The amplification of the faint image at the origin remains approximately constant as the bright images near the critical line,

$$M_3 \simeq M(0) = \left[\left[\beta^{-2} + 1 \right]^{1-\alpha} - 1 \right]^{-2} \qquad (A.10)$$

$$\rightarrow \begin{cases} \beta^{4(1-\alpha)} & \beta \ll 1 \\ \beta^4 (1-\alpha)^{-2} & \beta \gg 1. \end{cases} \qquad (A.11)$$

which results in amplification of the 'faint' image in the weak potential limit ($1 < \left[\beta^{-2} + 1 \right]^{1-\alpha} < 2$), and deamplification for strong potentials. The ratio of the amplification of the second brightest image to the amplification of the brightest image is calculable from a second order expansion of the amplification about the critical line,

$$\frac{M_2}{M_1} = \begin{cases} 1 - \dfrac{1 + 2(1-\alpha)(1-2\alpha)}{4(1-\alpha)^2} \dfrac{2}{M} & \beta \ll 1 \\[4mm] 1 - \dfrac{\beta^4}{4(1-\alpha)^2} \dfrac{2}{M} & \beta \gg 1. \end{cases} \qquad (A.12)$$

The cross section for amplification of the bright pair of images above the average value of M is

$$\hat{\sigma}(\bar{M}_{12} > M) \simeq \frac{\pi}{M^2} \frac{1}{(f_{\hat{r}}^+)^2} \qquad (A.13)$$

$$\simeq \frac{\pi}{M^2} \frac{(1 + \beta^2)^2}{4(1-\alpha)^2} \qquad (A.14)$$

$$\rightarrow \begin{cases} \dfrac{\pi}{4(1-\alpha)^2} \dfrac{1}{M^2} & \beta \ll 1 \\[4mm] \dfrac{\pi}{4(1-\alpha)^2} \dfrac{\beta^4}{M^2} & \beta \gg 1. \end{cases} \qquad (A.15)$$

The normalized time delay between the two brightest opposed images is estimated by

$$\hat{t}_{12} \simeq \left[\frac{1}{2} \frac{\hat{r}_{12}^2}{M} \right] \frac{1}{f_{\hat{r}}^2} \qquad (A.16)$$

$$\simeq \left[\frac{1}{2} \frac{\hat{r}_{12}^2}{M} \right] \frac{1 + \beta^2}{2(1-\alpha)}. \qquad (A.17)$$

Allied Images

The radial, or allied image, caustic is located where the radial magnification diverges. This implies

$$0 = f_{\hat{r}}^- = 1 - \left[\beta^{-2} + 1 \right]^{1-\alpha} \left[1 + \hat{r}_-^2 \beta^{-2} \right]^{\alpha-2} \left[1 + (2\alpha - 1)\hat{r}_-^2 \beta^{-2} \right]. \qquad (A.18)$$

This has exact analytic solutions for $\alpha = 0, 1/2$, which are

$$\hat{r}_- = \begin{cases} \beta \left[(\beta^{-2} + 1)^{1/3} - 1 \right]^{1/2} & \text{if } \alpha = 1/2 \\ \dfrac{\beta}{\sqrt{2}} \left[\sqrt{(3 + \beta^{-2})^2 + 4\beta^{-2}} - (3 + \beta^{-2}) \right]^{1/2} & \text{if } \alpha = 0, \end{cases} \qquad (A.19)$$

but for general α it has only approximate solutions for strong and weak lenses:

$$\hat{r}_- \simeq \begin{cases} \begin{cases} \beta^{2/3}(1 - \beta^{2/3})^{1/2} & \alpha = 1/2 \\ \dfrac{\beta}{\sqrt{1 - 2\alpha}} \left[1 - 2^{2-\alpha}(1 - \alpha)^{2-\alpha}(1 - 2\alpha)^{\alpha - 2}\beta^{2(1-\alpha)} \right]^{1/2} & \alpha \neq 1/2 \end{cases} & \beta \ll 1 \\ \dfrac{1}{\sqrt{3}} & \beta \gg 1. \end{cases}$$

$$(A.20)$$

The bright image separation can be found from a Taylor expansion about the critical line at $\hat{r}_-$,

$$\hat{r}_{12} \simeq \frac{2}{M} \frac{\hat{r}_-}{f - f_{\overline{rr}}} \qquad (A.21)$$

$$\simeq \begin{cases} \begin{cases} \dfrac{2}{3} \dfrac{\beta^{4/3}}{M} (1 - \beta^{2/3})^{-3/2} & \alpha = 1/2 \\ 2^{3-2\alpha}(1 - \alpha)^{3-2\alpha}(1 - 2\alpha)^{2\alpha - 7/2} \dfrac{\beta^{5-4\alpha}}{M} & \alpha \neq 1/2 \end{cases} & \beta \ll 1 \\ \dfrac{1}{M} \dfrac{\sqrt{3}}{2} \dfrac{\beta^4}{(1 - \alpha)^2} & \beta \gg 1. \end{cases} \qquad (A.22)$$

The cross section for amplification greater than M_0 is

$$\hat{\sigma}(\bar{M}_{12} > M) \simeq \frac{\pi \hat{r}^2}{M^2} \frac{1}{f - f_{\overline{rr}}} \qquad (A.23)$$

$$\simeq \begin{cases} \begin{cases} \dfrac{\pi}{3} \dfrac{\beta^2}{M_0^2} (1 - \beta^{2/3})^{-1} & \alpha = 1/2 \\ \pi 2^{2-2\alpha}(1 - \alpha)^{3-2\alpha}(1 - 2\alpha)^{2\alpha - 4} \dfrac{\beta^{6-4\alpha}}{M_0^2} & \alpha \neq 1/2 \end{cases} & \beta \ll 1 \\ \dfrac{\pi \beta^4}{4(1 - \alpha)^2} \dfrac{1}{M^2} & \beta \gg 1. \end{cases} \qquad (A.24)$$

The total three-image cross section is given by $\hat{\sigma} = \pi \hat{u}_-^2$ where $\hat{u}_- = f^-$ is the radius of the allied image caustic,

$$\hat{\sigma} = 4\pi (1 - \alpha)^2 r_-^6 \beta^{-4} \left[1 + (1 - 2\alpha)\hat{r}_-^2 \beta^{-2} \right]^{-2} \qquad (A.25)$$

$$\simeq \begin{cases} \pi 2^{2(\alpha - 1)}(1 - \alpha)^{2(\alpha - 1)}(1 - 2\alpha)^{1 - 2\alpha}\beta^{-2 + 4\alpha} & \beta \ll 1 \\ \dfrac{4\pi(1 - \alpha)^2}{27} \beta^{-4} & \beta \gg 1. \end{cases} \qquad (A.26)$$

When $\alpha \to 1/2$, it is necessary to go to the next order of the expansion,

$$\hat{\sigma} \simeq \pi(1 - \beta^{2/3})^3. \qquad (A.27)$$

The time delay between two bright allied images is

$$\hat{t}_{12} \simeq \left[\frac{1}{2}\frac{\hat{r}_{12}^2}{M}\right]\frac{\hat{r}_-}{3f^-} \qquad (A.28)$$

$$\simeq \left[\frac{1}{2}\frac{\hat{r}_{12}^2}{M}\right]\begin{cases} \begin{cases} \dfrac{1}{3}\beta^{2/3} & \alpha = 1/2 \\[2mm] \dfrac{1}{3}\left(\dfrac{2(1-\alpha)}{1-2\alpha}\right)^{1-\alpha}\beta^{2(1-\alpha)} & \alpha \neq 1/2 \end{cases} & \beta \ll 1 \\[6mm] \dfrac{1}{2(1-\alpha)}\beta^2 & \beta \gg 1. \end{cases} \qquad (A.29)$$

References

Blandford, R.D., and Jaroszyński, M., 1981, *Astrophys. J., 246*, 1.

Blandford, R.D., and Narayan, R., 1986, *Astrophys. J., 310*, 568.

Blandford, R.D., Phinney, E.S., and Narayan, R., 1987, *Astrophys. J., 313*, 23.

Bourassa, R.R., Kantowski, R., and Norton, T.D., 1973, *Astrophys. J., 185*, 747.

Bourassa, R.R., Kantowski, R., 1975, *Astrophys. J., 195*, 13.

Burke, W.L., 1981, *Astrophys. J. (Letters), 244*, L1.

Djorgovski, S., and Spinrad, H., 1984, *Astrophys. J. (Letters), 282*, L1.

Foy, R., Bonneau, D., and Blazit, A., 1985, *Astr. Astrophys., 149*, L13.

Gorenstein, M.V., Shapiro, I.I., Cohen, N.L., Corey, B.E., Falco, E.E., Marcaide, J.M., Rogers, A.E.E., Whitney, A.R., Porcas, R.W., Preston, R.A., and Rius, A., 1983, *Science, 219*, 54.

Gorenstein, M.V., Shapiro, I.I., Rogers, A.E.E., Cohen, N.L., Corey, B.E., Porcas, R.W., Falco, E.E., Bonometti, R.J., Preston, R.A., Rius, A., and Whitney, A.R., 1984, *Astrophys. J., 287*, 531.

Greenfield, P.E., *et al.*, 1980, *Science, 208*, 495.

Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., and Perley, R., 1985, *Astr. J., 90*, 691.

Kovner, I., 1986a, *Astrophys. J., 312*, 22.

Kovner, I., 1986b, *Princeton University preprint.*

Narasimha, D., Subramanian, K., and Chitre, S.M., 1982, *Mon. Not. R. astr. Soc., 200*, 941.

Narasimha, D., Subramanian, K., and Chitre, S.M., 1984, *Mon. Not. R. astr. Soc., 210*, 79.

Narasimha, D., Subramanian, K., and Chitre, S.M., 1984, *Astrophys. J., 283*, 512.

Narayan, R., Blandford, R., and Nityananda, R., 1984, *Nature, 310*, 113.

Nityananda, R., and Ostriker, J.P., 1984, *J. Astrophys. Astr., 5*, 235.

Paczyński, B., 1986a, *Astrophys. J., 301*, 503.

Paczyński, B., 1986b, *Astrophys. J., 304*, 1.

Phinney, E.S., and Blandford, R., 1986, *Nature, 321*, 569.

Schneider, D.P., Lawrence, C.R., Schmidt, M., Gunn, J.E., Turner, E.L., Burke, B.F., and Dhawan, V., 1985, *Astrophys. J., 294*, 66.

Schneider, D.P., Gunn, J.E., Turner, E.L., Lawrence, C.R., Schmidt, M., and Burke, B.F., 1986, *Astr. J., 91*, 991.

Schneider, P., 1986, *Astrophys. J. (Letters), 300*, L31.

Schneider, P., 1986, *Astr. Astrophys., in press.*

Schneider, P., 1986, *Astr. Astrophys., in press.*

Shaklan, S., and Hege, E., 1986, *Astrophys. J., 303*, 605.

Shaver, P.A., and Cristianii, S., 1986, *Nature,* **321**, 585.

Subramanian, K., and Chitre, S.M., 1985, *Astrophys. J.,* **276**, 440.

Subramanian, K., and Cowling, S.A., 1986, *Mon. Not. R. astr. Soc.,* **219**, 333.

Turner, E.L., Ostriker, J.P., and Gott, J.R., 1984, *Astrophys. J.,* **284**, 1.

Young, P.J., Gunn, J.E., Kristian, J., Oke, J.B., and Westphal, J.A., 1980, *Astrophys. J.,* **241**, 507.

Young, P.J., Deverill, R.S., Gunn, J.E., and Westphal, J.A., 1981, *Astrophys. J.,* **244**, 723.

Young, P.J., 1981, *Astrophys. J.,* **244**, 736.

TABLE 1
Integral Cross Sections and Branching Ratios for the Isothermal Potential

β	ϵ	Total Cross Section	Opposed	Allied	Five Image	$\bar{M}_{12} \geq 10$ Cross Section	Opposed	Allied	Five Image
	0.0	0.0013	0.551	0.449	0.000	0.0013	0.551	0.449	0.000
3.00	0.1	0.0588	0.000	1.000	0.000	0.0002	0.000	1.000	0.000
	0.2	0.3189	0.000	1.000	0.000	0.0002	0.000	1.000	0.000
	0.0	0.0550	0.791	0.209	0.000	0.0550	0.791	0.209	0.000
1.00	0.1	0.2670	0.592	0.231	0.177	0.0535	0.187	0.286	0.526
	0.2	0.1893	0.000	0.907	0.093	0.1226	0.000	0.856	0.144
	0.0	0.6140	0.984	0.016	0.000	0.0619	0.954	0.046	0.000
0.30	0.1	0.6009	0.899	0.023	0.079	0.0487	0.288	0.058	0.654
	0.2	0.5629	0.621	0.050	0.329	0.0470	0.058	0.203	0.739
	0.0	1.5311	0.999	0.001	0.000	0.0502	0.998	0.002	0.000
0.10	0.1	1.5185	0.968	0.001	0.031	0.0394	0.274	0.003	0.723
	0.2	1.4702	0.867	0.002	0.131	0.0427	0.153	0.004	0.843
	0.0	2.3060	1.000	0.000	0.000	0.0491	1.000	0.000	0.000
0.03	0.1	2.3004	0.979	0.000	0.021	0.0386	0.273	0.000	0.726
	0.2	2.2499	0.914	0.000	0.086	0.0423	0.148	0.003	0.849

TABLE 2
Integral Cross Sections and Branching Ratios for the Plummer Potential

β	ϵ	Total Cross Section	Opposed	Allied	Five Image	$\bar{M}_{12} \geq 10$ Cross Section	Opposed	Allied	Five Image
	0.0	0.005	0.567	0.430	0.000	0.0053	0.570	0.430	0.000
3.00	0.1	0.059	0.000	1.000	0.000	0.0588	0.000	1.000	0.000
	0.2	0.319	0.000	1.000	0.000	0.3189	0.000	1.000	0.000
	0.0	0.283	0.869	0.131	0.000	0.0670	0.782	0.218	0.000
1.00	0.1	0.267	0.592	0.231	0.177	0.0535	0.188	0.286	0.526
	0.2	0.269	0.000	0.480	0.520	0.0556	0.000	0.645	0.355
	0.0	7.419	1.000	0.000	0.000	0.0138	0.995	0.005	0.000
0.30	0.1	7.358	0.993	0.000	0.006	0.0118	0.157	0.005	0.838
	0.2	7.175	0.973	0.000	0.027	0.0139	0.109	0.000	0.891
	0.0	75.38	1.000	0.000	0.000	0.0089	1.000	0.000	0.000
0.10	0.1	75.27	0.999	0.000	0.001	0.0123	0.400	0.000	0.600
	0.2	74.04	0.997	0.000	0.003	0.0135	0.348	0.000	0.652
	0.0	313.9	1.000	0.000	0.000	0.0101	1.000	0.000	0.000
0.03	0.1	346.5	1.000	0.000	0.000	0.0169	0.250	0.000	0.750
	0.2	385.7	0.999	0.000	0.001	0.0046	0.000	0.000	1.000

TABLE 3
Amplification of the Faint Image for $\bar{M}_{12} \geq 10$

β	ϵ	Isothermal Opposed	Isothermal Allied	Five Image	Plummer Opposed	Plummer Allied	Five Image
	0.0	425.312	465.311	....	102.022	113.910	....
3.00	0.1	21.3568	20.7634	....	21.3568	20.7634	....
	0.2	....	8.07411	....	....	8.07411	....
	0.0	9.00360	15.4445	....	1.07700	3.66580	....
1.00	0.1	2.54422	2.79136	1.19969	2.54422	2.79136	1.19969
	0.2	....	4.73180	3.78601	....	2.93875	1.84264
	0.0	0.17110	5.53943	....	0.02735	1.07815	....
0.30	0.1	0.24902	3.63303	0.17827	0.02957	1.06770	0.00833
	0.2	0.34519	6.51348	0.41989	0.03730	1.06770	0.01340
	0.0	0.01259	2.84462	....	0.00012	1.06770	....
0.10	0.1	0.03123	2.56735	0.01291	0.00012	1.06770	0.00012
	0.2	0.05290	2.01740	0.01712	0.00012	1.06770	0.00012
	0.0	0.00366	2.01740	....	0.00003	1.06770	....
0.03	0.1	0.00604	2.01740	0.00111	0.00003	1.06770	0.00003
	0.2	0.00616	2.09214	0.00414	0.00003	1.06770	0.00002
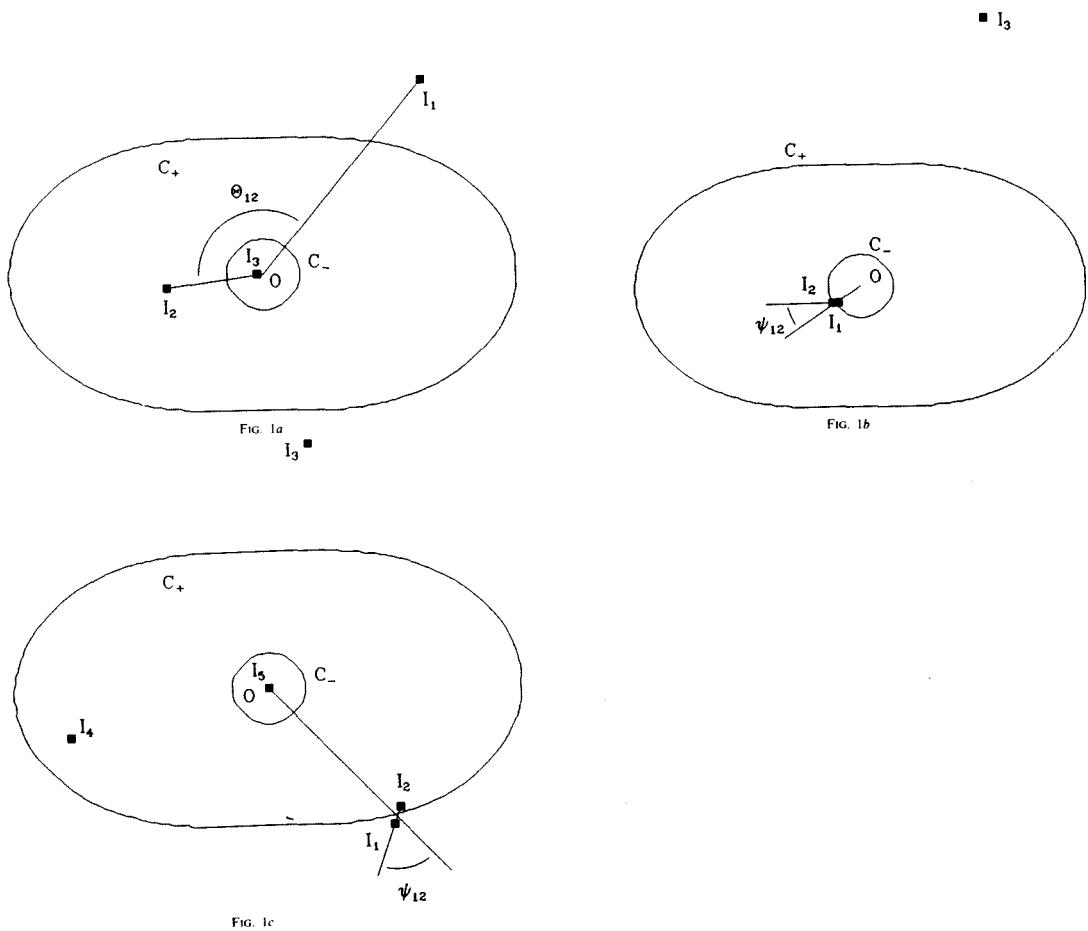
Figure 1. **(a).** Location and properties of opposed images for a source triply imaged by an elliptical lens centered at O. The tangential and radial critical lines are labeled C_+ and C_-. The three images are labeled I_1, I_2, and I_3 in order of decreasing amplification. The angle subtended by the two brightest images at O is designated θ_{12}. r_{12} (not shown) is the separation of the two brightest images (I_1 and I_2) in units of r_+. r_{O3} is the distance from the origin O to the faintest image. **(b).** Location and properties of allied images. Definitions as in (a) with the exception that ψ_{12} measures the angle between the separation of the two brightest images, and the radius vector. **(c).** Location and properties of images for a five-image configuration. Definitions as in (a) and (b). If the two brightest images are allied the angle ψ_{12} is measured, if they are opposed, the angle θ_{12} is measured.
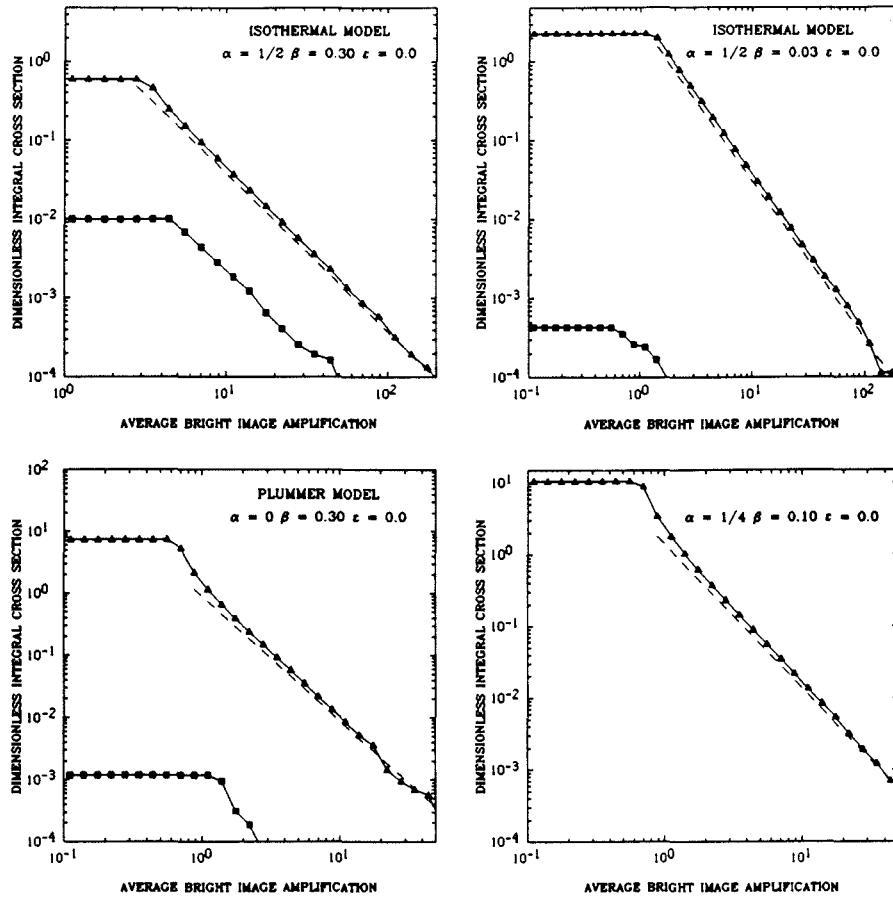
Figure 2. Dimensionless integral cross sections, $\hat{\sigma}(\bar{M}_{12} > M) = \sigma(\bar{M}_{12} > M)/r_+^2$, as a function of amplification $\bar{M}_{12}$ for four circular models. Lines marked by triangles are opposed cross sections, squares are allied cross sections. The dotted line is the asymptotic result calculated from a line integral around the caustics (Equation 4.2).

(a).	$\alpha = 1/2$	Isothermal model	$\beta = 0.30$
(b).	$\alpha = 1/2$	Isothermal model	$\beta = 0.03$
(c).	$\alpha = 0$	Plummer model	$\beta = 0.30$
(d).	$\alpha = 1/4$		$\beta = 0.10$

ISOTHERMAL MODEL
$\alpha = 1/2$ $\beta = 0.10$ $\epsilon = 0.0$

BRIGHTEST IMAGE SEPARATION

AVERAGE BRIGHT IMAGE AMPLIFICATION
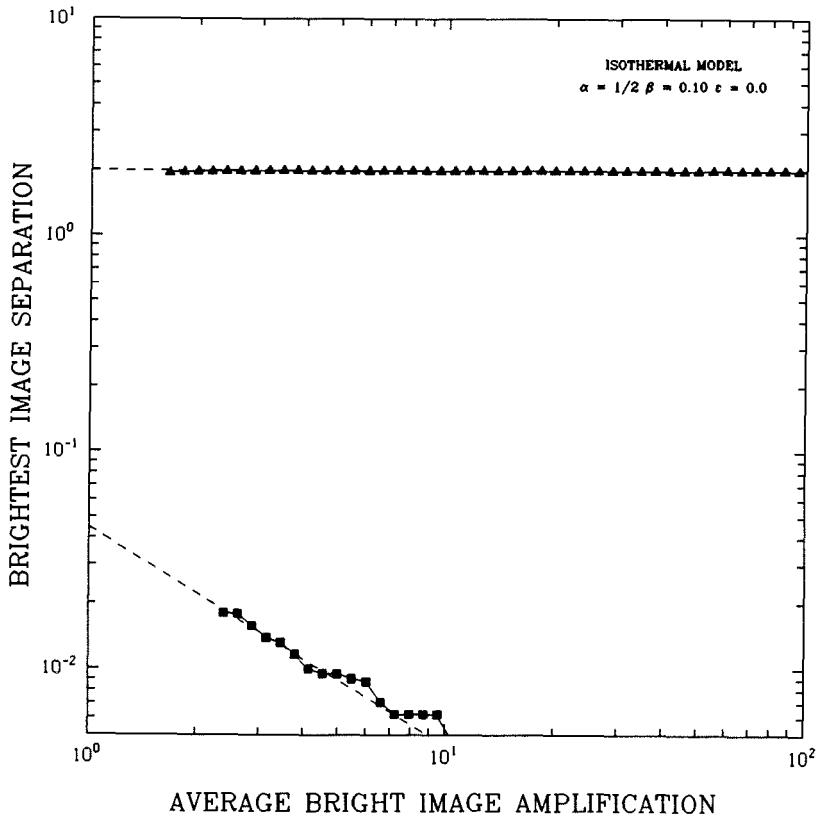
Figure 3. Dimensionless bright image separation $\hat{r}_{12} = r_{12}/r_+$ as a function of amplification $\bar{M}_{12}$ for the isothermal $\beta = 0.1$, $\epsilon = 0.0$ case. Shown for opposed (triangles) and allied (squares) image configurations. Error bars are the standard deviation in the value at the given magnification. Dashed lines are analytic fits from the Appendix (equations $A.6$ and $A.22$). Note that for opposed images $\hat{r}_{12}$ is insensitive to the magnification as it is fixed by the average diameter of the outer critical curve. However, the allied image separations decrease inversely with magnification.
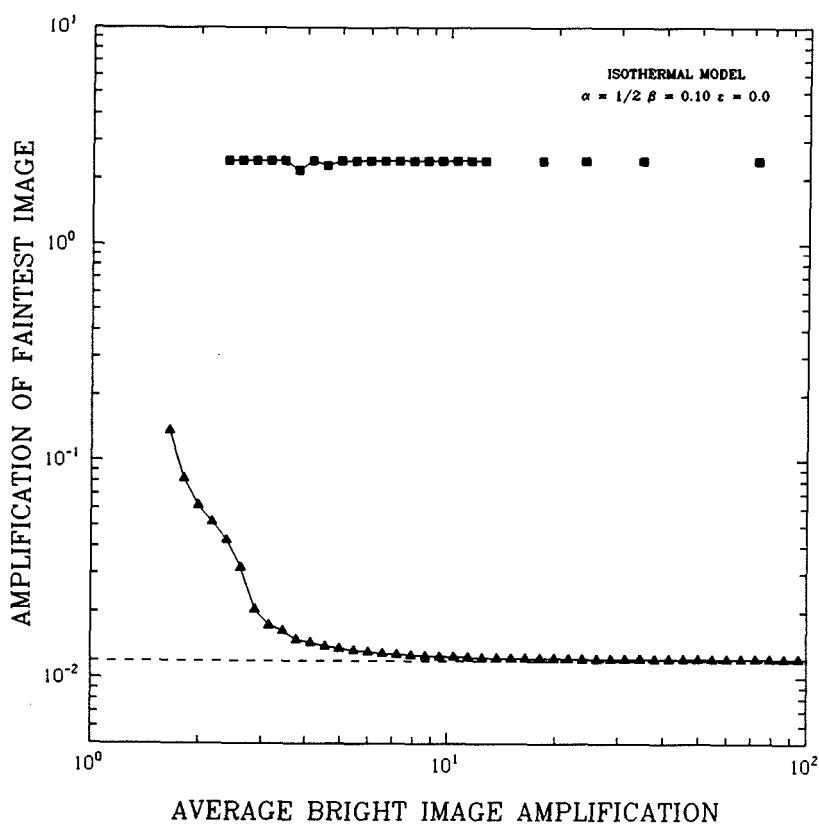
Figure 4. Amplification of the faintest image as a function of $\bar{M}_{12}$. Model and labeling as in **Figure 3**. Note that in the three-image allied configuration, the third image is not deamplified. No points are drawn in regions where the differential cross section vanishes due to the finite number of grid points in the simulation.
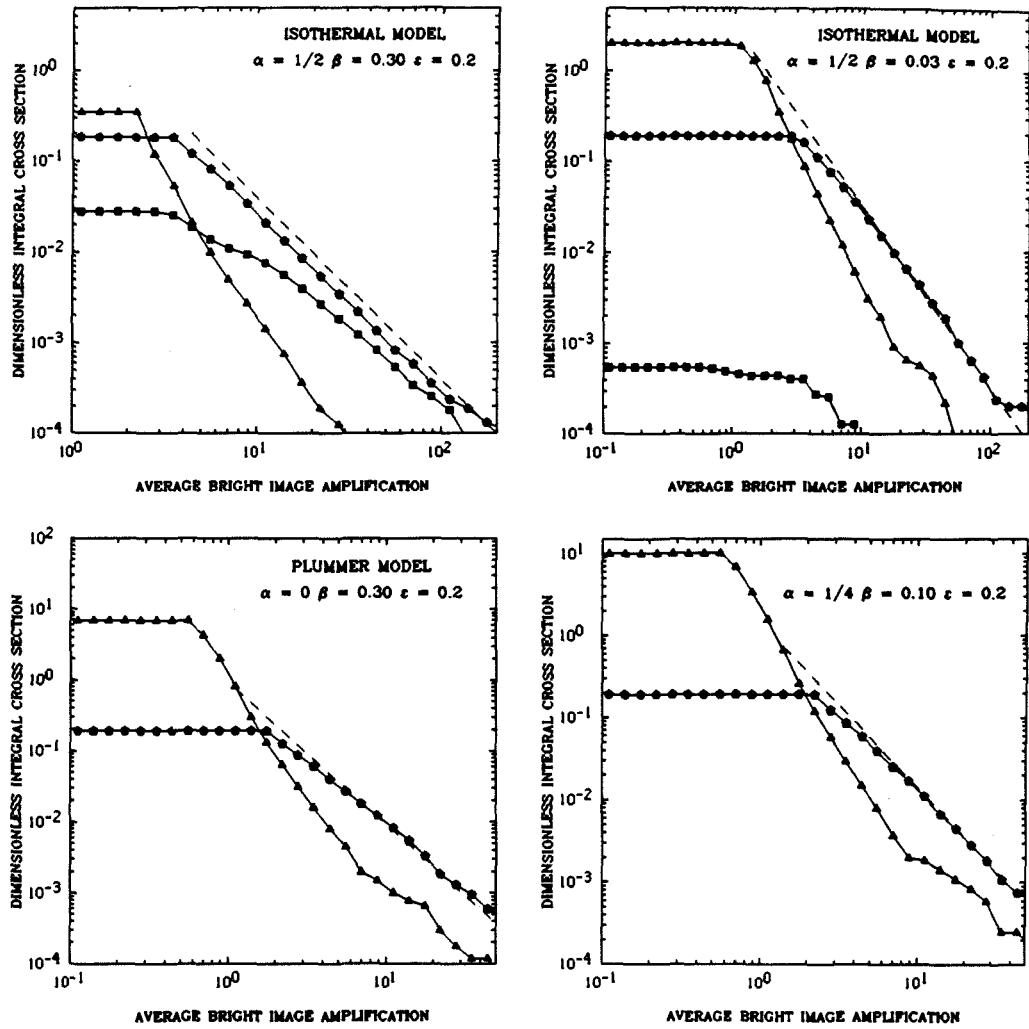
Figure 5. Dimensionless integral cross sections, $\hat{\sigma}(\bar{M}_{12} > M)$ as a function of amplification $\bar{M}_{12}$ for four eccentric models with $\epsilon = 0.2$. Definitions are as in **Figure 2**, with the addition that lines marked by pentagons are five-image cross sections.

(a).	$\alpha = 1/2$	Isothermal model	$\beta = 0.30$
(b).	$\alpha = 1/2$	Isothermal model	$\beta = 0.03$
(c).	$\alpha = 0$	Plummer model	$\beta = 0.30$
(d).	$\alpha = 1/4$		$\beta = 0.10$

Note that at large amplification, the five-image cross section is dominant.
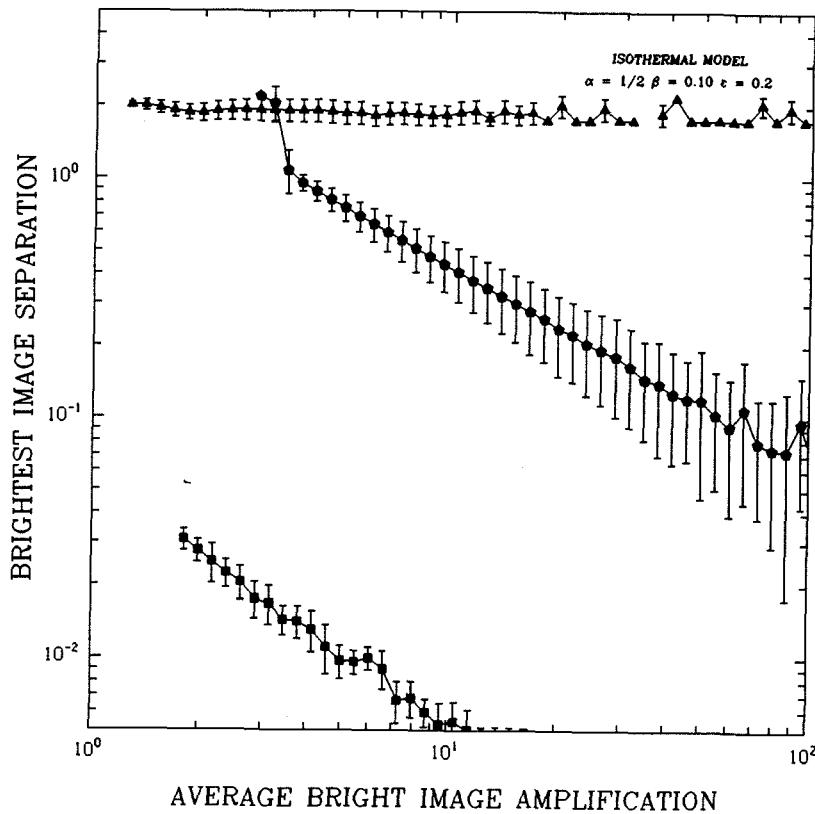
Figure 6. Dimensionless bright image separation $\hat{r}_{12} = r_{12}/r_+$ as a function of amplification $\bar{M}_{12}$ for the isothermal $\beta = 0.1$, $\epsilon = 0.2$ case. Shown for opposed-image (triangles), allied-image (squares), and five-image (pentagons) configurations. Error bars are the standard deviation in the value at the given magnification. Note that for opposed images r_{12} remains insensitive to the magnification.
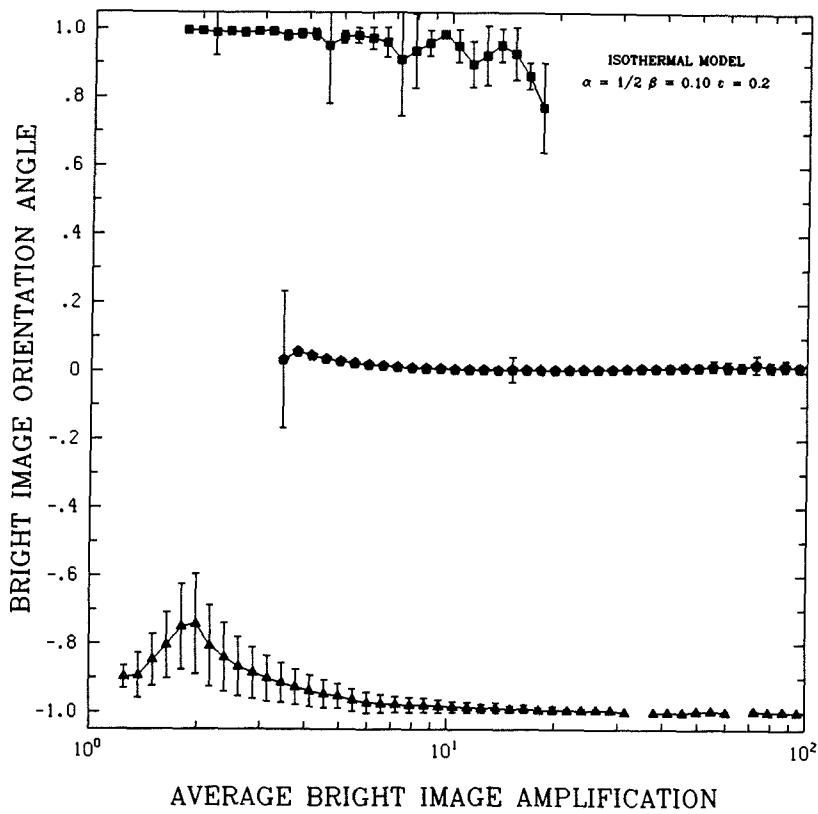
Figure 7. Cosines of orientation angles $\cos \theta_{12}$ and $\cos \psi_{12}$ for the same model as in **Figure 6** as a function of amplification $\bar{M}_{12}$. Positive values of the angle give ψ_{12}, while negative values are θ_{12}. Opposed images are diametrically opposed with respect to the center of the potential ($\theta_{12} \sim 180°$). If the brightest two images of a three image configuration are allied, they are displaced roughly radially ($\psi_{12} \sim 0°$). The brightest two images in a five-image configuration are generally also allied, but separated along a tangential direction ($\psi_{12} \sim 90°$).
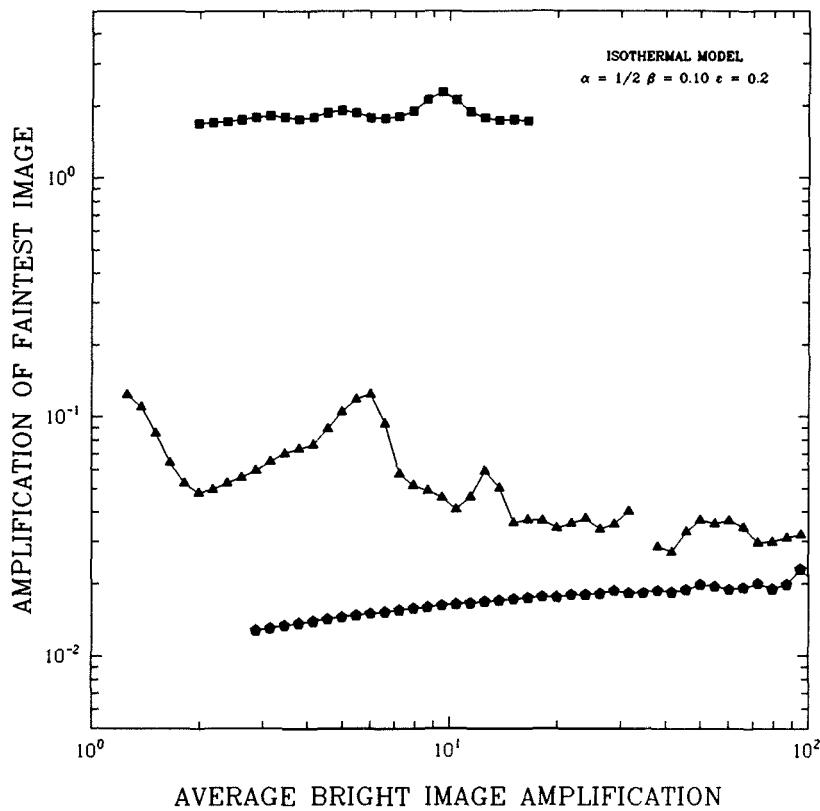
Figure 8. Amplification of the faintest image as a function of $\bar{M}_{12}$. Model and labeling as in **Figures 6** and **7**. Note that in the three-image allied configuration, the third image is not deamplified.
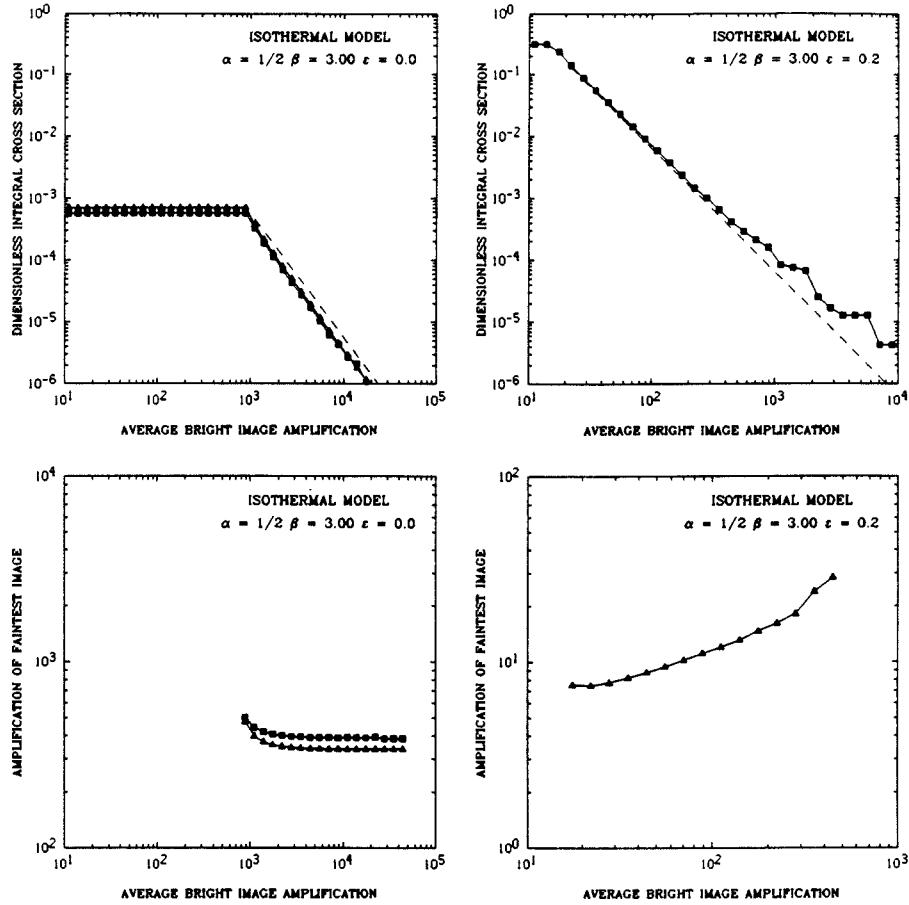
Figure 9. Dimensionless integral cross sections and faint image amplifications as a function of $\bar{M}_{12}$ for circular and elliptical cases of the marginal isothermal lens with $\beta = 3$.

(a).	Cross Section	$\epsilon = 0.0$
(b).	Cross Section	$\epsilon = 0.2$
(c).	Faint Image Amplification	$\epsilon = 0.0$
(d).	Faint Image Amplification	$\epsilon = 0.2$
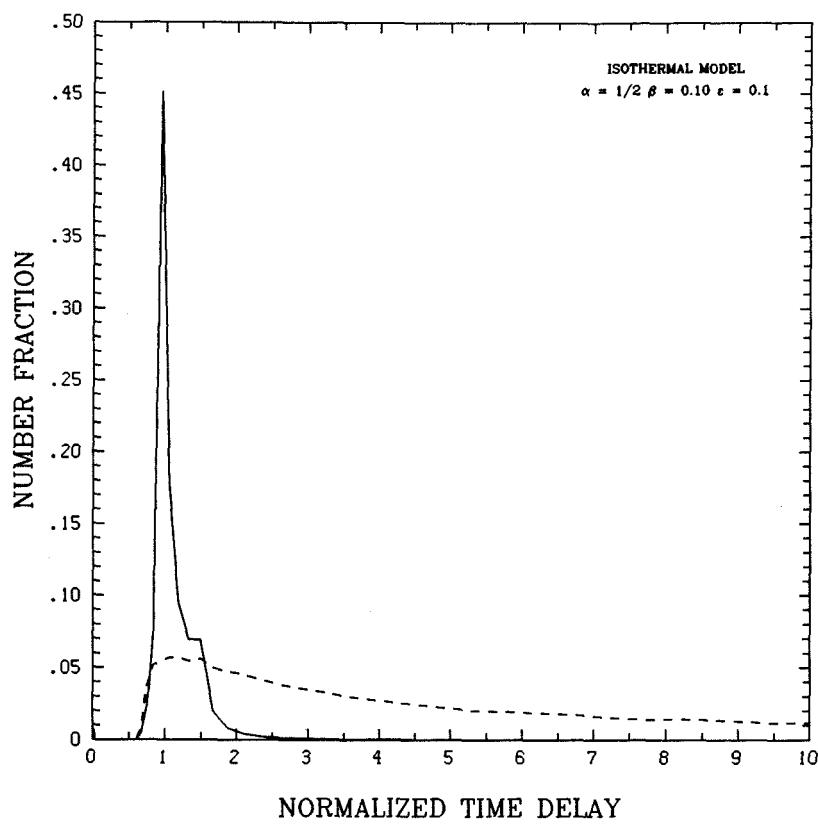
Figure 10. Number fraction distribution of time delays $N(T)$ where $T = (2\bar{M}_{12}/\hat{r}_{12}^2)\hat{t}_{12}$ for an isothermal model with $\beta = 0.1$ and $\epsilon = 0.1$. Solid line is distribution of three-image opposed time delays, dashed line is for five-image time delays. The three-image radial allied images are not shown because the distribution is not meaningful due to low statistics.

Chapter 3

GRAVITATIONAL IMAGING BY ISOLATED ELLIPTICAL POTENTIAL WELLS II: PROBABILITY DISTRIBUTIONS

Christopher S. Kochanek,

and

Roger D. Blandford

Theoretical Astrophysics 130-33,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

Abstract

Optical depths for the magnification and positions of multiple quasar images created by a cosmological distribution of isolated elliptical potential wells are computed. Introducing a core radius into a singular potential well can reduce the optical depth for multiple imaging by a large factor. Strong lenses predominantly produce three images of which two are opposed at low magnification and five images at high magnification when amplification bias may become important. In practice, amplification bias is important only for bright ($B \lesssim 18^m$) quasars. In most instances one image is deamplified by the core. Marginal lenses produce mainly three comparatively bright images, but with relatively small probability. As an application, we estimate a probability $\sim 10^{-6}$ that a given galaxy at $z_L \sim 0.4$ will lens a background quasar (e.g. $2237 + 0305$).

1. Introduction

In Blandford and Kochanek (1987; hereafter Paper 1), we computed cross sections for multiple imaging by a three-parameter (hardness, core radius and eccentricity) family of isolated elliptical potentials. We calculated the relative probabilities for creating different types of image arrangement and magnifications and also provided estimates of these quantities useful for rough estimates. In Paper 1, we confined our attention to the case of a single quasar and a single lens. In the present paper, we go on to consider the observable consequences of distributing lenses and quasars throughout the universe.

The nature of the lensing potentials responsible for the observed multiply imaged quasars is unclear. Several of the gravitational lens candidates are found in regions of the sky devoid of galaxies (e.g. Canizares 1986). In these cases, the deflecting matter appears quite sub-luminous and is perhaps a concentration of the "dark matter" strongly suspected to dominate the dynamics of the universe. The simplest non-degenerate form of potential well is elliptical in shape. (As we discuss in Paper 1, circularly symmetric potentials are not generic and their use in calculations like these can give misleading impressions.) In §2, we relate the physical parameters that describe our family of potentials to the velocity dispersion and core radius of an underlying isotropic particle distribution. For illustration purposes, we use galaxy-like and cluster-like parameters for our potential wells although it is possible to scale our results to other forms. We emphasize that our potentials are not intended to be accurate descriptions of real galaxies and clusters. In §3 we introduce the notion of optical depth and evaluate it under the assumption that there is a constant co-moving density of lenses and a source at redshift $z_S = 2$. In §4 we adopt a simple model of the quasar luminosity function and use this to calculate the expected observed distribution of image configurations. Our conclusions are collected in §5.

2. Model Potentials

We continue to use the model two-dimensional Newtonian potential introduced in Paper 1:

$$\psi(x,y) = \frac{A}{2\alpha} \left[\left(1 + (1-\epsilon)\left(\frac{x}{s}\right)^2 + (1+\epsilon)\left(\frac{y}{s}\right)^2 \right)^\alpha - 1 \right]. \tag{2.1}$$

The parameters α, A, s, and ϵ are the hardness, depth, core size and eccentricity of the potential. For simplicity, we restrict our attention to approximately isothermal potentials with $\alpha = 1/2$ and limit our study to the effects of A, s, and ϵ. The potential is reparametrized in terms of a dimensionless ratio $\beta \equiv s/r_+$ where r_+ equals the radius of the ring image formed on the lens plane by a source behind the origin of a circular potential. For the isothermal potential,

$$\beta = \left[\left(\frac{2AD_{OL}D_{LS}}{c^2 s^2 D_{OS}} \right)^2 - 1 \right]^{-1/2}. \tag{2.2}$$

If the potential is formed by small identical particles (e.g., stars) with an isotropic velocity distribution, then we can compute the observed central velocity dispersion in a formal manner by the procedure detailed in the Appendix to find that

$$\sigma_0^2 = \frac{3A}{4\pi s} \tag{2.3}$$

for the isothermal potential. The central velocity dispersion does not equal the local velocity dispersion well outside the core $\sigma^2 = A/2\pi s$. Direct observations of galaxies

usually cover a region substantially larger than the core. We therefore use σ, not σ_0 in contrasting our results with those obtained for a singular potential.

Our potential scale length s is also hard to relate unambiguously to observational parameters. Observationally, core radii for ellipticals and spiral bulges are generally measured by fitting the surface brightness at larger radii to a standard profile. A de-Vaucouleurs profile does not fit our potential; nor need it do so because the mass-to-light ratio need not be constant and velocity anisotropies may be present (e.g. Kormendy 1986). Alternatively, we could estimate s using the relation between central velocity dispersion, the central surface density, and an assumed mass-to-light ratio. This approach would then rest on the value of the mass-to-light ratio for galaxies (and clusters), which is not well determined. Because of these uncertainties, we will perform the calculations for a number of different scale lengths near the expected values. If we use equation (A.1) and simply assume that the mass-to-light ratio is constant over a few core radii, we can compute the radius at which the surface brightness has fallen to 1/2 of its central value to be $r_{1/2} \simeq 1.07s$. Recent observations reported by Lauer (1985), which have resolved the cores of nearby galaxies, have derived values for $r_{1/2}$ ranging from less than 0.1 to 1 kpc.

For the study of lensing by galaxy-like potentials, we will use the velocity dispersions and local proper number densities introduced by Turner, Ostriker, and Gott (1983, henceforth TOG) to model elliptical ($\sigma = 306$ km/s and $n_0 = 8.8 \times 10^{-3} h_{100}^3$ Mpc^{-3}) and spiral galaxies ($\sigma = 177$ km/s and $n_0 = 2.0 \times 10^{-2} h_{100}^3$ Mpc^{-3}) respectively. We use values of $s = 1/9, 1/3, 1$ and 3 kpc for the core scale length s. For clusters of galaxies, we shall use $\sigma = 1000$ km/s and $n_0 = 8.0 \times 10^{-6} h_{100}^3$ Mpc^{-3} (richness class $R \geq 1$ Kaiser 1984). There are indications that clusters have a cusp in the number density of galaxies (Beers, and Tonry, 1986). However, this does not necessarily imply a cusp in the gravitational potential. We choose a scale length $s = 33$ kpc, which is small enough to ensure that the lens can multiply image over a range of redshifts. Models will be referred to by (σ,s) where σ is the asymptotic velocity dispersion in km/s, and s is the core radius in kpc.

We can divide these lenses into two categories: intrinsically marginal, and intrinsically strong. A lens is *intrinsically marginal* if it is only just capable of generating multiple images (*cf.*, Narayan, Blandford, and Nityananda, 1983, Kovner, 1986). These lens are characterized by maximum values of the strength parameter $\beta \gtrsim 0.5$. Examples are the (306,3), (177,1), and (1000,33) models. A lens is *intrinsically strong* if $\beta \lesssim 0.5$ over a significant redshift range. The (306,1), (306,1/3), (306,1/9), (177,1/3), and (177,1/9) models are intrinsically strong lenses. Intrinsically strong lenses can, of course, be *locally marginal*, depending on the position of the lens and source.

Obviously, different choices for σ and s could be made. We note that the nature of the parametrization in terms of the ratio $A \propto \sigma^2/s$ and the fact that the lensing equations for an isolated potential can be made independent of the absolute physical scale means that each model introduced above actually describes a family of lenses. Under the transformation $s \to s'$ and $\sigma \to \sigma' = \sigma(s'/s)^{1/2}$, the lensing properties remain the same up to scale changes in the cross section and distances. The cross section must be multiplied by $(s'/s)^2$, and distances (such as the separation of the two brightest images) must be multiplied by (s'/s). Hence, the (306,3), (177,1), and (1000,33) models have roughly the same lensing properties provided all lengths are scaled by the appropriate ratio of core sizes.

3. Optical Depth Calculation

3.1. Optical Depths

Ultimately, we wish to calculate the probability of particular arrangements of multiple images from a cosmological distribution of lenses and quasars. In order to do this, it is convenient to define an angular cross section $d\sigma(\xi_L^i, z_L, \xi_I^i, z_S)$, which is the element of solid angle on the sky within which a source at redshift z_S can lie, behind a lens at redshift z_L, parametrized by ξ_L^i (in this case $\alpha, \beta, \epsilon, r_+$), in order to produce images with image characteristics ξ_I^i (such as the average amplification of the two brightest images exceeding 10). The differential cross section $d\sigma$ can be integrated to give an integral cross section σ in the usual manner. If the density of sources (of the specified type) on the sky is n_s, then the probability of a lensing event satisfying our requirements is just $n_s \sigma$.

For simplicity, we assume that the lenses do not evolve significantly over the redshift interval $0.5 \lesssim z_L \lesssim 1.5$ where they are most efficient at gravitational imaging. We can therefore define an optical depth τ for some property of the images (e.g., producing five images) by integrating the cross section over lens redshift.

$$\tau(\xi_L^i, \xi_I^i, z_S) = \int_0^{z_S} n_{L0}\sigma(\xi_L^i, z_L, \xi_I^i, z_S)(1 + z_L)^3 d_L^2 c\, dt_L \qquad (3.1)$$

where n_{L0} is the comoving density of lenses and dt_L is an interval of cosmic time. If $\bar{n}_{L0} = n_{L0}(c/H_0)^3$ is the comoving density of lenses of the required type per cubic Hubble radius and we adopt an Einstein-de Sitter cosmology, (cf., TOG $\Omega_0 = 1$ filled beam cosmology).

$$\tau(\xi_L^i, \xi_I^i, z_S) = 4\bar{n}_{L0} \int_0^{z_S} \sigma(\xi_L^i, z_L, \xi_I^i, z_S)[(1 + z_L)^{1/2} - 1]^2 (1 + z_L)^{-5/2} dz_L \qquad (3.2)$$

We confine our study to this simple cosmology because the work in TOG demonstrated that the particular cosmology used was less important than the internal structure of the lensing potential. We assume that $\tau \ll 1$ so that single scattering events completely dominate the optical depth. As we know of ~ 2000 quasars, the expected number of events would be $\sim 2000\tau$ in the absence of selection effects.

It is convenient to define an optical depth weighted average of a quantity f (e.g. mean bright image separation) to be

$$\langle f \rangle = \frac{1}{\tau} \int f(z_L)\frac{d\tau}{dz_L} dz_L. \qquad (3.3)$$

The integrals are performed over a set of cross sections generated by the method described in Paper 1 for varying strength parameters β. Each value of β corresponds to the two different lens redshifts through equation (2.2) relating β, the lens parameters and the angular diameter distances. The integrals for the optical depths were over 20 to 40 lens redshift points resulting in fractional errors of approximately 10%. Source redshifts ranged over six equally spaced points for $z_s = 0.5$ to 2.5.

3.2. Isothermal Galaxies and Clusters

As discussed in §2, we will restrict ourselves to a set of three simple models: isothermal potentials with central velocity dispersions of $\sigma = 306$, and 177 km/s, which are rough galaxy-like models for giant ellipticals and spirals respectively, and $\sigma = 1000$ km/s, which

is a cluster-like model. The most important result is that the introduction of a core results in a substantial reduction in the optical depth for lensing relative to the singular result. The exact result for the multiple image cross section by potential (2.1) in the circular limit is:

$$\sigma = \pi r_+^2 \beta^2 \left[(1 + \beta^{-2})^{1/3} - 1 \right]^3. \tag{3.4}$$

The parameter $\beta \equiv s/r_+$ is zero in the singular limit. When a core is introduced, the position of the C_+ (outer) critical line changes by only a small amount, so that r_+ remains approximately constant. The C_- (inner) critical line expands from the origin and consequently leads to a reduction in the cross section. The peak values of β for the $\sigma = 306$ km/s sequence $s = 1/9, 1/3, 1,$ and 3 kpc are $\beta = .02, .06, .18,$ and .65 for a source redshift of $z_S = 2$. Figure 1 shows the total lensing cross section and the differential optical depth $(1/\tau)d\tau/dz_L$ as a function of lens redshift for circular isothermal lenses. The cross section is substantially reduced relative to that for the singular isothermal sphere even for the smallest core size; factors of 1.8, 6, and 50 relative to the SIS cross section The largest contribution to the optical depth comes from $z_L \sim .5$ in all cases. When the source redshift is reduced to $z_S \simeq 1$, the position of the peak in the differential optical depth is only reduced to $z_S \simeq 0.4$.

Introducing a finite core also reduces the total multiple-imaging optical depth. For the elliptical galaxy model, the SIS optical depth of $\tau = 2.6 \times 10^{-3}$ is reduced by factors of 1.5, 1.8, 4.4, and 54 for $s = 1/9, 1/3, 1,$ and 3 kpc respectively. The effects are even more dramatic for the spiral galaxy model because the lens becomes intrinsically marginal for smaller core sizes. In this case the SIS optical depth is $\tau = 6.5 \times 10^{-4}$ and the reduction factors are 1.9, 4.4, and 52. (For $s = 3$ kpc the lens becomes subcritical.) Likewise for the marginal cluster-like potential, introducing the core ($s = 33$ kpc) reduces the optical depth from $\tau = 2.7 \times 10^{-4}$ to $\tau = 3.4 \times 10^{-6}$. Because the reduction in cross section occurs more strongly in regions where the cross section is already small, the lens redshift region contributing to the optical depth becomes more concentrated near the redshift of the peak. For the SIS case, the differential optical depth $(1/\tau)d\tau/dz_L$ is greater than one half of its peak value in the range $.2 \lesssim z_L \lesssim 1.0$ (for quasar at $z_S = 2.$, and a lens with $\sigma = 306$ km/s), whereas for the same model with a core size of 3 kpc, it is concentrated between $.25 \lesssim z_L \lesssim .7$. Large reductions in cross sections occur for elliptical potentials with cores as well, and lead to substantial decreases in the optical depth for lensing as can be seen from Table 1.

The integration over redshift can also alter the distribution of image configurations. The branching ratio for the three-image allied case is negligible except when the lens is intrinsically marginal. For intrinsically strong lenses the contributions from the marginal endpoints in the integration over lens redshifts are generally negligible. The three-image opposed and five-image branching ratios are sensitive to s and $\bar{M}_{12}$. For example, for the model elliptical galaxy with a scale length of 1/3 kpc, the ratio of the optical depths for the three-image opposed geometry to that for five images at a magnification of 1 (10) is 7.5 (0.2), whereas for $s = 1$ kpc it is 3.0 (0.05) (see Figure 2a). The ratio is a strong function of amplification because the cross sections are dominated by the five-image cross section if the mean amplification of the bright images $\bar{M}_{12} \gtrsim (1 - \alpha)^{-1}\epsilon^{-1} = 10$ for the isothermal potential with $\epsilon = 0.2$ (cf., Paper 1).

The average separation of the brightest images $< r_{12} >$ formed by the lens depends largely on the velocity dispersion. In the three-image opposed case, images appear roughly on either side of the galaxy at fixed separation (independent of magnification) with values

of roughly $1\pm.5$, 3 ± 1, and 24 ± 8 arc-seconds for $\sigma_0 = 177, 306$ and 1000 km/s respectively (see Figure 2b). By contrast, the two brightest images in three-image allied geometry have separations of $s = .07\,h_{100}s_{kpc}$ arc-seconds at low amplification, and the separation decreases slightly less rapidly than the $1/\bar{M}_{12}$ law seen in the cross sections because the marginal lenses contribute at high but not low amplification. In general the two images will be too close together to be optically resolvable for galactic potentials with typical separations being several tenths of an arc-second or less. The third image, however, is also amplified for the three-image allied case, and will typically be placed about one arc-second from the lens on the opposite side from the bright pair. The separation of brightest pair of images in the five-image configuration is approximately equal to that of the three-image opposed case for $\bar{M}_{12} \sim 1$, but decreases with an approximately $1/\bar{M}_{12}$ dependence for larger magnifications (*cf.*, Paper 1).

For the strong three-image opposed and five-image cases, the faintest image is usually within .1 arc-seconds of the center of the lens and is typically deamplified by $2-5$ magnitudes depending on the core size (see Figure 2c, d). However, the average deamplification is less than $\sim \beta^2$ (*cf.*, Paper 1) because of the contributions from the marginal lenses in which the faint image is amplified. Nevertheless, the distribution is asymmetric, and in most observed cases the amplification will be close to β^2.

4. Calculated Image Properties

We have investigated a variety of model potentials of the form described in Paper 1 and computed the expected properties of multiply-imaged quasars as discussed in §3. Here we describe and illustrate the main features of our results.

4.1. Number Counts

For simplicity, we shall adopt the approximation that quasars of a given B magnitude are evenly distributed in redshift for $0.5 \leq z_S \leq 2.5$ with no quasars outside that range. We model the total number of quasars per square degree brighter than magnitude B_s by

$$\log N(B_s) = \begin{cases} 0.5 + 0.9(B_s - 19) & 14.5 < B_s < 19 \\ 0.5 + 0.9(B_s - 19) - 0.14(B_s - 19)^2 & 19 < B_s < 22 \end{cases} \qquad (4.1)$$

The bright quasar counts have a slope of eight per magnitude, but the counts for faint quasars flatten to a slope of ~ 4 per magnitude. There are approximately three quasars per square degree at 19^m and one hundred quasars per square degree at 22^m. This prescription is a reasonable approximation to the results of recent quasar surveys (*e.g.*, Shanks *et al.*, 1986). From the optical depth at a given source redshift as a function of image magnification $\tau(\xi_L^i, \xi_I^i, z_S)$, we compute expected magnitude and redshift distribution of lensed quasar images for lenses of a given type and normalized density $\bar{n}_{L0}$. Suppose that the average flux from the two brightest images corresponds to a magnitude $\bar{B}_{12}$, and the intrinsic source magnitude is B_S, corresponding to an amplification $2.5 \log \bar{M}_{12} \simeq B_S - \bar{B}_{12}$. The observed distribution of multiply-imaged quasars $N^*(\bar{B}_{12}, z_S)$ then satisfies

$$\frac{dN^*}{dz_S} = \int \frac{1}{2} \frac{dN}{dB_S}(B_S)dB_S\tau(\xi_L^i, \bar{M}_{12} > 10^{0.4(B_S - \bar{B}_{12})}, z_S). \qquad (4.2)$$

Equation (4.2) embodies the principle of amplification bias (TOG); i.e., the possibility that highly amplified faint quasars outnumber modestly amplified bright quasars. Now

asymptotically, $\tau \propto A^{-2}$ for large amplifications, which implies that amplification bias is important when the integral counts increase with decreasing flux S faster than $N(S) \propto S^{-2}$. Equivalently, the slope of the number magnitude relation must satisfy $d \log N / dB > 0.8$ (6.3 per magnitude), which according to equation (4.1) is only true for $B < 19.5^m$. In other words, we do not expect amplification bias to be very important for faint quasars, and even for the brightest $\sim 16^m$ quasars, amplification of $< 19^m$ quasars by factors $\lesssim 10$ can only be expected to double the number of multiply-imaged quasars. Our detailed computations verify this. (See Vietri, and Ostriker, 1983, and Vietri, 1985, for a more formal discussion limited to the singular isothermal sphere.)

4.2. Simulations of Intrinsically Strong Lenses

In Figure 3a, we show the expected number of multiply-imaged quasars over the sky as a function of the magnitude corresponding to the average flux of the two brightest images, $\bar{B}_{12} = B_S - 2.5 \log(\bar{M}_{12})$ for the (306,1/3) model, an intrinsically strong lens. While 90% of all lensing events are associated with the three-image opposed geometry, the fraction of five-image events increases substantially below 18^m and dominates below 14^m. This is due to amplification bias as we discussed in §4.1. The separation of the two brightest images reflects this amplification bias (see Figure 3b). Recall that at low amplifications, the brightest images in the five-image case have roughly the same separation as the brightest images of the opposed three-image case but then merge at higher amplifications with separation $\propto 1/\bar{M}_{12}$ (cf., Paper 1). When $\bar{B}_{12} \lesssim 18^m$, the contribution from amplified, fainter quasars lowers the mean separation of the two brightest images. When $\bar{B}_{12} \gtrsim 18^m$ the mean separation converges to that expected for $\bar{M}_{12} \lesssim 3$. The opposed bright images are separated by approximately 3 arc-seconds on average, and the mean square scatter about the mean is ~ 1 arc-second. This reflects the fact that the bright images in the opposed geometry have roughly fixed separation independent of magnification even if the potential is not circular. For the five-image case, the separation is usually very small for the brightest quasars because of amplification biasing. The scatter of the distribution is enormous, because the number distribution at any magnitude has contributions from events at low amplification with separations ~ 2 arc-seconds, and contributions from highly amplified events with separations $\lesssim 0.1$ arc-seconds. The dispersion decreases for the lensing events above 17^m where there are only small contributions from the highly amplified images with small separations. For the relatively rare allied case, the brightest of two images are always separated by $\lesssim 0.1$ arc-seconds. Because the two brightest images are so close together, a three-image allied lens may be confused with a three-image opposed lens which has a strongly deamplified faint image. The third image in the lens is amplified and on the opposite side of the potential.

For opposed three-image and five-image geometries, the faintest image is usually located near the core within the C_- critical curve. It is deamplified roughly by the square of the ratio of the critical surface density for lensing divided by the actual surface density. For five-image configurations we find that the mean amplification of the faintest image M_5 ($\ll 1$) and its displacement from the center of the potential is correlated with the mean amplification of the brightest images $\bar{M}_{12}$ ($\gtrsim 1$). For the three-image opposed geometry, on the other hand, we find that the amplification of the faintest image M_3 ($\ll 1$) and its displacement from the center is anti-correlated with the mean amplification of the brightest images. Thus, for both three-image allied and five-image configurations, as the two brightest images of the lens brighten, the positions and amplifications of the faint images approach each other (Figure 3cd). The faint image is deamplified by three to six

magnitudes in both cases. The third image for the three-image allied case, however, is located about one or two arc-seconds from the center of the potential, roughly opposite from the two bright images. It is also amplified by about one magnitude so that it is typically only a little fainter than the bright pair.

4.3. *Simulations of Intrinsically Marginal Lenses*

In Figure 4a, we show the expected number of multiply-imaged quasars for the (306,3) model, which is an intrinsically marginal lens. In this case, the three-image opposed geometry is suppressed, and the lenses are 70% three-image allied and 30% five-image. The branching ratio between these two geometries is roughly independent of amplification, so there is no amplification bias effect in the relative numbers of the two geometries as a function of the magnitude of the brightest images. The total expected number of multiple-imaging events, however, is much less than for the intrinsically strong case (~ 10 events over the sky as compared to ~ 1000 events over the sky for the intrinsically strong lenses.)

The brightest pair of images have roughly the same separation independent of the geometry (allied three-image or five-image). (See Figure 4b.) This occurs because the potential is essentially quadratic in the region where the multiple imaging takes place. The mean separation, however, increases with $\bar{B}_{12}$, again because of amplification bias. The brightest events are strongly amplified faint quasars for which the bright image separation is small. The separation ranges from 0.5 arc-second at $\sim 15^m$ to two arc-seconds at 20^m. In the intrinsically weak case (Figure 4c, d), there are only three-image allied and five-image events, and the faintest images behave similarly (as did the brightest pair). The faint image is approximately one arc-second from the core, and is amplified by one to two magnitudes.

5. Conclusions

In this paper we have computed the range of image properties to be expected from a cosmological distribution of isolated lenses and quasar sources. For a given source, at $z_S \sim 2$, there is a range of lens redshifts z_L within which multiple images can be created. For the Einstein-De Sitter universe that we adopt, this range is generally centered on $z \sim 0.4 - 0.5$. We have considered intrinsically strong and intrinsically marginal lenses and quantified our results using optical depths and predicted quasar distributions.

Our results are mostly straightforward generalizations of those presented in Paper 1. If we compare with the earlier work of TOG who used a singular, isothermal sphere for their potential, we found that the introduction of a finite core size can significantly reduce the overall cross section for multiple imaging. For a galaxy-like potential, a central core radius of ~ 1 kpc similar to that believed to be present in galaxies gives a fivefold reduction in the cross section and a similar decrease in the optical depth. As with the single lenses, three image opposed geometries predominate at modest amplifications, but five-image geometries, in which the two brightest are situated close to one another, predominate at larger magnification.

An important feature of gravitational lenses, termed amplification bias by TOG, is the tendency for relatively rare high amplification events to dominate a flux-limited survey, simply because the density of intrinsically faint quasars increases so rapidly. However, in order for this to occur, the slope of the quasar counts must exceed ~ 6 per magnitude. Recent quasar surveys strongly suggest that faint quasar counts have a flatter slope fainter than $B \sim 19.5^m$ and so amplification bias can only be of importance for the brightest

quasars and here it will only roughly double the expected number of lensing events. One corollary is that most lensing events would be expected to contain only modestly amplified images and that three-image opposed cases should predominate except for $B \lesssim 15^m$. It is of interest that the only five-image case we have is the brightest example, Q1115+080 (Weymann et al., 1980). Another corollary is that if our lens models are appropriate, then the large separation lens candidates, 2345+007 and 1635+267 (Weedman et al., 1982, Djorgovski, and Spinrad, 1984) would probably be produced by potential wells with velocity dispersion $\sim 500 \mathrm{kms}^{-1}$, rather larger than those normally associated with galaxies. (However, these quasar pairs may also be distinct objects; cf., Phinney, and Blandford, 1986.)

As discussed in Paper 1, the quasar 2237+0305, which appears to be split into two (or maybe three) images by a $z_L = 0.04$ Zwicky galaxy is probably an example of a locally marginal, but intrinsically strong lens. We estimate the probability P_{lens} that a given galaxy at $z_L = 0.04$ multiply images a quasar. Now from the image arrangement (Tyson, and Gorenstein, 1986) we might estimate that $r_+/D_{OL} \sim 1$ arc-seconds (equivalent to $r_+ \sim 0.6$ kpc) and choosing $\beta \sim 3$ gives $s \sim 2$ kpc. This implies that the equivalent velocity dispersion of our model potential is $\sigma \sim 300$ km/s. Inspecting the integral cross section of Figure 9b of Paper 1, we see that the minimum image amplification is $\bar{M}_{12} \sim 15 \equiv 3^m$ and that the cross section falls $\propto \bar{M}_{12}^{-2}$ for greater magnifications. Now the quasar has magnitude $\sim 16.5^m$ and so it is most probable that we are lensing a $\sim 19.5^m$ quasar, as the source counts flatten at this point. There are approximately ten 19.5^m quasars per square degree and the lensing cross section is ~ 1 square arc-seconds implying that $P_{lens} \sim 10^{-6}$. Now roughly 10^4 galaxies have been examined in the redshift survey by which this example was discovered (Huchra, private communication) and this is the only gravitational lens. While we are undeniably rather fortunate to have observed a quasar in a nearby galaxy, the a priori odds against this happening may be as large as ~ 0.01.

The potentials used in this work were deliberately chosen to be of a simple form in order to facilitate an understanding of the physical effects contributing to gravitational lensing. The next step is to explore the properties of potentials that model real galaxies more closely. In particular, it will be important to use galaxies that model pairs, groups, and clusters of galaxies in order to assess the need to invoke non-galactic lenses. The advent of reportedly efficient automated optical quasar search techniques (e.g., Smith 1986) is a strong inducement to complete this exercise.

Appendix: Conversion to Astronomical Observables

We can relate the surface density to a velocity dispersion in the following manner. Given the potential as defined in equation (2.1) with ellipticity $\epsilon = 0$, we derive the surface mass density

$$\Sigma(x) = \Sigma(0)\left[\alpha(1 + r^2/s^2)^{\alpha-1} + (1 - \alpha)(1 + r^2/s^2)^{\alpha-2}\right] \qquad (A.1)$$

where

$$\Sigma(0) = A/2\pi G s^2. \qquad (A.2)$$

The mass density is

$$\rho(x) = -\frac{1}{\pi s}\int_x^\infty \frac{d\,\Sigma(x')}{d\,x'}\frac{d\,x'}{\sqrt{x' - x}} \qquad (A.3)$$

where $x = r^2/s^2$, which becomes

$$\rho = \frac{\Sigma(0)}{s}\frac{\Gamma(3/2 - \alpha)}{\Gamma(1/2)\Gamma(1 - \alpha)}\left[\alpha(1 + r^2/s^2)^{\alpha-3/2} + (3/2 - \alpha)(1 + r^2/s^2)^{\alpha-5/2}\right] \qquad (A.4)$$

for the surface density (A.1) where $\Gamma(\mu)$ is the gamma function. This determines the mass as a function of radius

$$M(r) = 4\pi\int_0^r \rho(r')r'^2 dr' \qquad (A.5)$$

$$= \frac{2\Gamma(1/2)\Gamma(3/2 - \alpha)}{\Gamma(1 - \alpha)}\frac{\Sigma(0)}{s}\frac{r^3}{(1 + r^2/s^2)^{3/2-\alpha}} \qquad (A.6)$$

The observed central velocity dispersion is

$$\sigma^2(0) = \frac{2}{\Sigma(0)}\int_0^\infty \frac{G\,\rho(r)\,M(r)}{r}\,dr \qquad (A.7)$$

$$= \frac{3}{2}G\Sigma(0)s\frac{\Gamma(3/2 - \alpha)\Gamma(3/2 - \alpha)\Gamma(1/2)\Gamma(3/2 - 2\alpha)}{\Gamma(1 - \alpha)\Gamma(1 - \alpha)\Gamma(3 - 2\alpha)}. \qquad (A.8)$$

This result is true for non-singular potentials ($s \neq 0$), but does not reduce to the singular isothermal sphere result ($\alpha = 1/2, s = 0$) because of a subtlety in the ordering of limits in the integral (A.7) and the limit of the scale length becoming zero. This allows us to relate the lens parameter β to physical parameters through equation (3.7),

$$\beta = \left[\left(\frac{D_{OL}D_{LS}}{D_{OS}}\frac{8\pi}{3}\frac{\Gamma(1 - \alpha)\Gamma(1 - \alpha)\Gamma(3 - 2\alpha)}{\Gamma(3/2 - \alpha)\Gamma(3/2 - \alpha)\Gamma(1/2)\Gamma(3/2 - 2\alpha)}\frac{\sigma^2(0)}{sc^2}\right)^{1/(1-\alpha)} - 1\right]^{-1/2}. \qquad (A.9)$$

This result for the central velocity dispersion is larger than the velocity dispersion in a singular isothermal sphere by a factor of 3/2. This is not unrealistic in that a number of observations have indicated that the velocity dispersion is higher in the cores of elliptical galaxies (Lauer, 1985, Kormendy, 1986). The asymptotic result (i.e., that from the singular isothermal sphere) is probably closer to the tabulated values which did not come from very high resolution imaging of the galactic core.

References

Beers, T., and Tonry, J., 1986, *Astrophys. J.*, **300**, 557.

Blandford, R.D., Phinney, E.S., and Narayan, R., 1986, *Astrophys. J.*, **313**, 23.

Blandford, R.D., and Kochanek, C.S., 1987, *Astrophys. J.*, **321**, 658.

Canizares, C.R., 1986, *MIT preprint.*

Djorgovski, S., and Spinrad, H., 1984, *Astrophys. J. (Letters)*, **282**, L1.

Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., and Perley, R., 1985, *Astr. J.*, **90**, 691.

Kaiser, N., 1984, *Astrophys. J. (Letters)*, **284**, L9.

Kormendy, J., 1986, *Dominion Astrophysical Observatory preprint.*

Kovner, I., 1986b, *Princeton University preprint.*

Lauer, T.R., 1985, *Astrophys. J.*, **292**, 104.

Narayan, R., Blandford, R., and Nityananda, R., 1984, *Nature*, **310**, 113.

Phinney, E.S., and Blandford, R., 1986, *Nature*, **321**, 569.

Shanks, T., Fong, R. and Boyle, R.J., *Quasars, Proc. IAU Symp. 119* , ed. G. Swarup, and V.K. Kapahi, (Reidel: Dordrecht), 1986.

Smith, M., *Quasars, Proc. IAU Symp. 119*, ed. G. Swarup, and V.K. Kapahi, (Reidel: Dordrecht), 1986.

Turner, E.L., Ostriker, J.P., and Gott, J.R., 1984, *Astrophys. J.*, **284**, 1.

Tyson, T., and Gorenstein, M., 1985, *Sky and Telescope*, **70**, 319.

Vietri, M., and Ostriker, J.P., 1983, *Astrophys. J.*, **267**, 488.

Vietri, M., 1985, *Astrophys. J.*, **283**, 343.

Weedman, D.W., Weymann, R.J., Green, R.F., and Heckman, T.M., 1982, *Astrophys. J. (Letters)*, **255**, L5.

Weymann, R.J., Latham, D., Angel, J.R.P., Green, R.F., Liebert, J.W., Turnshek, D.A., Turnshek, D.E., and Tyson, J.A., 1980, *Nature*, **285**, 641.

TABLE 1
Integral Optical Depths and Branching Ratios for a Source at $z_S = 2$

σ km/s	s kpc	ϵ	$\bar{M}_{12} \geq 1$ Optical Depth	Opposed	Allied	Five Image	$\bar{M}_{12} \geq 10$ Optical Depth	Opposed	Allied	Five Image
177	1	0.0	1.25×10^{-5}	0.884	0.116	....	1.12×10^{-5}	0.871	0.129	....
	1	0.2	2.04×10^{-5}	0.000	0.655	0.345	8.20×10^{-6}	0.000	0.762	0.238
	1/3	0.0	1.50×10^{-4}	0.991	0.009	....	8.51×10^{-6}	0.963	0.037	....
	1/3	0.2	1.41×10^{-4}	0.727	0.028	0.245	6.53×10^{-6}	0.044	0.151	0.804
	1/9	0.0	3.53×10^{-4}	0.999	0.001	....	7.93×10^{-6}	0.995	0.005	....
	1/9	0.2	3.22×10^{-4}	0.881	0.002	0.118	6.44×10^{-6}	0.134	0.010	0.855
306	3	0.0	4.84×10^{-5}	0.884	0.116	....	4.33×10^{-5}	0.870	0.130	....
	3	0.2	7.99×10^{-5}	0.000	0.658	0.342	3.22×10^{-5}	0.000	0.763	0.237
	1	0.0	5.91×10^{-4}	0.990	0.010	....	3.35×10^{-5}	0.962	0.038	....
	1	0.0	5.53×10^{-4}	0.726	0.029	0.245	2.58×10^{-5}	0.044	0.152	0.804
	1/3	0.0	1.39×10^{-3}	0.999	0.001	....	3.12×10^{-5}	0.995	0.005	....
	1/3	0.0	1.27×10^{-3}	0.881	0.002	0.118	2.54×10^{-5}	0.134	0.010	0.855
	1/9	0.0	1.64×10^{-3}	0.999	0.000	....	2.65×10^{-5}	0.995	0.005	....
	1/9	0.0	1.79×10^{-3}	0.858	0.000	0.142	2.52×10^{-5}	0.139	0.001	0.860
1000	33	0.0	3.40×10^{-6}	0.852	0.148	....	3.40×10^{-6}	0.852	0.148	....
	33	0.2	7.04×10^{-6}	0.000	0.738	0.262	3.11×10^{-6}	0.000	0.785	0.215
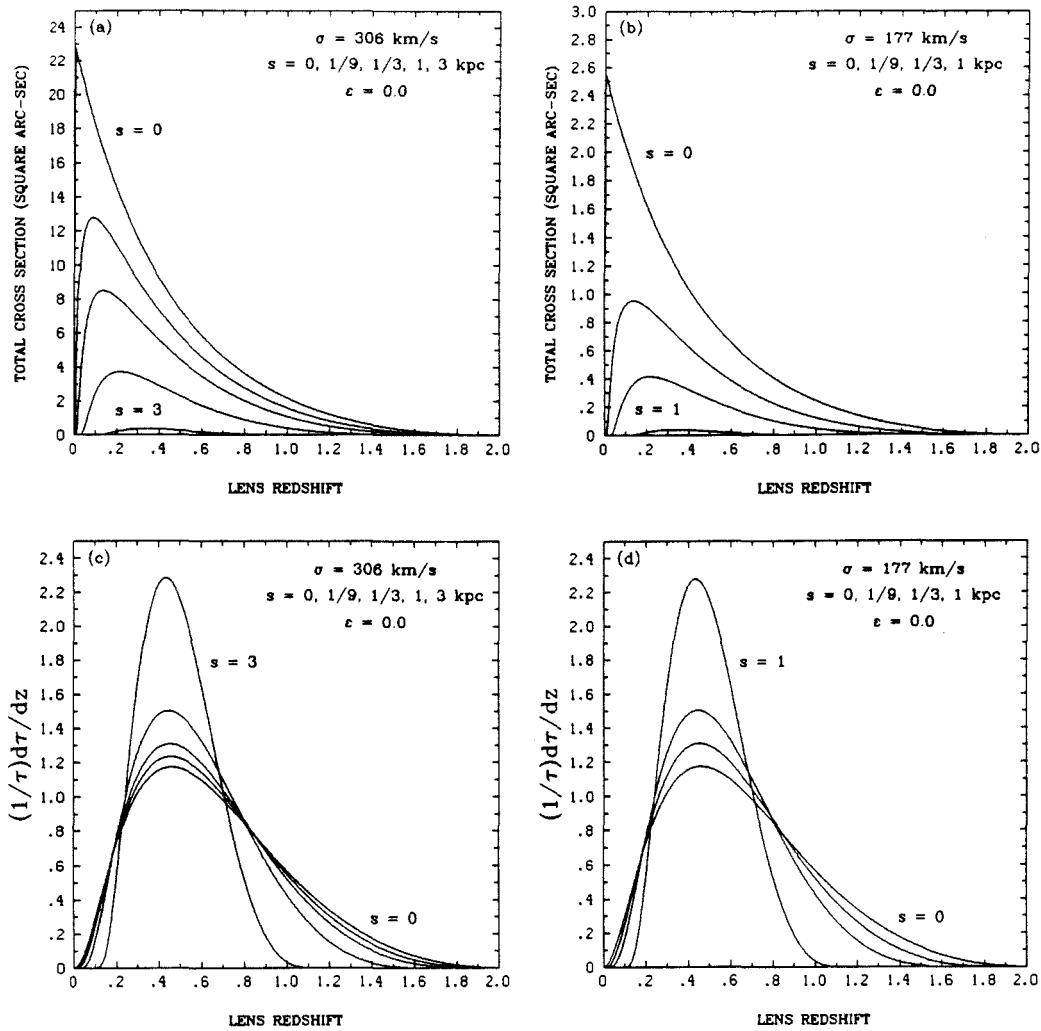
Figure 1. (a). Cross sections in square degrees for circular lenses with central velocity dispersions $\sigma = 306$ km/s and core sizes of $s = 0$, 1/3, 1, and 3 kpc. (b). Cross sections in square degrees for circular lenses with central velocity dispersions $\sigma = 177$ km/s and core sizes of $s = 0$, 1/3, and 1 kpc. (c). Normalized differential optical depth $(1/\tau)d\tau/dz_L$ for the circular lenses in (a). (d). Normalized differential optical depth $(1/\tau)d\tau/dz_L$ for the circular lenses in (b).
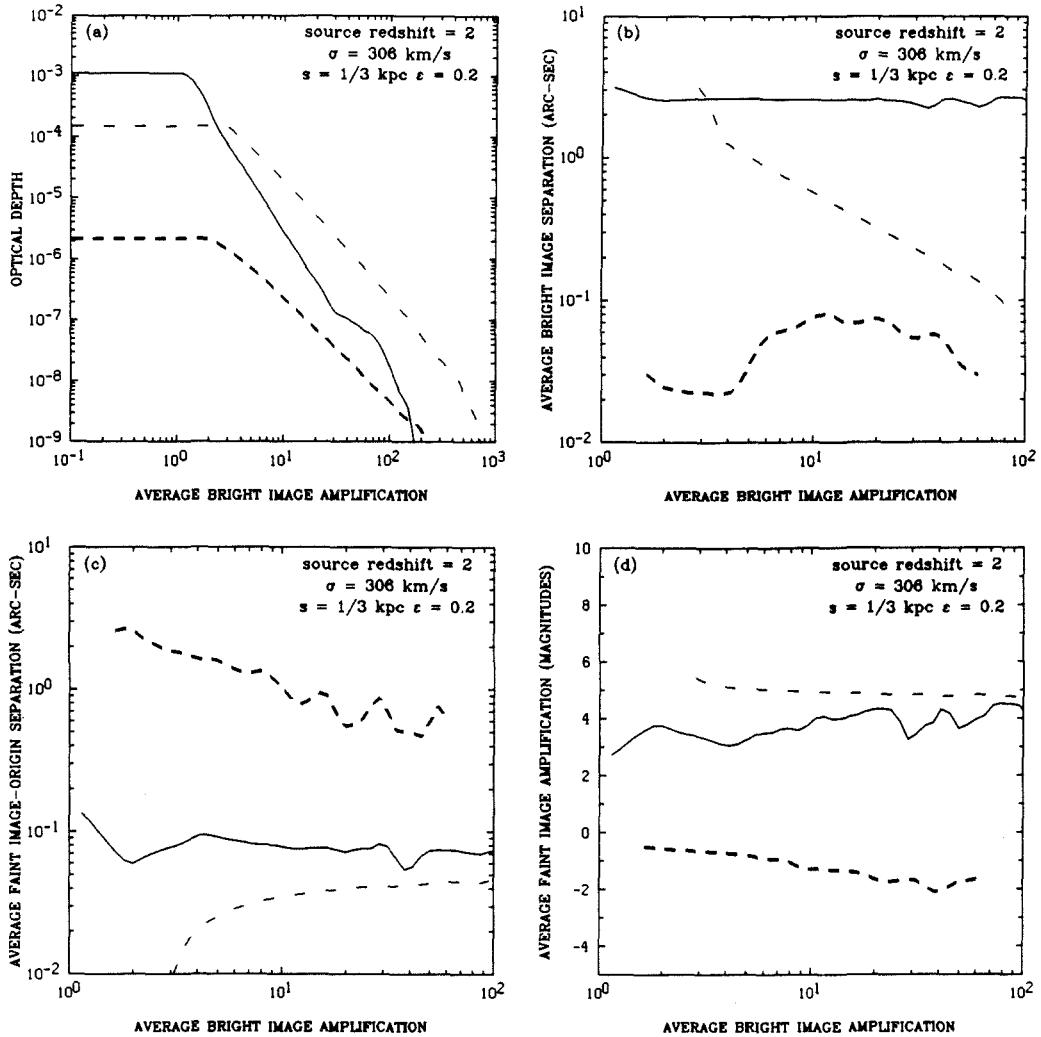
Figure 2. Optical depth weighted averages of image properties for a quasar at $z_S = 2$ behind a lens with $\sigma = 306$ km/s, $s = 1/3$ kpc, and $\epsilon = 0.2$. Data is presented for three-image opposed (continuous lines), three-image allied (heavy dashed lines), and five-image (faint dashed lines) geometries.

(a).   Integral Optical Depth for Magnification $\bar{M}_{12} > M$

(b).   Average Bright Image Separation $< r_{12} >$ in Arc-Seconds

(c).   Average Faint Image-Origin Separation in Arc-Seconds

(d).   Average Amplification of Faintest Image in Magnitudes

In each case, these average quantities are plotted as a function of the mean amplification of the brightest pair of images $\bar{M}_{12}$. Note that in (d), a positive number represents a *deamplification.*
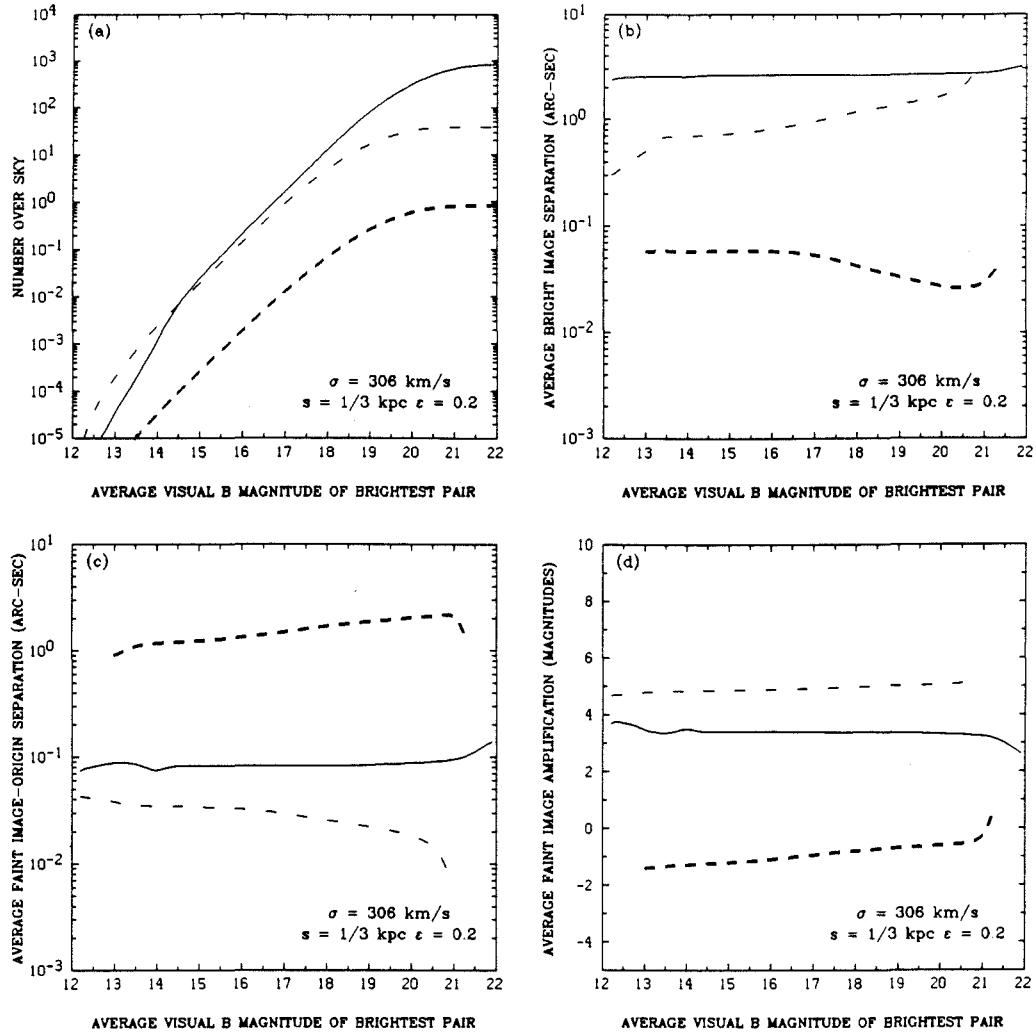
Figure 3. Image properites for the observed distribution of quasars and a lens model with $\sigma = 306$ km/s, $s = 1/3$ kpc, and $\epsilon = 0.2$. This is an example of an intrinsically strong lens. Notation as in **Figure 2**.

(a). Expected Number of Lenses of Magnitude $\bar{B}_{12} < B$

(b). Average Bright Image Separation $< r_{12} >$ in Arc-Seconds

(c). Average Faint Image-Origin Separation in Arc-Seconds

(d). Average Amplification of Faintest Image in Magnitudes

Each plotted as a function of the mean magnitude $\bar{B}_{12}$ of the brightest two images making up the lens.
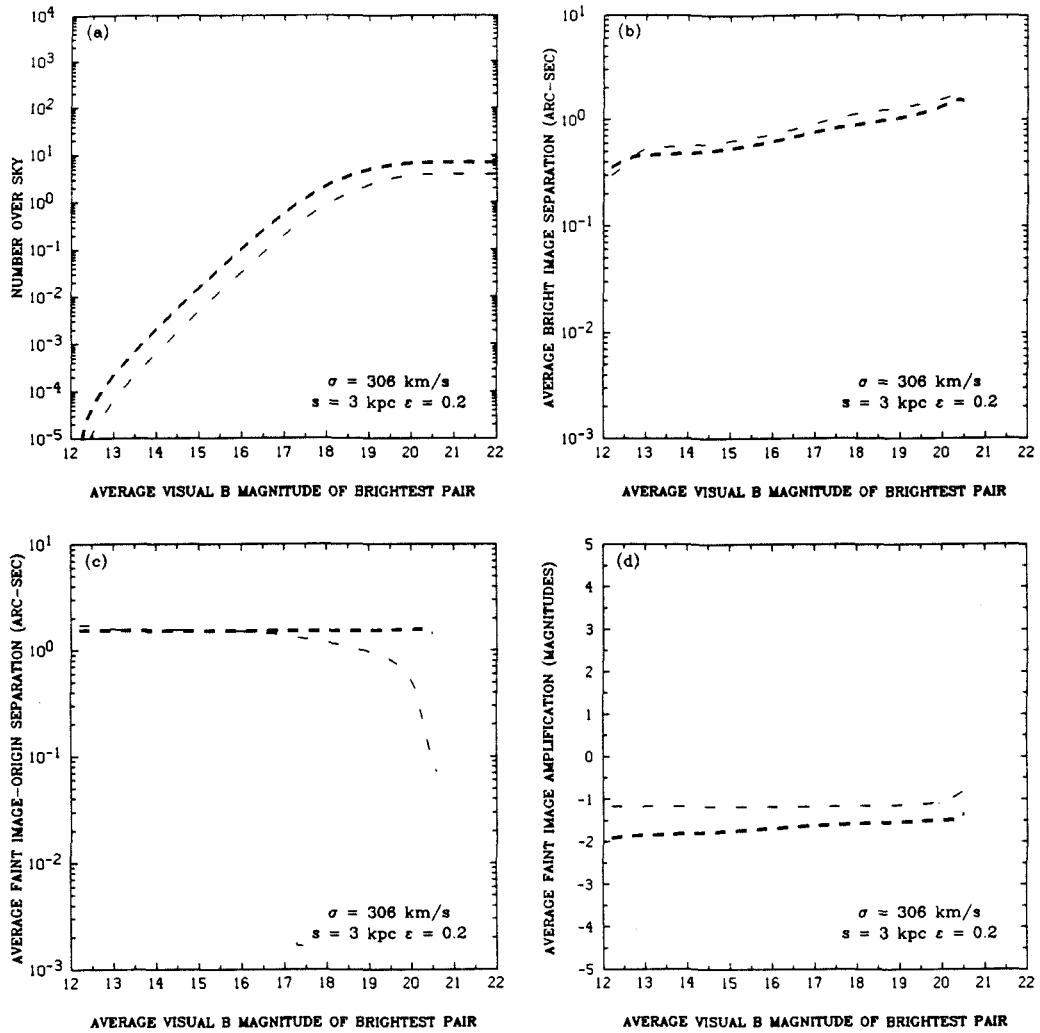
Figure 4. Image properties for the observed distribution of quasars and a lens model with $\sigma = 306$ km/s, $s = 3$ kpc, and $\epsilon = 0.2$. This is an example of an intrinsically marginal lens. This model also corresponds to the $\sigma = 177$ km/s, $s = 1$ kpc and the $\sigma = 1000$ km/s, $s = 33$ kpc models if the lengths and number densities are rescaled. Notation as in **Figure 2**.

(a).   Expected Number of Lenses of Magnitude $< \bar{B}_{12}$

(b).   Bright Image Separation r_{12} in Arc-Seconds

(c).   Faint Image-Origin Separation in Arc-Seconds

(d).   Amplification of Faintest Image in Magnitudes

Each plotted as a function of the mean magnitude $\bar{B}_{12}$ of the brightest two images making up the lens.

Chapter 4

THE TWO-SCREEN GRAVITATIONAL LENS

Christopher S. Kochanek,

and

John Apostolakis

Theoretical Astrophysics 130-33,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

Abstract

Gravitational lensing by systems of two galaxy-like potentials located at different redshifts and separations is studied. We show that there is roughly a 1-10% chance of a gravitational lens system being affected by two lenses at the same redshift, and that there is roughly a 1-10% chance of a gravitational lens system being affected by two lenses at different redshifts. We show that two lenses interact significantly for transverse separations $\lesssim 4r_{+,0}$ where $r_{+,0}$ is the radius of the outer critical line of the singular potential. Nearly aligned lenses at different redshifts show interactions over almost all of redshift space. The interactions can lead to either increases or decreases in the total cross section relative to the cross section of two isolated lenses depending on the system's geometry, and the amplification of the brightest images. The behavior of the image structures is dominated by the effects of caustics, and in particular the fold caustics, as in the single screen case. Typical image separations are $\sim 2r_{+,0}$ as in the isolated lenses; larger separations are found only as the distance between the most widely separated images in the lens, in which case the separations approach $\sim 4r_{+,0}$. This means that "chaining" marginal lenses together cannot explain the so called dark matter lenses $2345+007$, $1635+267$, and $0023+171$ for which there are only two widely separated images. Highly amplified images are almost always part of a high image multiplicity system (5 or 7 images rather than 3 images). This implies that amplification bias is associated with high image multiplicities.

1. Introduction

At present, there is no observed lens system that has been unequivocally shown to require two lenses at appreciably different redshifts. There are, however, two systems that show indications that any final model may need to take this into account. The two-screen lens has been proposed as a means of modeling the $2016 + 112$ lens system, for which existing models using a single screen model have either failed to match other features of the system, or seem overly contrived (Narasimha et al., 1986). The lens system $2237 + 0305$ consists of four images of a quasar lying inside a large nearby spiral galaxy ($\sigma = 300$ km s^{-1}, $z = 0.04$). However there is an absorption line in the quasar spectra corresponding to a redshift of $z = 0.6$, which may indicate the presence of a second lens in the system (Huchra et al., 1985). Multiple screen geometries have also been invoked to explain the "dark matter" lenses ($2345+007$ Weedman et al., 1982, $1635+267$, Djorgovski, and Spinrad, 1984, and $0023 + 171$, Hewitt et al., 1987), by using aligned, non-luminous, galactic halos at different redshifts to generate the lensing effect (Subramanian, Rees, and Chitre, 1987). On the theoretical plane, the multiple screen lensing equations have been derived and discussed by Blandford and Narayan (1987), Schneider and Borgeest (1986), Kovner (1987), and Padmanabhan and Subramanian (1988). The case of two lenses on the same plane is frequently used in efforts to model observed lens candidates, and there has been one in-depth study in the case of a two point mass lens (Schneider, 1986).

In this paper we will give a broad overview of the two-screen gravitational lens, and how it relates to the known behavior of the single screen lens. The work consists primarily of a survey of the statistical properties of the double lens similar to the work in Blandford and Kochanek (1987, hereafter, BK) and Kochanek and Blandford (1987, hereafter KB) on the isolated lens. Because there is no overall picture of the behavior of these systems we also show examples of the image geometries that occur, and the caustic structures that characterize the lenses. In §2 we introduce the lensing equations for a two-screen lens and comment on how they differ from the single screen equations. In §3 we introduce the model potential that will be used to characterize the lenses in this survey. In §4 we discuss the probability with which a two-lens geometry occurs either for two lenses at the same redshift, or for two lenses at different redshifts. In §5 we discuss the numerical technique, and the limitations of the survey. In §6 we discuss the behavior of two lenses at the same redshift. While interesting in itself, this serves as a baseline from which to discuss the effects of putting the second lens at a different redshift. In §7 we discuss the characteristics of two aligned lenses at different redshifts. This provides a second baseline for the general case, and shows where the deviations from simple analytic models begin to develop. In §8 we discuss the results from studies of lensing with two unaligned lenses at different redshifts. These are compared to the results form §6 and §7. In §9, we consider the effect of relaxing the symmetry restrictions on the potentials for a few restricted cases to show that the results are not strongly dependent on the symmetry of the potentials used in §6, §7, and §8. Finally in §10 we summarize the results and discuss the implications of the two-screen lens for modeling observed lens systems.

2. The Compound Lens

Suppose we have a system of two lenses located on "screens" called 1 and 2 at redshifts z_1 and z_2 ($z_1 < z_2$) with the observer at point O and the source at S. Then the time delay along a virtual path parametrized by the angles of the points where the rays intersect the

screens $(\vec{\theta}_i)$, is the sum of the contributions from each screen

$$t = (1 + z_1)\left[(\vec{\theta}_1 - \vec{\theta}_2)^2 \frac{D_{O1}D_{O2}}{D_{12}} - 2\phi_1(\vec{\theta}_1)\right]$$
$$(1 + z_2)\left[(\vec{\theta}_2 - \vec{\theta}_S)^2 \frac{D_{O2}D_{OS}}{D_{2S}} - 2\phi_2(\vec{\theta}_2)\right]$$

(2.1)

where the D_{ij} are angular diameter distances and ϕ_i is the two-dimensional Newtonian potential of the ith screen (Bourassa and Kantowski, 1975, but for more modern formulations, see Blandford and Narayan, 1986 or Schneider and Weiss, 1987). We will work in an $\Omega = 1$ Einstein-DeSitter universe for which the angular diameter distance between points i and j at redshifts z_i and z_j $(z_i < z_j)$ is

$$D_{ij} = \frac{2c}{H_0} \frac{x_j x_i^{1/2} - x_j^{1/2} x_i}{x_j^2 x_i} \qquad x_i = 1 + z_i \qquad x_j = 1 + z_j$$

(2.2)

We will use the Einstein-DeSitter cosmology throughout, as the effects of changing the cosmology tend to be small relative to changes in the lenses or their redshifts. The position of the image for a given source is found by solving the four variational equations so that the virtual time delay is stationary with respect to variations in the four coordinates $\vec{\theta}_1$ and $\vec{\theta}_2$. This implies

$$\vec{\theta}_2 = \vec{\theta}_1 - \frac{2D_{12}}{D_{O2}}\vec{\alpha}_1(\vec{\theta}_1)$$
$$\vec{\theta}_S = \vec{\theta}_2 - \frac{2D_{2S}}{D_{OS}}\vec{\alpha}_2(\vec{\theta}_2) - 2\frac{1+z_1}{1+z_2}\frac{D_{O1}D_{2S}}{D_{O2}D_{OS}}\vec{\alpha}_1(\vec{\theta}_1)$$

(2.3)

where

$$\vec{\alpha}_i \equiv \vec{\nabla}_{x_i}\phi_i(\vec{x}_i) = \frac{1}{D_{Oi}}\vec{\nabla}_{\theta_i}\phi_i(\vec{\theta}_i)$$

(2.4)

The gradient operator $\vec{\nabla}$ is the two-dimensional gradient operator in the plane of the screen. This allows a full solution for the source position $\vec{\theta}_S$ in terms of the observed image position $\vec{\theta}_1$

$$\vec{\theta}_S = \vec{\theta}_1 - \frac{2D_{1S}}{D_{OS}}\vec{\alpha}_1(\vec{\theta}_1) - \frac{2D_{2S}}{D_{OS}}\vec{\alpha}_2(\vec{\theta}_2)$$

(2.5)

Here we have used the requirement that the equation reduce to the single screen equation in the absence of a second lens screen to simplify the cosmological terms. This reduces to the single screen lens equation in the limit that $D_{12} \rightarrow 0$, $D_{O2} = D_{O1}$, and $D_{2S} = D_{1S}$. The cosmological terms appearing in equations (2.3) and (2.5) are shown in Figure 1. Note that the equation for $\vec{\theta}_2$ in (2.5) also governs the relation between the true position of the second lens $(\vec{\theta}_2)$ and the images of it $(\vec{\theta}_1)$ seen by the observer after the rays interact with the first lens.

The important difference between the two-screen lens equation and the one-screen equation is that the net deflection angle cannot be described by the gradient of a single two-dimensional potential. If we try to reduce the problem to a single deflection by an effective potential on screen 1 we must include a vector potential $\vec{\beta} = \beta \, \mathbf{e}_z$ in addition to the effective potential ϕ_e.

$$\vec{\theta}_S = \vec{\theta}_1 - 2\frac{D_{1S}}{D_{OS}}\left[\vec{\nabla}\phi_e + \vec{\nabla} \times \vec{\beta}\right]$$

(2.6)

We can determine the source terms for the two potentials by taking the divergence and curl of the equation (note that $\nabla \cdot \vec{\beta} = 0$ by definition because $\vec{\beta}$ is perpendicular to the lens plane, and $\vec{\nabla}$ lies in the lens plane) so that

$$\sigma_e = \sigma_1(\vec{x}_1) + \frac{D_{2S}}{D_{1S}}\sigma_2(\vec{x}_2) - \frac{D_{2S}}{D_{1S}}\left(\frac{2D_{O1}D_{12}}{D_{O2}}\right)\left[\phi_1{}_{,xx}\,\phi_2{}_{,xx} + \phi_1{}_{,yy}\,\phi_2{}_{,yy} - \phi_1{}_{,xy}\,\phi_2{}_{,xy}\right]$$
(2.7)

and

$$\nabla^2\beta = \frac{D_{2S}}{D_{1S}}\left(\frac{2D_{O1}D_{12}}{D_{O2}}\right)\left[(\phi_1{}_{,yy} - \phi_1{}_{,xx})\phi_2{}_{,xy} + (\phi_2{}_{,xx} - \phi_2{}_{,yy})\phi_1{}_{,xy}\right]$$
(2.8)

where σ_i is the surface mass density in screen i, and σ_e is the effective surface mass density $\nabla^2\phi_e = 4\pi G\sigma_e$. The potentials on the first screen are evaluated at $\vec{x}_1$ and the potentials on the second screen are evaluated at $\vec{x}_2$. Notice that in the limit of the two screens merging into a single screen at some redshift, the cross terms between the potentials vanish, and

$$\begin{aligned}\nabla^2\phi_e &= 4\pi G(\sigma_1 + \sigma_2)\\ \nabla^2\beta &= 0,\end{aligned}$$
(2.9)

which is exactly the result for a single screen. In Figure 2, we show the contours of the effective potentials of the second screen for two cases. In one case the background lens is not multiply-imaged by the foreground galaxy, and in the second case, it is multiply imaged by the foreground galaxy. In general, the second order terms in $\phi_2{}_{,ij}\,\phi_1{}_{,kl}$ are small compared to the first order terms. The second order terms are most important in the core of the foreground galaxy where the effective surface density of the background galaxy is typically negative.

The addition of a vector potential to the deflection makes the magnification tensor asymmetric

$$M^{-1} = \begin{vmatrix} 1 - \beta_{,xy} - \phi_{,xx} & -\beta_{,yy} - \phi_{,xy} \\ -\beta_{,xx} - \phi_{,xy} & 1 - \beta_{,xy} - \phi_{,yy} \end{vmatrix}$$
(2.10)

so that the amplification now consists of a rotation in addition to the expansion and shear of the images that occur for a single lens screen. Note that a second screen is trivial if its second derivatives are constant (for a proof, see Kovner, 1987).

3. The Model Potential

The introduction of a second lens at a different redshift causes a vast expansion of parameter space. Because the primary goal of this paper is to study the effects of two deflecting screens rather than the properties of different potentials (see BK and KB) we will restrict ourselves to a simple spherical potential which at large distances from the core asymptotically approaches the potential of a singular isothermal sphere. The two-dimensional Newtonian potential is

$$\phi = A\left(1 + r^2/s^2\right)^{1/2}$$
(3.1)

where the potential depth A is related to the velocity dispersion σ_v and core size s by $A = 2\pi s\sigma_v^2$. While the ellipticity of lensing potentials is important both qualitatively and quantitatively to the cross sections for gravitational lensing, we will not make the lenses

intrinsically elliptical in this paper except for a brief discussion in §9. The degenerate symmetry of the spherical potentials is broken by the tidal fields of the other lens in the regimes where there is significant interaction between the potentials.

Typical galaxies have core sizes, s of approximately 0.1 to 1.0 kpc and velocity dispersions σ_v of 100 to 300 km s^{-1}. As isolated lenses, galaxies can be divided into two classes, the intrinsically strong and intrinsically marginal lenses. A lens is *intrinsically marginal* if it is only marginally capable of generating multiple images for all redshifts, and it is *intrinsically strong* otherwise. This can be measured by a dimensionless strength parameter

$$\beta = [b^2 - 1]^{-1/2} \qquad b = [\beta^{-2} + 1]^{1/2} \tag{3.2}$$

where

$$b = \frac{2AD_{OL}D_{LS}}{c^2 s^2 D_{OS}} = 4\pi \frac{D_{OL}D_{LS}}{s D_{OS}} \left(\frac{\sigma_v}{c}\right)^2 \tag{3.3}$$

where β is the ratio s/r_+ and r_+ is the radius of the outer critical line of the isolated potential. Strong lenses have $\beta \lesssim 0.5$ for some redshifts, while marginal lenses have $\beta \gtrsim 0.5$ for all redshifts. Intrinsically strong lenses can, of course, be *locally* marginal depending on the position of the observer, the lens, and the source. For the purposes of illustration, we will use a $\sigma_v = 220$ km s^{-1}, $s = 1/4$ kpc galaxy as a model of an intrinsically strong lens, and a $\sigma_v = 220$ km s^{-1}, $s = 2$ kpc galaxy as a model of an intrinsically marginal lens.

There are several useful quantities from the isolated circular lens, which appear as scales in the two-screen lenses. The first is the radius of the outer critical line, and its limiting value for the case of a singular lens,

$$\begin{aligned} r_+ &= s[b^2 - 1]^{1/2} = s\beta^{-1} \\ r_{+,0} &= sb = 4\pi \frac{D_{OL}D_{LS}}{D_{OS}} \left(\frac{\sigma_v}{c}\right)^2 \end{aligned} \tag{3.4}$$

The inner critical line is located at

$$r_- = s[b^{2/3} - 1]^{1/2} \tag{3.5}$$

The critical lines are images of caustics on the source plane. For the circular lens, the u_+ caustic corresponding to the r_+ critical line is degenerate, $u_+ = 0$. The outer caustic is located at

$$u_- = s[b^{2/3} - 1]^{3/2} \tag{3.6}$$

and it is related to the total cross section of the isolated lens by

$$\sigma = \pi u_-^2 = \pi s^2 [b^{2/3} - 1]^3 \tag{3.7}$$

Note that up to a trivial scaling of the length scale, lens systems are identical if the quantity σ_v^2/s is held constant. This holds true for the two-screen lens if both lenses are scaled identically so that the ratios of the length scales (r_+/s) of the two lenses are not changed. Thus, a pair of $\sigma_v = 220$ km s^{-1}, $s = 2$ kpc is in principal identical to a pair of $\sigma_v = 110$ km s^{-1} galaxies with $s = 0.25$ kpc if all lengths in the problem are reduced by a factor of 4.

4. The Probability of Superpositions

Two lenses may be superposed either at redshifts that are nearly the same, or at redshifts that differ substantially. We will consider the likelihood of the two types of superpositions separately. The first is the likelihood of two lenses lying at the same redshift being separated by less than angle θ, and the second is the likelihood of two lenses at different redshifts being separated by less than angle θ. In the first case, the correlations of lenses must be included, whereas in the second case, the distribution of lenses at very different redshifts is assumed to be uncorrelated. In all cases, we assume a constant co-moving density of lenses (on average) so that the proper number density as a function of redshift is

$$n(z_L) = n_0 x^3 \qquad n_0 \simeq 10^{-2} h^3 \text{ Mpc}^{-3} \qquad x = 1 + z_L \qquad (4.1)$$

for galaxies, where $h = H_0/100$ km s^{-1}Mpc^{-1}. Note that probabilities, P, will be independent of the value of H_0 because $P \sim ndV$ where $n \sim h^3$ and the volume element $dV \sim h^{-3}$. The galaxy-galaxy correlation function is fairly well known, so that relative to a galaxy at redshift z the local density of galaxies is (*cf.*, Phinney and Blandford, 1986)

$$n(r) = n_0 x^3 \left[1 + \xi(r) \right] \qquad (4.2)$$

where the correlation function is

$$\xi(r) = (r/r_0)^{-\alpha} \qquad r_0 \simeq 5x^{-1}h^{-1} \text{ Mpc} \qquad \alpha \simeq 1.8 \qquad (4.3)$$

The underlying cosmology is again assumed to be an Einstein-DeSitter $\Omega = 1$ universe.

To obtain the probability of a second lens galaxy lying within angle θ of the first lens galaxy, we will split the discussion into the correlated N_c and uncorrelated N_{uc} parts of the expectation value. Near the lens the forward and backward light cones going to the observer and source planes form a region that is nearly cylindrical (because the distances are much smaller than the cosmological distances over which the cone changes size) and locally Euclidean. The expected number of galaxies within a cylindrical region of radius $r = D_{OL}\theta$ and length $2l$ due to the uncorrelated part of $n(r)$ is

$$
\begin{aligned}
N_{uc} &= 2\pi n_0 x^3 r^2 l \\
&= 2.66 \, 10^{-5} \, x \left(1 - \frac{1}{x^{1/2}} \right)^2 \theta^2(") \, l_{\text{Mpc}} h^{-3} \frac{n_0}{10^{-2} \text{Mpc}^{-3} h^{-3}}
\end{aligned}
\qquad (4.4)
$$

For the correlated part we will make use of the fact that the correlations are local and peaked on scales smaller than r_0 so that we can allow the limits of the line of sight integral to extend to infinity. The expected number of galaxies due to the correlated part of the probability is

$$
\begin{aligned}
N_c &= \frac{2\pi}{3-\alpha} n_0 x^3 r_0^3 \left(\frac{D_{OL}\theta}{r_0} \right)^{3-\alpha} B\left(\frac{1}{2}, \frac{1-\alpha}{2} \right) \\
&= 0.050 \left(1 - \frac{1}{x^{1/2}} \right)^{1.2} \theta^{1.2}(")h^{-3} \frac{n_0}{10^{-2}\text{Mpc}^{-3}h^{-3}}
\end{aligned}
\qquad (4.5)
$$

where $B(\mu, \nu)$ is the Beta function. Note that for a lens at $z_L = 0.5$ and on scales of $\theta = 10''$, the length l in equation (4.4) must be equal to ~ 230 Mpc for N_{uc} to be as large

as N_c. If we expand the angular diameter distance (2.2) for small redshift separations Δz, we can relate the distance l in N_{uc} to a separation in redshift,

$$\Delta z \simeq (1+z)^{5/2} \frac{l\,\mathrm{Mpc}}{3000 h^{-1}} \tag{4.6}$$

so that a separation of $100\,\mathrm{Mpc}$ at a redshift of $z = 0.5$ corresponds to a $\Delta z \sim 0.1$. The angular scale for lensing by galaxies is $\theta \sim 1''$, which means that for scales l in equation (4.4) that are not cosmological, the correlated probability is always the dominant contribution to the probability of having a second lensing galaxy at the same redshift.

We can estimate the probability of there being a second lens at a different redshift by calculating the expected number of galaxies inside the double cone formed by a photon path which leaves the source, goes to the lens plane of the first lens, and is deflected so as to reach the observer with angular distance θ from the center of the lens. The result of this calculation is

$$\begin{aligned}
N(\theta) &= \frac{8\pi}{3} n_0 \left(\frac{c}{H_0}\right)^3 \theta^2 \left(1 - \frac{1}{x^{1/2}}\right)^2 \left(1 - \frac{1}{y^{1/2}}\right) \\
&= 0.053\,\theta^2('') \frac{n(0)}{10^{-2} h^{-3}\mathrm{Mpc}^{-3}} h^{-3} \left(1 - \frac{1}{x^{1/2}}\right)^2 \left(1 - \frac{1}{y^{1/2}}\right)
\end{aligned} \tag{4.7}$$

The probability of the second galaxy lying at a different redshift equals the probability of the second galaxy lying at the same redshift when $\theta \simeq 15''$ when $z_L = 0.5$ and $z_S = 2$. The angle that should be used to estimate the probability that any particular lens has a second lens should roughly be the radius of the region encompassing the images plus roughly a critical radius for a typical lens at that redshift. The critical radius of the isolated lens is roughly the distance at which an image is strongly affected by the presence of the lens. Recall from §2 that the critical radius for the singular isothermal sphere is

$$\frac{r_+}{D_{OL}} = .29''\,\sigma_{100}^2\,\frac{x^{-1/2} - y^{-1/2}}{1 - y^{-1/2}} \tag{4.8}$$

In Figure 3, we show the probability contours for finding a second lens at the same redshift or at a different redshift as a function of the distance from the first lens as compared to the scale of the critical radius r_+. Note that for the fiducial value of n_0, the typical probability on a scale of $5r_{+,0}$ is approximately one percent. However, this neglects the effects of changes in the cross section due to aligning the lenses, and the uncertainties in n_0.

5. The Computational Method and the Parameter Survey

The algorithm used to study the two-screen lenses is the same as that described in BK. Rays are traced from the observer back to the source rather than *vice versa* because the map from image to source is one-to-one whereas the map from source to images is not. For the two-screen lens this is done in two steps: first the ray is traced to the second lens screen so that the deflection at the second screen can be computed, and then the total deflection is computed to find the position on the source plane. The amplification is determined by the ratio of the areas of a triangle formed from the three corners of a grid

zone on the image plane and the area of its projection on the source plane. Unlike the "ray-shooting method" of Schneider and Weiss (1988), which only computes the amplification as a function of source position, we retain all of the information about the image structure so that the statistical structure of the image population can be studied.

The price for retaining all of the information about the lens is that it is considerably more difficult to use this algorithm at high spatial resolutions because of the large amount of memory required to retain the data structures. To completely resolve the structure of the lens, we require a minimum dynamical range of 10^6 in amplification, and 10^3 in spatial resolution. This requires a grid on the image and source planes of at least 1000 by 1000 in size. For a maximum of M_i images in a lens system on an N by N grid, we require $4N^2(2 + M_i)$ bytes of memory, or 36 Mbytes for the minimum resolution with $N = 1000$ and $M_i = 7$. This is too large to be run efficiently on any conventional computer available to us because of the large parameter space that must be surveyed. For this reason, the algorithm was rewritten to use the Caltech/JPL Mark III hypercube, a parallel computer, which has 32 processors, each of which is an 8 MHz 68020/68881 microprocessor pair with 4 Mbytes of memory. The survey was performed on grids with a resolution of 1536 by 1536 zones, which translates into 85 Mbytes of memory for the main data structures. There are a number of interesting programming issues involved in making the ray-tracing algorithm work on the hypercube–interested readers should see Apostolakis and Kochanek (1988). We did not have access to a parallel contour generating routine, which forced us to use low resolution (256 by 256 instead of 1536 by 1536) runs on a conventional machine to generate the contour plots and half-tone images. This leads to some spurious results in the figures such as the polygonal caustics for zero lens separation seen in Figure 5. The caustics are real, but they should be circular- however, even the slightest perturbation leads to the formation of theses "proto-cusp" structures.

The parameter space in which this survey could be performed is enormous. There are four parameters defining the individual potentials: the velocity dispersion σ, the core size s, the ellipticity ϵ, and the orientation angle θ. Then there are the three extrinsic parameters: the redshifts of the two lenses z_1 and z_2, and the angular separation of the potentials l. The parameters we wish to study are the effects of separation and redshift. For simplicity we will ignore ellipticity until §9, and we will then discuss it only briefly. From the work in BK, we know that there is a fundamental dichotomy between intrinsically strong and intrinsically marginal potentials, so we will restrict our study to a pair of identical strong potentials, and a pair of identical weak potentials. A strong-weak combination is unlikely to be interesting because the weak potential can only have a major effect if the intrinsically strong potential is in a redshift region where it is locally weak.

The survey is then broken up into four parts. The first part, in §6, discusses the behavior of two identical lenses at the same redshift. In §7, we discuss the behavior of two aligned lenses at different redshifts. These two cases are useful because they contain some analytically tractable features, and because they will provide a baseline from which to study non-aligned lenses at different redshifts, the results of which are presented in §8. For these cases we have restricted the space studied still further by examining only two sequences in §8: a sequence in which the lenses are fixed at different redshifts but with varying separations, and a sequence in which the lenses are at fixed separation but with varying redshifts. In all cases, one lens is held fixed at a redshift of 0.5. These restrictions are necessitated by the amount of detail that appears in the behavior of the two-screen lens. Over 100 sets of computations had to be performed for the sequences that are discussed in §6-8; it is simply not practical to cover the parameter space more

throughly. Finally, in §9 we will discuss the extent to which these results are generic when the symmetry conditions are relaxed, or if a different region of redshift space is sampled. In Figures 4a, b, we show the parameter space studied in this survey as compared to several relevant scales in the problem.

At this point we want to draw the reader's attention to a series of figures that summarize the results of this survey. In Figures 5 and 6 we show the changes in the caustic structure of the lenses as the parameters are varied for the strong and marginal lenses. Figures 7 and 8 show the total cross sections, and Figures 9 and 10 show the branching ratios for the strong and marginal lenses. Typically, the behavior of the lenses is different if only strongly amplified images are examined. Figures 11 and 12 show the cross sections, and Figures 13 and 14 show the branching ratios for lenses in which the brightest image has been amplified by more than a factor of ten. We will make frequent reference to these figures in the next three sections.

6. Two Lenses at the Same Redshift

There are two effects due to the superposition of two lensing potentials at a fixed redshift. The first is the change in cross section due to the non-linear interaction of the two potentials, and the second is the qualitative change in the lensing behavior due to the increased asymmetry of the overall potential. If the potentials described in §3 are isolated, they have a cross section for multiple imaging given by equation (3.7), $\sigma_0 = \pi s^2 (b^{2/3} - 1)^3$ where b, defined in §3 is the potentials strength. If $b < 1$ then the lens is subcritical and cannot generate multiple images. If the separation of the two lenses is large, then the lenses will behave as if they were isolated, with total cross section $\sigma_{l \gg 1} = 2\sigma_0$. If the potentials are exactly superposed $(l = 0)$, the strength of the combined lens system is $2b$, and the total cross section is $\sigma_{l=0} = \pi s^2 ((2b)^{2/3} - 1)^3$. The relative gain in cross section for the aligned lenses over the isolated lenses is

$$\frac{\sigma_{l=0}}{\sigma_{l \gg 1}} = \frac{((2b)^{2/3} - 1)^3}{2(b^{2/3} - 1)^3}, \tag{6.1}$$

which has the interesting property that the gain is 2 in the limit of a singular lens ($b \to \infty$, $s \to 0$). For the cases studied here, the strong lens has $b = 11.2$ when both lenses are at $z = 0.5$ which leads to a gain of 2.6, whereas the marginal lens has $b = 1.4$ under the same conditions, for a gain of 30 when both lenses are at $z = 0.5$.

If we now separate the lenses by distance $l \sim r_+$, the lensing equations become analytically intractable except along the symmetry axis of the configuration. We can gain some understanding by considering the sequence of events that occur as the lenses approach one another. We must first understand how far away the lenses must be before they act like isolated lenses. In the far field of the isothermal potentials, the deflection angle is a constant equal to the deflection due to a singular isothermal sphere. We call this deflection $r_{+,0} = sbD_{OL}^{-1} = 4\pi(\sigma/c)^2 D_{LS}/D_{OS}$ in terms of angular units. If the two lenses are at large separations, the effect of one lens on the other is to deflect all rays through the lens by a fixed amount, or equivalently, to move the region of the source plane associated with the lens from a position axisymmetrically behind the lens to $r_{+,0}$ in the direction of the second lens. In this simple picture, the lenses will start to interact in a non-perturbative fashion when the two multiply-imaged source plane regions of the isolated lens start to overlap, which occurs when $l \simeq 4r_{+,0}$ for a pair of singular lenses.

The second effect is the tidal perturbation by the distant second lens. A reasonable measure of this is to look at the change in the potential of the distant lens across the lens

relative to the potential–this is a measure of the deformation of the time delay surface by the second potential. We are interested in the second order effect because the first order effect is simply the uniform deflection discussed above. Therefore, at potential 2 we can estimate the ellipticity induced by potential 1 to be

$$
\epsilon \simeq \frac{|\partial_x^2 \psi_1(l)|(2r_{+,0})^2}{\psi_2(0)} \simeq \frac{s(2r_{+,0})^2}{(s^2 + l^2)^{3/2}}
$$
$$
\rightarrow \left(\frac{2r_{+,0}}{l}\right)^2 \left(\frac{s}{l}\right) \qquad l \gg s \tag{6.2}
$$

When we perform the same expansion for small separations, we find

$$
\epsilon \simeq \left| \frac{[\psi_{xx}(l) - \psi_{xx}(0)](2r_{+,0})^2}{\psi(0)} \right| \simeq 12\frac{l^2 s}{(2r_{+,0})^3} \tag{6.3}
$$

From BK, we know that the effects of ellipticity are always important for their effects on the branching ratios, and that for marginal lenses they are important for determining the total cross section of the lenses. This is a crucial fact for interpreting the marginal lens case because a small increase in ellipticity results in a dramatic increase in cross section for the marginal lenses. From BK, an ellipticity of 0.1 in a marginal lens comparable to the one used in this study led to a five-fold increase in the multiple imaging cross section relative to the circular lens.

The exact deflection equations can be solved easily only on the line perpendicular to the separation vector, along which the transverse deflections of the two potentials cancel. In this case, the equation for the position of the outer critical line, y_c, and the outer caustic, u_c, can be solved exactly,

$$
y_c = s_l \left[\left(\frac{2r_{+,0}}{s_l}\right)^{2/3} - 1 \right]^{1/2} \qquad u_c = s_l \left[\left(\frac{2r_{+,0}}{s_l}\right)^{2/3} - 1 \right]^{3/2} \qquad s_l^2 = s^2 + (l/2)^2 \tag{6.4}
$$

Note that there is no caustic connecting the two potentials if the term in brackets becomes negative (imaginary solutions), which means that the caustics of the isolated lenses merge when

$$
l \leq 2(4r_{+,0}^2 - s^2)^{1/2} \rightarrow 4r_{+,0} \qquad \text{if } s \ll r_{+,0} \tag{6.5}
$$

This confirms the simple picture of the interactions of the potentials becoming "non-linear" when the near uniform deflections by the second lens leads to having the previously isolated source regions overlap.

Unfortunately the equivalent equations for solutions along the direction of separation have no analytic solutions for non-singular lenses. If we restrict ourselves to the singular potential, then for $l > 4r_{+,0}$ we have two separate pairs of caustics corresponding to the two separate lenses. The separation of the centers of the caustics is $l - 2r_{+,0}$ and the width of the regions is $r_{+,0}$. This is in keeping with the picture of two perturbed but separate lenses in this region. When $2r_{+,0} < l < 4r_{+,0}$ the caustic systems are merging with an outer caustic at distance $l/2$, and an inner caustic at $2r_{+,0} - l/2$. Finally, for $l < 2r_{+,0}$ the outer caustic is at $2r_{+,0} - l/2$ and the inner caustic is at l. The position of the second interior caustic gives an estimate of the area of the lens that will generate more than three images. When the lenses are isolated they are mostly unaffected by the

second lens so that to first order they will generate only three-image configurations. When the caustics overlap, the overlap region generates images in both lenses so that the area of overlap is roughly the fraction of the area that will generate more than three images. Once $l < 2r_{+,0}$ we can use the position of the inner caustic of the singular lens to estimate the area of higher imaging multiplicities. This gives a simple model for the variation of the lens cross section with separation, and a simple model for the branching ratios for higher multiplicities,

Separation	Total Cross Section	Branching Ratio ($N > 3$)	
$l > 4r_{+,0}$	$2\pi r_{+,0}^2$	0	
$2r_{+,0} < l < 4r_{+,0}$	$\pi r_{+,0}^2(2 - (2 - (l/2r_{+,0}))^2)$	$(2 - (l/2r_{+,0}))^2$	(6.6)
$l < 2r_{+,0}$	$\pi(2r_{+,0} - l/2)^2$	$(l/2r_{+,0})^2$	

This estimate is approximately correct for strong lenses and qualitatively correct for marginal lenses. This estimate can be improved, particularly for $l < 2r_{+,0}$, by using the position of the transverse caustic in equation (6.4) so that $\sigma \simeq \pi u_-(u_- + l/2)$. In Figure 15, we show a comparison of the cross sections and branching ratios from this simple model to the results of the numerical simulations.

While this characterizes the total cross sections of the potential it tells us little about the cross section for generating bright images. In BK, we found that the cross section for the mean amplification of the brightest two images to be greater than M to be approximately $\pi r_+^2(1 + \beta^2)^2 M^{-2}$ where β was defined in §3. This result is derived from a power expansion about the critical lines of the circular potential. For strong lenses ($\beta \ll 1$) the high magnification cross section shows the same degree of enhancement as the total cross section, while for aligned marginal lenses the increase in r_+ is partially compensated for by a decrease in β. In this case, the cross section is increased by a factor ~ 3.5 relative to the isolated lenses, compared to the factor of 30 for the total cross section. This difference is in accord with the changes in r_+ and β.

Unlike the total cross section the high amplification cross section shows a constructive resonance at separations $l \simeq 2r_{+,0}$. The cross section as a function of amplification can be derived from a line integral about the critical lines of the lens. Defining $H = M^{-1}$, the inverse of the amplification, then the cross section for amplifying the brightest two images to greater than amplification M_0 is

$$\sigma(M_0) \simeq \frac{1}{2M_0^2} \int_{H=0} \frac{ds}{|\vec{\nabla} H|} \tag{6.7}$$

At the destructive resonance, the overall length of the critical lines ($\int ds$ along the critical lines) does not decrease as much as the change in the cross section suggests. Moreover, the magnitude of the gradient of the Hessian is reduced over much of the length of the fold because the average distance to the fold is now greater than r_+, and the gradients are sharply decreasing functions of the distance from the potential: $|\nabla H| \sim \psi_{xxx} \sim (x \pm l/2)(1 + (x \pm l/2)^2/s^2)^{-5/2}$. In general, the cross section for high amplification events will almost always be increased by the interactions of a second lens. The changes in the total cross section are the result of changing the low amplification end of the distribution where the M^{-2} rule breaks down.

As an example of two lenses at the same redshift, consider the case of two strong lenses separated by $r_{+,0}$. In Figure 16 we show the caustics, critical lines, and multiplicity

diagram for the lens. Notice the way the multiplicity diagram has a "banded" structure that confines images to very narrow regions of the lens plane for a given multiplicity. This is the cause of the nearly fixed inter-image distances seen in Figure 17b. On these diagrams we also show a typical seven image configuration for the lens. This case is actually somewhat simpler than others (such as the $l = 3r_{+,0}$ case in Figure 5) in which there are several different ways in which images can merge to reduce the total multiplicity. The seven-to-five image transition is characterized by tangential mergers on the outer critical line, the five-to-three image transition is characterized by radial mergers on the inner critical line, and the three-to-one image transition is characterized by radial mergers on the middle critical line.

The cross section for each multiplicity as a function of the amplification of the brightest image is shown in Figure 17a. At low amplifications, the cross section is dominated by the three-image configurations, while at high amplifications it is dominated by seven-image configurations. This behavior is generic both to the two screen lens and to the isolated elliptical lenses studied in BK. At high amplification, the cross section closely follows the M^{-2} law for the cross section derived from the expansions about a fold catastrophe (cf., equation (6.7)); the deviations from this rule occur only at low amplifications. If we examine the seven-image behavior more closely, we can see that the seven-image configuration is dominated by mergers on a fold caustic. While most of the images remain at fixed positions (relative to the brightest image) and amplifications, the two brightest images merge with increasingly high amplifications. The separation of the brightest images is proportional to M^{-1} where M is the amplification of the brightest image, and the amplification of the second brightest image approaches that of the brightest image. There is some indication of the cusps in the seven image caustic because the third brightest image shows some signs of approaching the brightest image, and having an increasing amplification as M increases. Even in this case, where the cusps form a large part of the caustic, the fold behavior dominates, and the statistics are dominated by that of the fold. The two faintest images are trapped in the cores of the two galaxies, and they have the characteristic amplification of the core of the galaxy $M(0) \simeq [b-1]^{-2}$. For the strong lens, this gives $M(0) \simeq 0.01$, which is approximately correct for the double lens, but there are corrections from the presence of the second lens that slightly reduce the deamplification by the core. The image separations relative to the brightest image, and the image amplifications as a function of the amplification of the brightest image are shown in Figures 17b and 17c.

7. Two Aligned Lenses At Different Redshifts

The next problem we will consider is that of two aligned lenses at varying redshifts. For the simple case of the singular isothermal sphere, this problem is analytically tractable. If we define the deflections by the singular potentials as

$$r_{12} = \frac{2D_{12}D_{O1}}{D_{O2}}2\pi\sigma^2 \qquad r_{1S} = \frac{2D_{1S}D_{O1}}{D_{OS}}2\pi\sigma^2 \qquad r_{2S} = \frac{2D_{2S}D_{O2}}{D_{OS}}2\pi\sigma^2 \qquad (7.1)$$

where r_{12} is the deflection by the foreground lens at the background lens, and r_{1S} and r_{2S} are the deflections by the foreground and background lenses at the source plane, the the lensing equations become

$$r_2 = r_1 \pm r_{12}$$
$$r_S = r_1 \pm r_{1S} \pm r_{2S} \qquad (7.2)$$

The choice of signs depends on which side of the center of the potential the ray is passing.

Given these relations it is easy to build an effective one-screen potential that characterizes the two-screen lens system (this is possible because of the axial symmetry of the system). If we construct a bending angle diagram for this system, there are three discontinuities in the bending angles that correspond to the positions of critical lines. Similarly, we can compute the positions of the caustics associated with the critical lines to be

$$u_1 = |r_{1S} + r_{2S} - r_{12}| \qquad u_2 = |r_{12} - |r_{1S} - r_{2S}|| \qquad u_3 = |r_{1S} - r_{2S}| \qquad (7.3)$$

The total cross section of the system is given by the area inside the outermost of these three caustic positions,

$$\sigma = \pi \max(u_1^2, u_2^2, u_3^2) \qquad (7.4)$$

The two interior caustics represent transitions to higher image multiplicities, which allows us to estimate the branching ratio for image multiplicities greater than three to be the ratio of the area inside the second largest caustic to the total cross section. The singular potential can produce only three or five images (with one, two, or three of the images captured in the singular cores). The cross sections and branching ratio for higher multiplicities for this simple model and for the simulations are compared in Figure 18. (Also see Figures 9b, 10b, 11b, and 12b.)

At high amplifications, however, the cross section is almost always larger than that of two isolated lenses. This is best understood as simply the same effect as that discussed in the previous section: the length of the critical lines stays constant or increases, while the slope of the Hessian is reduced at the critical line, leading to an increase in the high amplification cross section. High amplifications are again associated with high image multiplicities, so that the branching ratios for five and seven images are significantly enhanced. When the second lens is at small redshifts the high amplification cross section is nearly divergent because the small core size causes the lens to become subcritical only at very small redshifts. As a result the lens at low redshifts acts like a near critical sheet of density in front of the background lens leading to a strong amplification of the images. The marginal lens does not have this extreme peak at low redshifts because the lens becomes subcritical much faster than the strong lens. (See figures 11b, 12b, 13b, and 14b.)

As a particular example of this case, let us examine the properties of two strong lenses located at redshifts of 0.5 and 1.0 respectively. In Figure 19 we show the caustics, critical lines and image multiplicity diagram for this case. In Figure 20a, we show the cross sections for different multiplicities as a function of the amplification of the brightest image. As in the case examined in §6, at low amplifications, three-image systems dominate the cross section, and at high amplifications, seven-image systems dominate the cross section as can be seen in Figure 20a. This time, we will examine the behavior of the three-image configurations in more detail. At low amplifications ($M \lesssim 7$) the two brightest images have a large separation, and are roughly unamplified, while the third image is strongly deamplified. This configuration is what was called the three-image opposed geometry, which dominated the low amplification three-image cross section. At higher amplifications, ($M \gtrsim 7$) the geometry changes to the three-image allied geometry where two of the images are merging on the caustic while the third image remains approximately fixed in position and amplification. This is partially obscured in Figure 20b because no distinction is made in the statistical routine between different varieties of three-image configurations, but the transition between the two different geometries is clearly marked by the discontinuity in the separation of the two brightest images. Label the images in

order of increasing brightness by the numbers 1, 2, and 3, but in terms of configuration by A, B, and C where B and C are the two images that merge when the transition from three images to one image is made. In the opposed geometry, the correspondence between the two is that $A = 3$, $B = 2$, and $C = 1$, but in the allied geometry, the correspondence is $A = 1$, $B = 3$ and $C = 2$. Once we are in the three-image allied configuration, the brightest images merge according to the M^{-1} law, and the second brightest image is nearly as bright as the brightest image. The image statistics are again dominated by the properties of the fold catastrophe.

8. Non-Aligned Lenses at Differing Redshifts

Virtually all of the analytic results cease to be applicable for the non-aligned lenses at different redshifts. Moreover the range of caustic structures varies greatly, from what are essentially two isolated caustic structures to complex mixtures of higher order catastrophes. Some of the generic behavior of the caustics observed for the single isolated lens is lost. For the isolated strong lens, the transition from five to three images is always a tangential merger at the outer critical line (sometimes a three-image merger due to the cusps, but usually only two images), and the transition from three to one image is always a radial merger on the inner critical line. For the two screen lenses, the transition from seven images to five images or five images to three images can occur in many different ways: radial mergers in either core on the inner critical line, radial mergers on the outer critical line, and tangential mergers on the outer critical line. Even sources starting in the same region of the source plane can undergo qualitatively different mergers depending on the caustic crossed in leaving the region.

We can estimate the regions over which the two lenses interact significantly by expanding the arguments used in §6 and §7. The simplest case is to restrict the discussion to the singular isothermal sphere, in which case the deflection equations can be solved for along the axis of separation. If we restrict ourselves to the region in which the background lens is not multiply-imaged by the foreground lens ($l > r_{12}$) and define an effective separation vector $l' = l + r_{12}$, the lensing equations along the separation vector become

$$
\begin{aligned}
\theta_S &= \theta_1 + |r_{1S} + r_{2S}| & \theta_1 &< -l'/2 \\
&= \theta_1 - |r_{1S} - r_{2S}| & -l'/2 < \theta_1 &< l'/2 \\
&= \theta_1 - |r_{1S} + r_{2S}| & l'/2 < \theta_1 &
\end{aligned}
\tag{8.1}
$$

where we are assuming that $r_{1S} > r_{2S}$. The caustics of the two lenses merge when $l' < 2|r_{1S} + r_{2S}|$, or

$$
l < 2|r_{1S} + r_{2S}| - r_{12}
\tag{8.2}
$$

In the limits that the lens moves to the same plane ($r_{12} \to 0$ and $r_{1S} \to r_{2S}$) or that the second lens moves to the source redshift ($r_{2S} \to 0$ and $r_{12} \to r_{1S}$) equation (8.2) produces the expected results. However, in the limit that the foreground galaxy approaches the observer, the result diverges in angular coordinates because the singular isothermal sphere is always supercritical and hence able to generate multiple images. On the other hand, there is no consistent way of correcting the equations that will be valid for both strong and marginal lenses. The curves showing the approximate positions of the cross section minimum and the point where the isolated caustics begin to overlap as a function of redshift are shown in Figure 4.

We can also generalize the equation (6.2) to estimate the angular separation at which the induced ellipticities become important. We want to measure the ellipticity induced

by the lens which is being held fixed (because it will be responsible for most of the cross section of the compound lens as the cross section peak for the isolated lenses is near $z = 0.5$). The only required changes are $l \rightarrow l'$, and that we weight the results by the ratio of the lens strengths, so that the induced ellipticity in lens 1 by lens 2 is approximately

$$\epsilon \simeq \left[\frac{r_{2S}}{r_{1S}} \left(\frac{2r_{1S}}{l'} \right)^2 \frac{s}{l'} \right] \tag{8.3}$$

The curve showing the separation at which there is a significant elliptical perturbation in the fixed lens is shown in Figure 4.

As a case in point for this section, consider the case of two marginal lenses separated by slightly less than $r_{+,0}$, but with one lens at redshift 0.5, and the other lens at redshift 0.9. We show the caustics, critical lines, and image multiplicity diagram of this lens in Figure 21. This lens shows a higher degree of complexity than the other two cases we have examined in detail. The lens has several crossed fold catastrophes where transitions can be made in which the number of images changes by four rather than two. The transition from five images to three images can be made in two ways depending on whether the transition is made across the cusped inner caustic (shown in Figure 21) or across the outer elliptical caustic. The elliptical caustic is associated with radial mergers in the cores of the galaxy rather than tangential mergers on the outer critical line. Despite all of this complexity, the lens has many of the same generic properties of the previous cases. For this case, we shall examine the five-image configurations in detail. This is one of the rare cases in which the seven-image cross section does not eventually dominate the total cross section at high amplifications, (although it may at higher amplifications than we studied). The cross sections are shown in Figure 22.

The structure of the five-image separations and amplifications is complicated at low amplifications both because of the different regions associated with five images, and because of changes such as the allied/opposed geometry change for three images discussed in the last section. The important fact is that at high amplifications everything can again be characterized in terms of the properties of fold catastrophes, as is seen in Figure 22a. There are some signs of the cusp catastrophes in the behavior of the fainter images, but they have vanished when the amplification of the brightest image is greater than 10. This is simply due to the differential cross section as a function of amplification near cusps and any other higher order catastrophe decreasing more steeply than M^{-3}, which characterizes the fold catastrophe.

9. The Effects of Asymmetries in the Potentials

All of the results shown so far have used two identical spherical potentials. While this assumption is necessary to limit the size of the parameter space to be explored, we will show in this section that the results are essentially generic except in the extremely high symmetry configurations of §7, but that in this case the changes that breaking the symmetry introduces are entirely predictable. The first step is to modify the potential introduced in §3 to make it an elliptical potential. We do this simply by introducing an ellipticity into the equation for the potential (3.1), where if ϵ is the ellipticity,

$$\begin{aligned} \phi &= A \left(1 + (1 - \epsilon)x_1^2/s^2 + (1 + \epsilon)y_1^2/s^2 \right)^{1/2} \\ x_1 &= \ x \cos\theta + y \sin\theta \\ y_1 &= -x \sin\theta + y \cos\theta \end{aligned} \tag{9.1}$$

where, like the circular potential, $A = 2\pi s\sigma^2$ and x_1 and y_1 define the principal axis of the potential. This potential has positive surface densities for all values of ϵ. The form of this potential is somewhat unrealistic because the potential contours remain elliptical at large radii, unlike real galactic potentials, which become circular at large radii. However, we do not wish to introduce extra parameters beyond those needed to break the symmetry of the circular potential.

We examined several cases in which both lenses have ellipticities $\epsilon = 0.1$ and position axis x' at angles of 25 and 115 degrees with respect to the separation vector. We did not vary the strengths and core sizes of the potentials. The caustic structures of the lenses, which are not highly symmetric to begin with (i.e., except for the aligned lenses at varying redshifts), are not strongly perturbed by the introduction of ellipticity into the lenses. The axisymmetric lenses behave like the elliptical lenses considered in BK: the introduction of ellipticity leads to the formation of cusped caustics, and increases in the cross sections for higher image multiplicities. The cross section for the aligned marginal lenses does not show as much of a dramatic variation as the isolated marginal lens. This implies that the perturbation of a marginal lens supplied by a lens at a second redshift leads to roughly the same increase in cross section as an elliptical perturbation. As a series of examples of the consequences of introducing ellipticity into the potentials, we show the caustics, critical lines, and multiplicity diagrams for the elliptical counterparts of the special cases discussed in §6, §7, and §8 in Figures 23, 24, and 25. As predicted, the strong lens system shown in Figure 23 is largely uneffected by the introduction of ellipticity because the tidal effects of the two lenses have already broken the circular symmetry of the potentials. The case of the aligned strong lenses shows that in cases of high symmetry, the ellipticity can cause a large change in the lens structure. The inner caustics are nowhere near having circular symmetry, and the five-image region has grown at the expense of the three-image region. The total cross section and the banded structure of the image multiplicity diagram are relatively unchanged. The marginal lens shown in Figure 25 is strongly perturbed from the circular case even though the circular symmetry was already broken. This results from the extreme sensitivity of the marginal lens to enhancements in surface density as seen in the dependence of the cross section on the strength (6.1). On the other hand the five-image geometry associated with the same piece of the fold used in Figure 21 remains largely unchanged.

10. Conclusions

(1) The two-screen lens, like the isolated lens, is dominated by the behavior of images near fold caustics. In all cases, the integral cross section scales with the magnification of the brightest images as M^{-2} (at sufficiently large magnifications), which is the characteristic scaling of the fold catastrophe. This effect is also shown by the fact that the brightest two images are generally images that are merging at a caustic with the separation between the two brightest images decreasing with amplification, $\propto M^{-1}$. The amplification of the second brightest image typically has nearly the same degree of amplification as the brightest image. Cusp catastrophes are common in the overall caustic structure of the lenses usually as part of a caustic line containing a total of four or six cusps. Cusps do not, however, seem to contribute to the statistical properties of the lenses primarily because their space filling properties are of a lower dimension than those of folds (they are points instead of lines). A common caustic structure is that of two crossed fold catastrophes where transitions can be made in which four images are destroyed along two different critical lines – these are not,

however, examples of higher order catastrophes.

(2) Aside from the two brightest images, the other images typically maintain a nearly fixed pattern relative to the brightest image. This is seen in the statistical results from the fact that the distance of these images relative to the brightest image remains nearly fixed, and it is seen in the banded structure of the image multiplicity diagrams. As in the single screen results, the cores of the galaxies will generally capture some of the images. These are then strongly deamplified by the strong lenses, or somewhat amplified in the marginal lenses. Typically for three-image and five-image geometries, one image is captured in the core of one of the galaxies. For the seven-image case, typically one image is captured in each of the galaxy cores.

(3) There is a cross section peak relative to two isolated lenses of the same type when the lenses are aligned and at the same redshift. The magnitude of the increase strongly depends on the nature of the isolated lenses. The gain is two for singular lenses. As the lenses become marginal, the gain increases, and it will of course diverge if two aligned sub-critical lenses act together to generate multiple images. The area over which the peak extends is approximately $r_{+,0}$ in separation, and $\Delta z \simeq 0.3$ in redshift, on both sides of the fixed lens. The most "efficient" use of any mass for the purpose of lensing is always to combine the mass into one lens at intermediate redshifts. This conclusion agrees with that of Padmanabhan and Subramanian (1988).

(4) In addition to constructive interference, the lenses can interfere destructively. This occurs in two ways. The first is associated with the lenses being separated by approximately $2r_{+,0}$, in which case for each lens the field of the second lens deflects the rays associated with the images of the first lens so that the source regions overlap. In this case, the cross section is essentially the same as that of a single lens, but the multiplicity of the images is greatly increased. Associated with the minimum in the cross section is a maximum for the branching ratios for five and seven images. The second form of destructive interference is associated with over-focusing the rays as seen in the cross section minimum for the aligned lenses at different redshifts. While the most "efficient" use of any mass for lensing is to put it all at one redshift, the "Grand Designer" must be sure to build the lens correctly or the cross section will be less than that of two isolated lenses.

(5) Regions of destructive interference in the total cross section tend to be regions of *constructive* interference in the cross section for highly amplified images. This is a simple consequence of the properties of the fold caustic. The highly amplified systems also tend to be the systems with high image multiplicities, so that amplification bias is equivalent to multiplicity bias. Two-screen lenses, like elliptical lenses, provide information on the presence or absence of amplification bias in the quasar population from the observed branching ratios for higher image multiplicities in multiply-imaged systems.

(6) Image separations can be slightly greater than $2r_{+,0}$, which characterized the maximum image separations of the isolated lens. Systems with separations greater than $2r_{+,0}$, however, are typically high image multiplicity systems with $l \lesssim 3r_{+,0}$. While the maximum image separations can be approximately $4r_{+,0}$, this tends to describe the most widely separated pair of images in a more complicated system. The distance from any one image to its nearest neighbor is still of order $r_{+,0}$. This suggests that large image separations are unlikely to be caused by chaining many weak lenses together.

(7) Marginal lenses have an enormous gain in cross section over large regions of redshift

space and separations. Note that the length scales of the separations over which the interaction matters are governed by the characteristic deflection of the potential $r_{+,0}$, which is associated with the overall velocity dispersion of the potential, rather than length scale of the cross section u_-, which is very small for the marginal lenses. The image separations in such a compound lens are also $\sim r_{+,0}$ so that chaining marginal lenses leads to a larger number of images, not a larger separation of images. Hence if there is a much higher number density of non-luminous potential wells with large core radii at earlier epochs, it may be possible to chain them together to get "dark matter" lenses, as was suggested by Subramanian, Rees, and Chitre (1988). However, such a lens is still a marginal lens, and therefore the "third" image is not deamplified in the core of the galaxy, but it is instead amplified and ought to be visible. This proposal cannot be used to explain the existing "dark matter" lenses in which only two images are seen. The best explanation for $2345 + 007$, $1635 + 267$, and $0023 + 171$ still seems to be that they are random correlations, rather than gravitational lenses. (If they are lenses, a far more likely solution is lensing by a galaxy and a distant cluster, which are too faint to be seen by current telescopes.)

(8) The overall probability of two-screen lenses is low – approximately one in ten to one in fifty lens systems should be systems with more than one lens. However, this neglects the selection effects coming from the changes in the probability of finding lenses. Note that while there are significant destructive interference effects in the total cross section for multiple imaging, the cross section for strongly amplified images is almost always increased. For marginal lenses the destructive interference effects are small even in the total cross section. Hence while the probability of alignment may be of order 0.01 to 0.1, the cross section may be increased by orders of magnitude, so that the change in optical depth can be significant even if the chances of alignment are small. A reasonable estimate is that the optical depth for two-screen lenses is one tenth of the total optical depth for multiple imaging and possibly greater. Note that for more massive galaxies, the size of the interaction region grows $\propto \sigma^2$, so that for a giant elliptical the increase in cross section easily dominates the drop in abundance relative to the smaller galaxies.

References

Apostolakis, J.A., and Kochanek, C.S., 1988, in *The Third Conference on Hypercube Concurrent Computers and Applications*, (New York: Assoc. for Computing Machinery), p.963.

Blandford, R.D., and Narayan, R., 1986, *Astrophys. J.*, **310**, 568.

Blandford, R.D., and Kochanek, C.S., 1987, *Astrophys. J.*, **321**, 658.

Bourassa, R.R., and Kantowski, R., 1975, *Astrophys. J.*, **195**, 13.

Djorgovski, S., and Spinrad, H., 1984, *Astrophys. J. (Letters)*, **282**, L1.

Hewitt, J.N., Turner, E.L., Lawrence, C.R., Schneider, D.P., Gunn, J.E., Bennett, C.L., Burke, B.F., Mahoney, J.H., Langston, G.I., Schmidt, M., Oke, J.B., and Hoessel, J.G., 1987, *Astrophys. J.*, **321**, 706.

Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., and Perley, R., 1985, *Astr. J.*, **90**, 691.

Kochanek, C.S., and Blandford, R.D., 1987, *Astrophys. J.*, **321**, 676.

Kovner, I., 1987, *Astrophys. J.*, **316**, 52.

Narasimha, D., Subramanian, K., and Chitre, S.M., 1986, *Astrophys. J.*, **315**, 434.

Padmanabhan, T., and Subramanian, K., 1988, *Mon. Not. R. astr. Soc.*, **233**, 265.

Phinney, E.S., and Blandford, R.D., 1986, *Nature*, **321**, 569.

Schneider, D.P., Gunn, J.E., Turner, E.L., Lawrence, C.R., Hewitt, J.N., Schmidt, M., and Burke, B.F., 1986, *Astr. J.*, **91**, 991.

Schneider, P., and Borgeest, U., 1986, *Max Planck Institude, MPA 237 preprint*.

Schneider, P., and Weiss, A., 1986, *Astr. Astrophys.*, **164**, 237.

Schneider, P., and Weiss, A., 1987, *Max Planck Institude, MPA 311 preprint*.

Subramanian, K., Rees, M.J., and Chitre, S.M., 1987, *Mon. Not. R. astr. Soc.*, **224**, 283.

Weedman, D.W., Weymann, R.J., Green, R.F., and Heckman, T.M., 1982, *Astrophys. J. (Letters)*, **255**, L5.
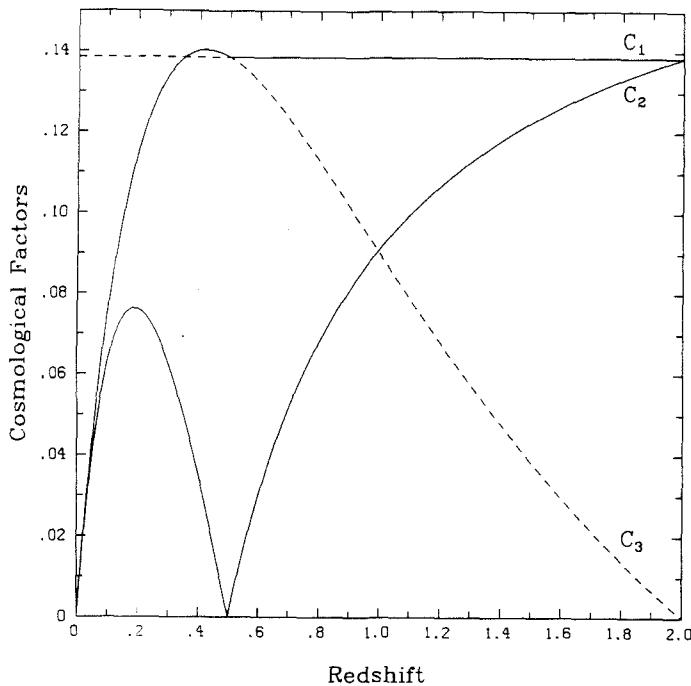
Figure 1. The cosmological weighting factors appearing in the two-screen lensing equation are shown for an Einstein-DeSitter universe. One lens is held fixed at a redshift of 0.5, the second lens is allowed to have a varying redshift. The source is fixed at redshift $z = 2.0$. C_1 is $D_{O1}D_{1S}/r_H D_{OS}$, which measures the ability of the foreground galaxy to lens the source. C_2 is $D_{O2}D_{2S}/r_H D_{OS}$, which measures the ability of the background galaxy to lens the source. C_3 is $D_{O1}D_{12}/r_H D_{O2}$, which measures the ability of the foreground galaxy to lens the background galaxy. The Hubble radius $r_H = c/H_0$ is inserted to make the quantity dimensionless.

(a)

(b)
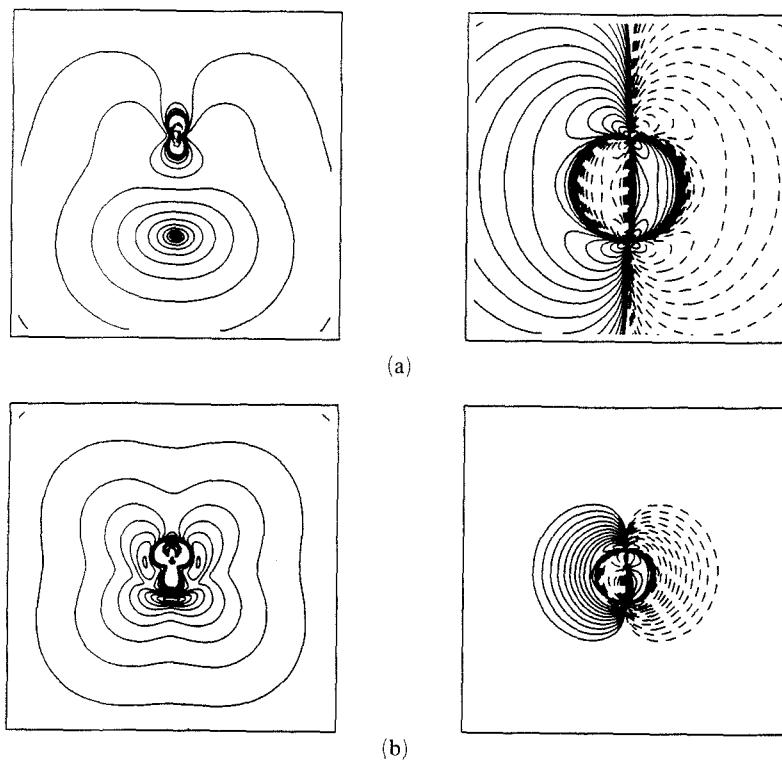
Figure 2. Contour plots of the effective surface density (right) and vector potential source (left) for two different lenses. In case (a). the background lens is not multiply-imaged by the foreground lens. The position of the background lens is at the lower center, and of the foreground lens at the upper center. In case (b) the background lens is multiply-imaged by the foreground lens leading to three "images" of the core in the surface density. Solid lines represent positive contours, and dashed lines are negative contours. Note the negative effective surface density at the position of the core of the foreground galaxy. Both cases show two strong lenses, for case (a) $z_1 = 0.5$, $z_2 = 0.8$, and $l = 2r_{+,0}$, and for case (b) $z_1 = 0.5$, $z_2 = 1.0$, and $l = r_{+,0}/4$.
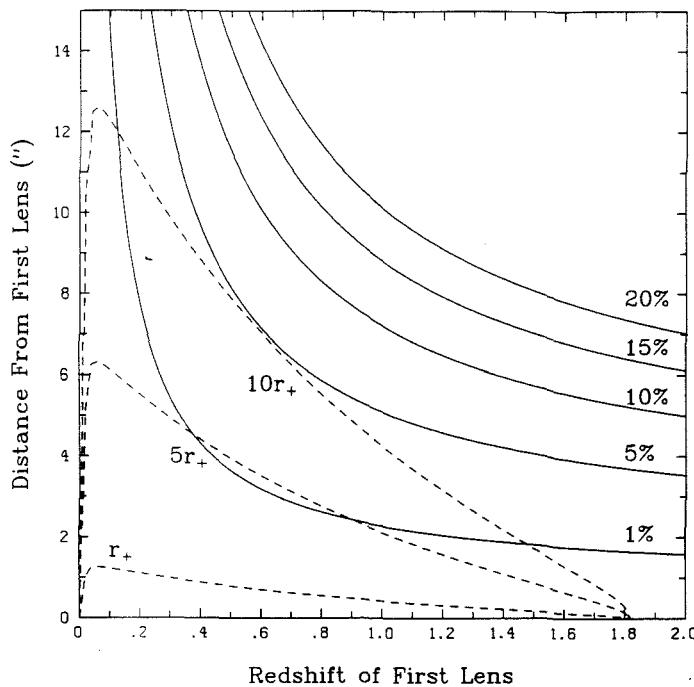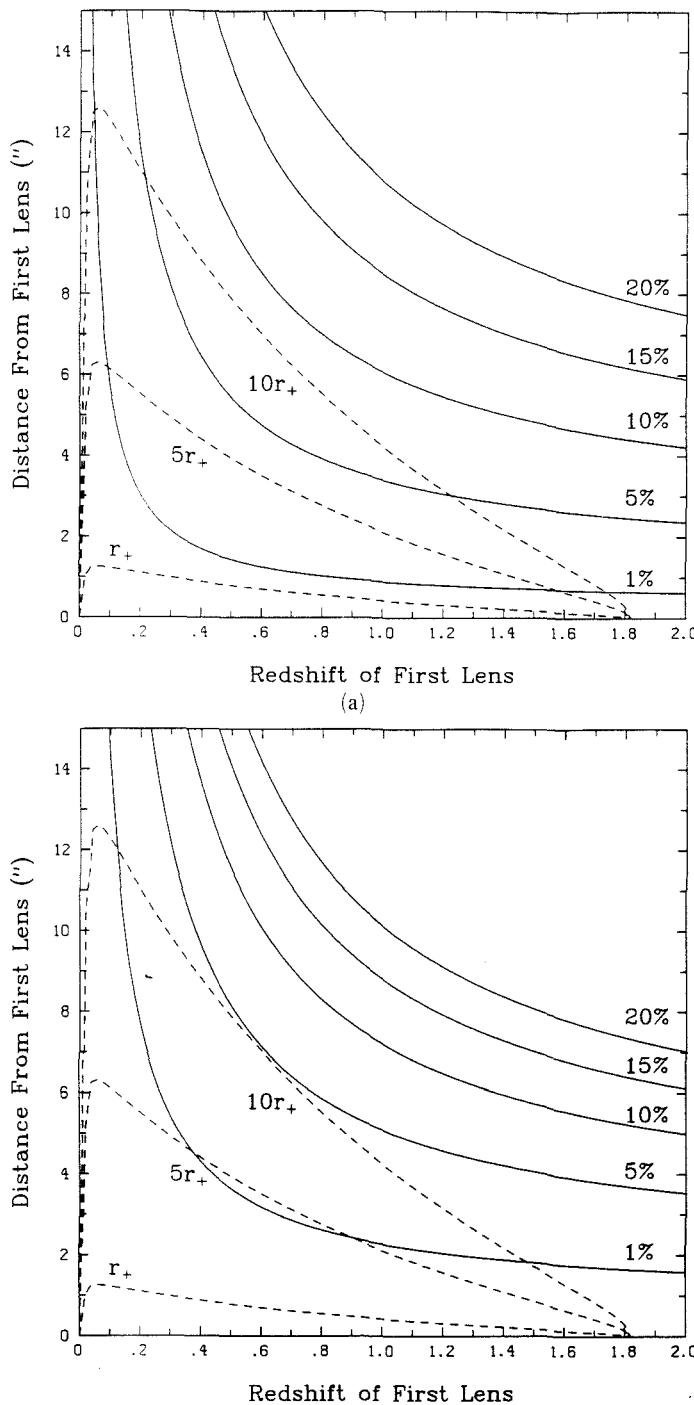
Figure 3. (a). The probability contours (solid lines) as a function of angular separation and redshift for finding a second galaxy at the same redshift within a given angular distance from a galaxy. (b). The probability contours (solid lines) as a function of angular separation and redshift for finding a second galaxy at a different redshift within a given angular distance from a galaxy. In both cases, the dashed lines are a multiple of the radius of the outer critical line r_+ for the intrinsically strong lens.
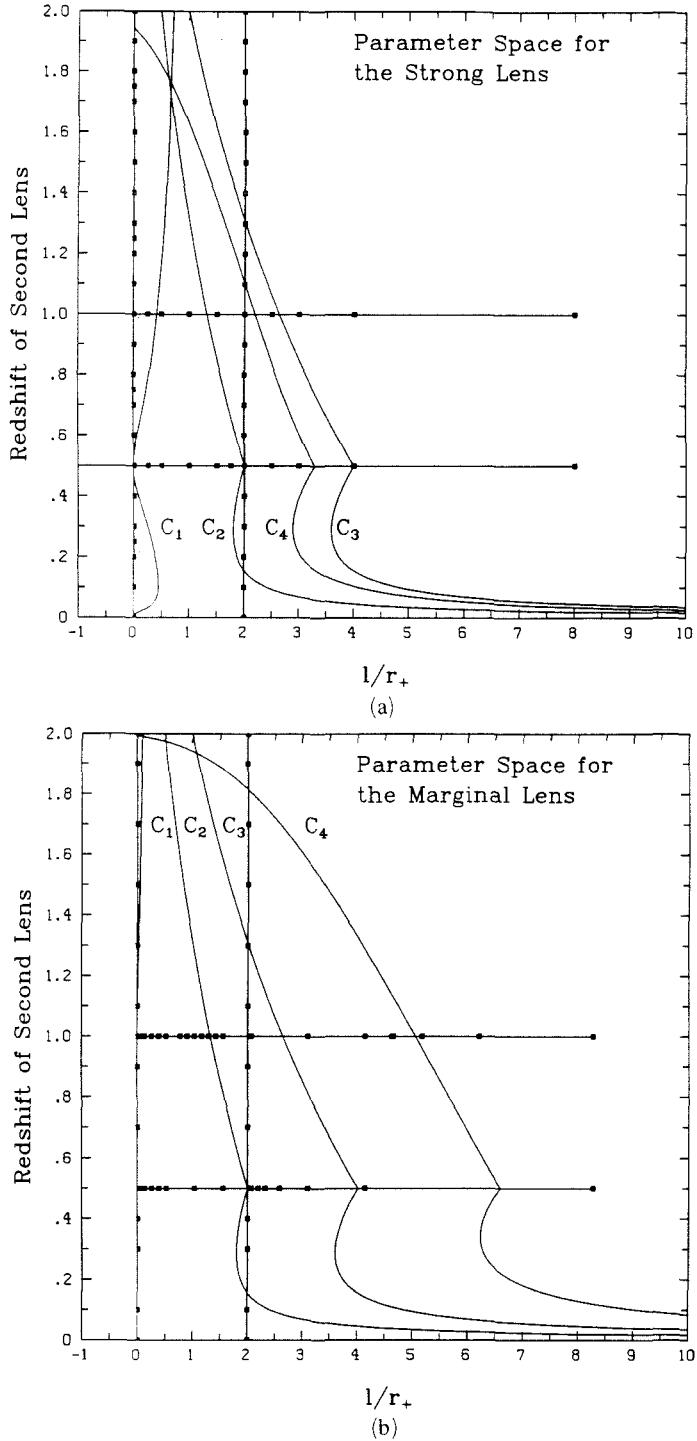
Figure 4. The parameter space surveyed for the strong (a) and marginal (b) lenses. Curve C_1 shows the separation within which the background lens is multiply-imaged by the foreground galaxy. Curve C_2 shows the approximate separation at the cross section minimum which occurs when the caustics of the two lenses completely overlap. Curve C_3 shows the approximate separation at which the lenses cease to have overlapping caustics. Curve C_4 shows the separation at which the induced ellipticity becomes significant.
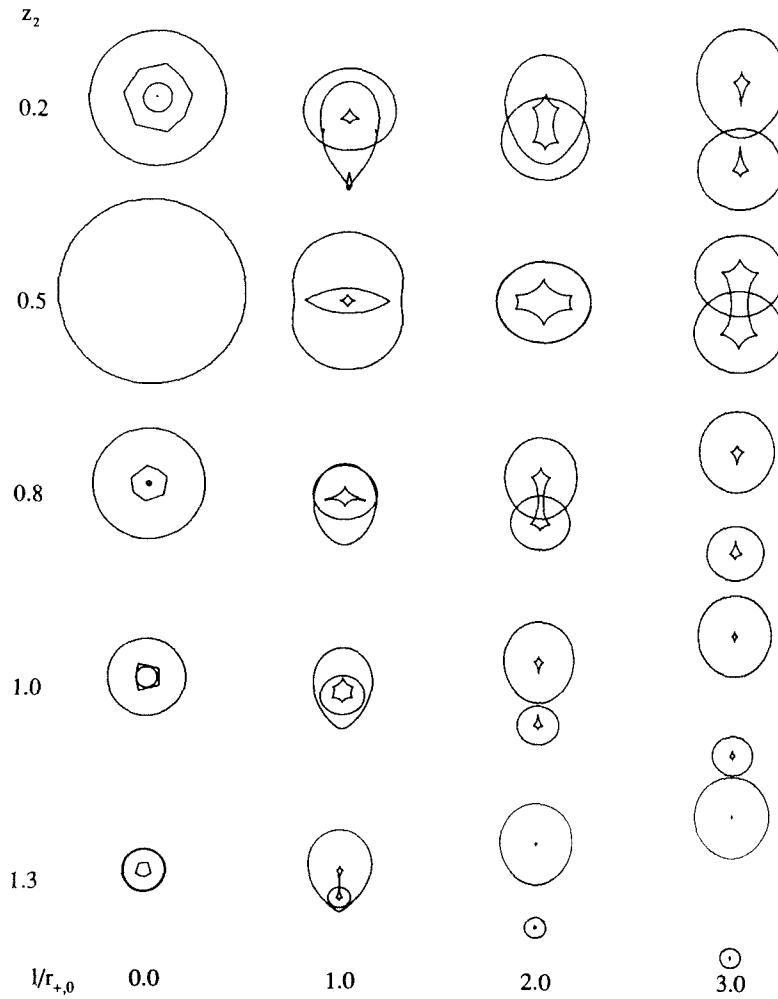
Figure 5. The behavior of the caustics as a function of redshift and separation for the strong lenses. One lens is held fixed at $z = 0.5$. The position and redshift of the second lens is varied. The caustics are shown for the second lens at redshifts of $z = 0.2$, 0.5, 0.8, 1.0, and 1.3 and at separations of $l = 0$, $l = r_{+,0}$, $l = 2r_{+,0}$, and $l = 3r_{+,0}$. The caustics consist of folds (the lines) and cusps (the cusps in the lines). The aligned series ($l/r_{+,0} = 0$) is strongly sensitive to small perturbations leading to the polygonally shaped caustics–these represent a symmetry problem at the low resolutions used to generate the contour diagrams (see §5).
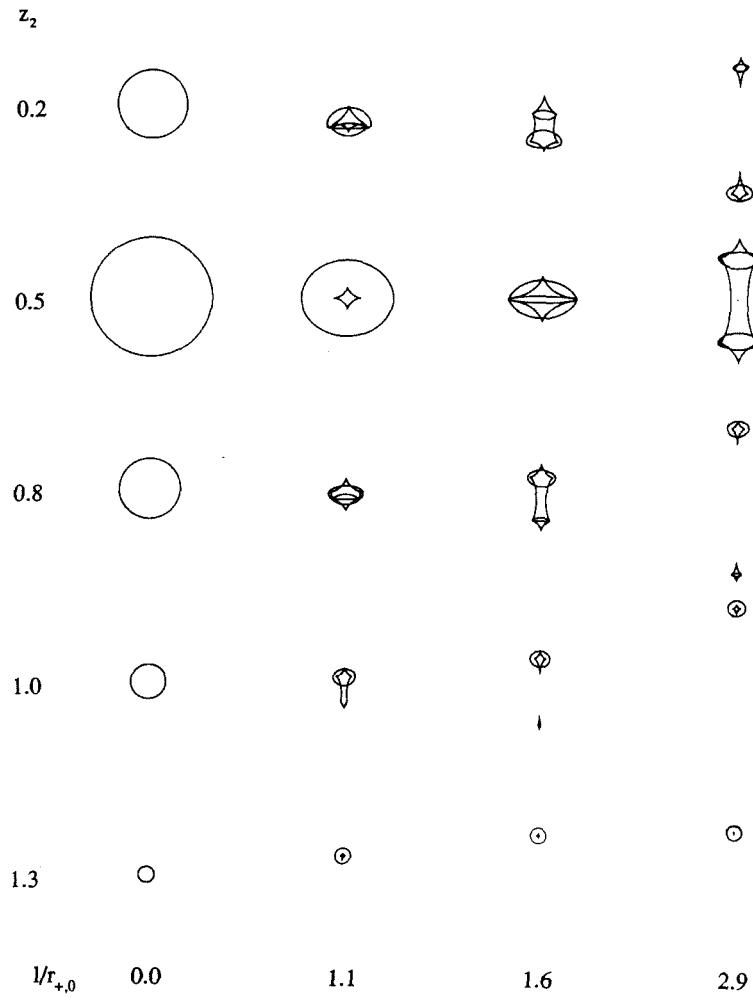
Figure 6. The behavior of the caustics as a function of redshift and separation for the marginal lenses. One lens is held fixed at $z = 0.5$. The position and redshift of the second lens is varied. The caustics are shown for the second lens at redshifts of $z = 0.2$, 0.5, 0.8, 1.0, and 1.3 and at separations of $l = 0$, $l = 12u_- \simeq 1.1r_{+,0}$, $l = 18u_- \simeq 1.6r_{+,0}$, and $l = 32u_- \simeq 2.9r_{+,0}$. The caustics consist of folds (the lines) and cusps (the cusps in the lines).
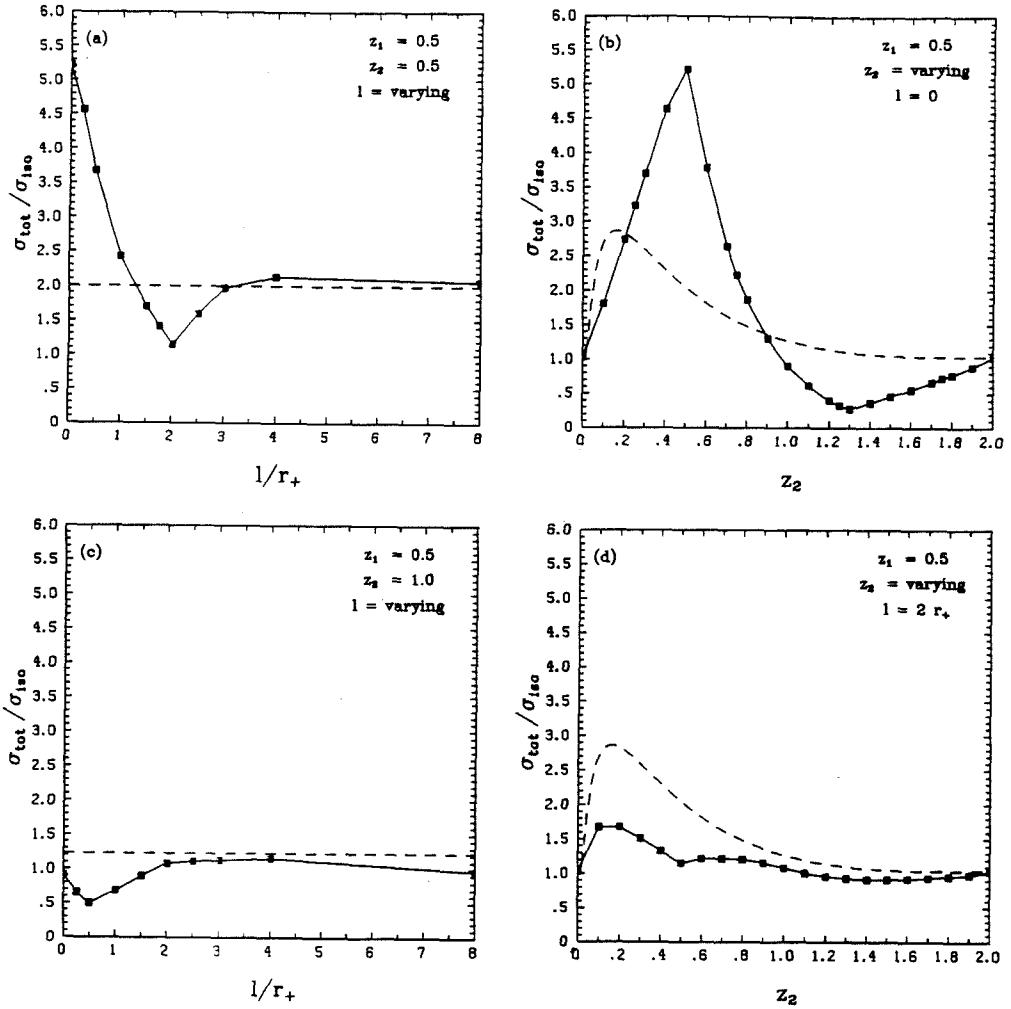
Figure 7. The total multiple imaging cross sections for the strong lens are shown for four cuts through the redshift-separation parameter space. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied. The simulations are shown with the solid lines, and the black squares are the data points. The dashed lines are the cross section that would be expected from two completely independent lenses at the two redshifts. If the solid line is above the dashed line the two-screen lens is interfering constructively, giving a greater cross section than the isolated lenses, and if the solid line is below the dashed line, the two-screen lens is interfering destructively. All cross sections are measured in units of the cross section of an isolated lens at a redshift of $z = 0.5$.

**Figure 8.** The total multiple imaging cross sections for the marginal lens are shown for four cuts through the redshift-separation parameter space. In (**a**) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (**b**) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (**c**) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (**d**) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied. The simulations are shown with the solid lines, and the black squares are the data points. The dashed lines are the cross section that would be expected from two completely independent lenses at the two redshifts. If the solid line is above the dashed line the two-screen lens is interfering constructively, giving a greater cross section than the isolated lenses, and if the solid line is below the dashed line, the two-screen lens is interfering destructively. All cross sections are measured in units of the cross section of an isolated lens at a redshift of $z = 0.5$.

Figure 9. The total branching ratios for the strong lens are shown for four cuts through the redshift-separation parameter space. The branching ratio for n images is the cross section for finding n images divided by the total cross section for multiple imaging. The branching ratio for three images is indicated by the triangles, the branching ratio for five images is indicated by the squares, and the branching ratio for seven images is indicated by the pentagons. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied.

Figure 10. The total branching ratios for the marginal lens are shown for four cuts through the redshift-separation parameter space. The branching ratio for n images is the cross section for finding n images divided by the total cross section for multiple imaging. The branching ratio for three images is indicated by the triangles, the branching ratio for five images is indicated by the squares, and the branching ratio for seven images is indicated by the pentagons. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied.

Figure 11. The multiple imaging cross sections when the brightest image is amplified by more than ten for the strong lens are shown for four cuts through the redshift-separation parameter space. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied. The simulations are shown with the solid lines, and the black squares are the data points. The dashed lines are the cross section that would be expected from two completely independent lenses at the two redshifts. All cross sections are measured in units of the total cross section of an isolated lens at a redshift of $z = 0.5$.

Figure 12. The multiple imaging cross sections when the brightest image is amplified by more than ten for the marginal lens are shown for four cuts through the redshift-separation parameter space. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied. The simulations are shown with the solid lines, and the black squares are the data points. The dashed lines are the cross section that would be expected from two completely independent lenses at the two redshifts. All cross sections are measured in units of the total cross section of an isolated lens at a redshift of $z = 0.5$.

Figure 13. The branching ratios for the strong lens when the brightest image has been amplified by more than a factor of ten are shown for four cuts through the redshift-separation parameter space. The branching ratio for n images is the cross section for finding n images divided by the total cross section for multiple imaging. The branching ratio for three images is indicated by the triangles, the branching ratio for five images is indicated by the squares, and the branching ratio for seven images is indicated by the pentagons. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied.

Figure 14. The branching ratios for the marginal lens when the brightest image has been amplified by more than a factor of ten are shown for four cuts through the redshift-separation parameter space. The branching ratio for n images is the cross section for finding n images divided by the total cross section for multiple imaging. The branching ratio for three images is indicated by the triangles, the branching ratio for five images is indicated by the squares, and the branching ratio for seven images is indicated by the pentagons. In (a) the lenses are both at redshift $z = 0.5$ and the separation is allowed to vary. In (b) the lenses are aligned with separation $l = 0$. One lens is held fixed at $z = 0.5$ and the redshift of the second lens is varied. In (c) one lens is at $z = 0.5$, the other lens is at $z = 1.0$, and the separation is varied. In (d) the separation is held fixed at $l = 2r_{+,0}$, one lens is at $z = 0.5$, and the redshift of the second lens is varied.

Figure 15. A comparison between the simple analytic model derived in the text (6.6) and the numerical results for the cross sections and branching ratios of two lenses at the same redshift as a function of separation. The solid lines are the analytic model, and the dashed lines are the numerical results for the strong lens. In (a) the cross sections are compared, and in (b), the branching ratios are compared.

(A): Source Plane - Caustics

☐ 1 Image
▨ 3 Images
▨ 5 Images
▨ 7 Images
■ The Source

(B): Image Plane - Critical Lines

☐ Positive Amplification
▨ Negative Amplification
■ The Images
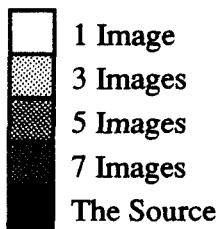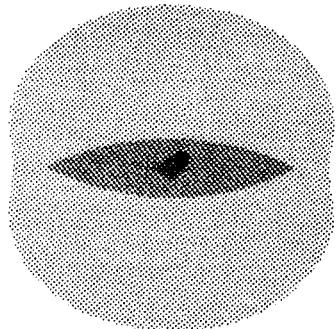
(C): Image Plane - Multiplicities
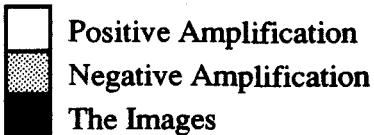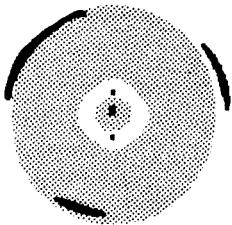
☐ 1 Image
▨ 3 Images
▨ 5 Images
■ 7 Images
■ The Images
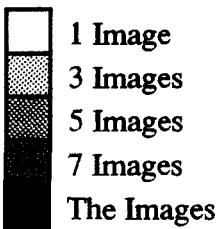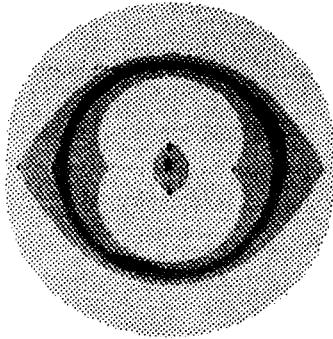
Figure 16. The caustics, critical lines, and image multiplicity diagram for two strong lenses at the same redshift ($z = 0.5$) and a separation of $l = r_{+,0}$ are shown. In (a) we show the cross sections, in (b) the critical lines, and in (c) the image multiplicity diagram. The image multiplicity diagram shows where the images that compose a lens of a given multiplicity can lie: in a five-image lens, all five images must lie in the five image regions of the diagram. Also shown is an example of a characteristic seven-image configuration. The longest arc of the images is really two images that are connected at the critical line.
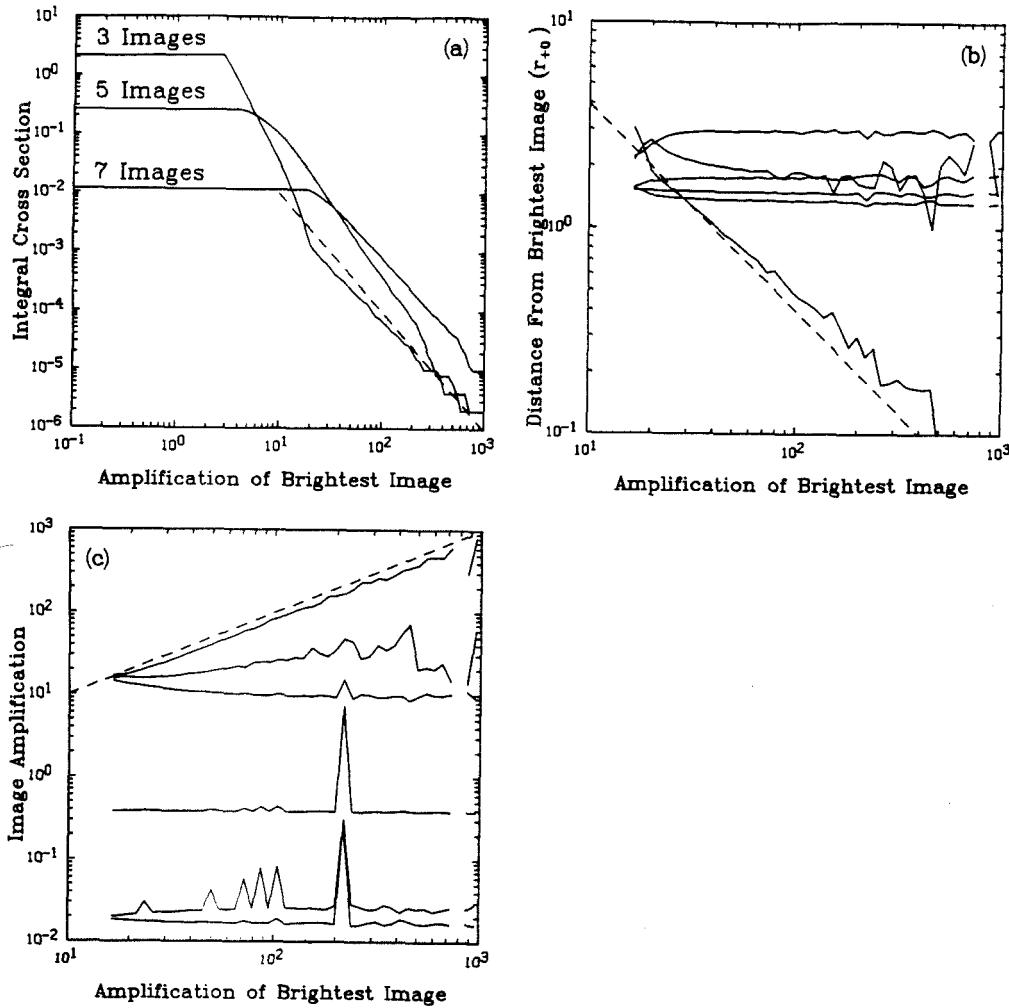
Figure 17. A detailed look at the statistics of the seven-image configurations for two strong lenses at the same redshift ($z = 0.5$) and a separation of $l = r_{+,0}$. In (a) we show the integral cross sections as a function of amplification. Note that at sufficiently high amplification, the high image multiplicities dominate the cross section. The dashed line is to show the M^{-2} dependence of the cross sections at high amplification predicted from the behavior of images near fold catastrophes. In (b) we show the average separation between the brightest image and the subsidiary images as a function of the brightest image's amplification. The second brightest image approaches the brightest image as the amplification increases. The average separation decreases proportional to M^{-1}, which is shown by the dashed line. The images are labeled in order of increasing amplification: image 1 is the faintest, and image 7 is the brightest. In (c) we show the average amplification of the images as a function of the amplification of the brightest image. The second brightest image's average amplification approaches that of the brightest image, as shown by its approach to the dashed line. The two faintest images are captured in the two cores of the galaxies and are strongly deamplified.
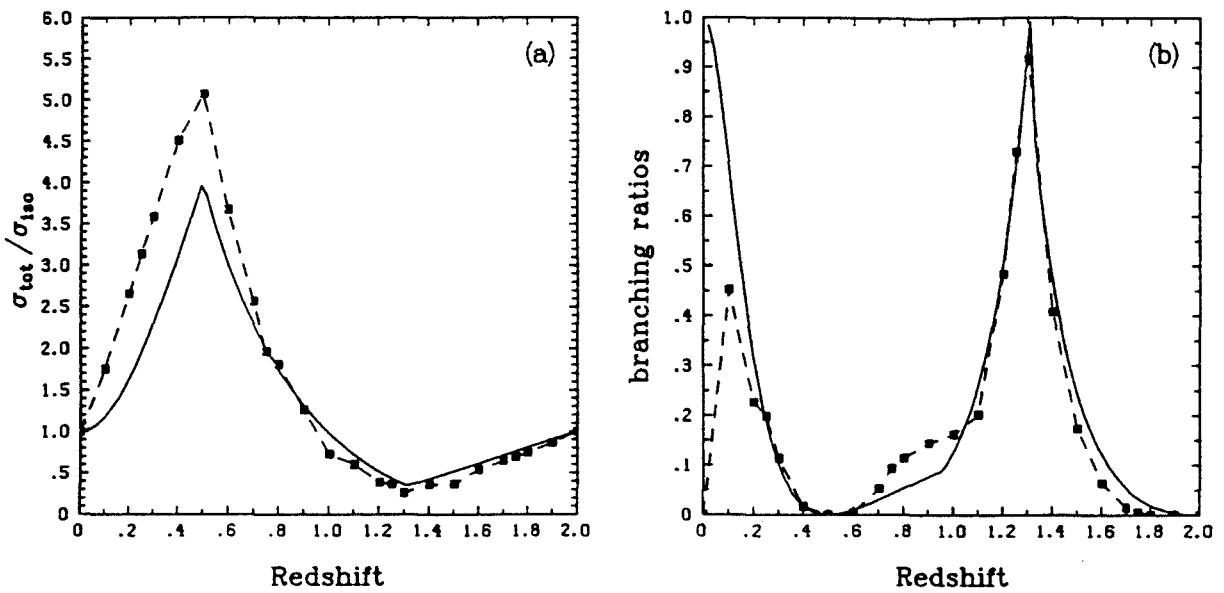
Figure 18. A comparison between the simple analytic model derived in the text (7.4) and the numerical results for the cross sections and branching ratios of two lenses at the same redshift as a function of separation. The solid lines are the analytic model, and the dashed lines are the numerical results for the strong lens. In (a) the cross sections are compared, and in (b), the branching ratios for image multiplicities greater than three are compared.

(A): Source Plane - Caustics

- 1 Image
- 3 Images
- 5 Images
- 7 Images
- The Source

(B): Image Plane - Critical Lines

- Positive Amplification
- Negative Amplification
- The Images
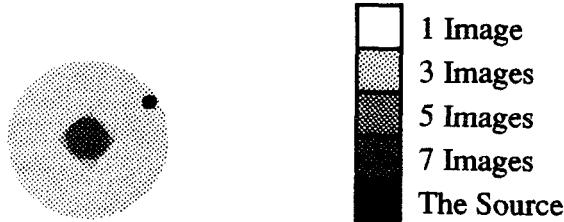
(C): Image Plane - Multiplicities
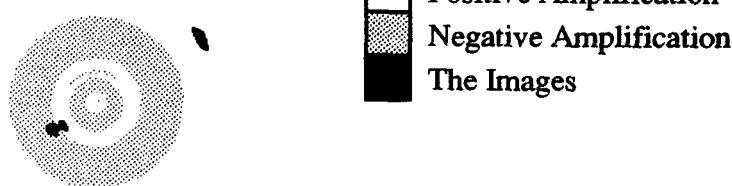
- 1 Image
- 3 Images
- 5 Images
- 7 Images
- The Images
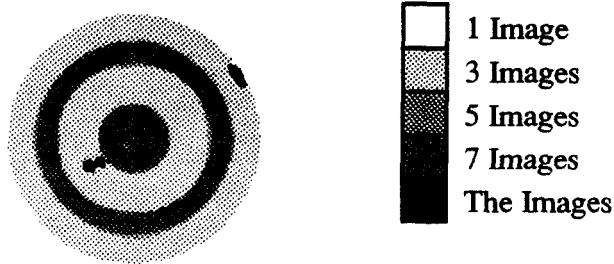
Figure 19. The caustics, critical lines, and image multiplicity diagram for two aligned strong lenses at redshifts of $z_1 = 0.5$, and $z_2 = 1.0$ are shown. In (a) we show the cross sections, in (b) the critical lines, and in (c) the image multiplicity diagram. The image multiplicity diagram shows where the images that compose a lens of a given multiplicity can lie: in a five-image lens, all five images must lie in the five-image regions of the diagram. Also shown is an example of the typical three-image configuration showing the radial "allied" merger of two images.
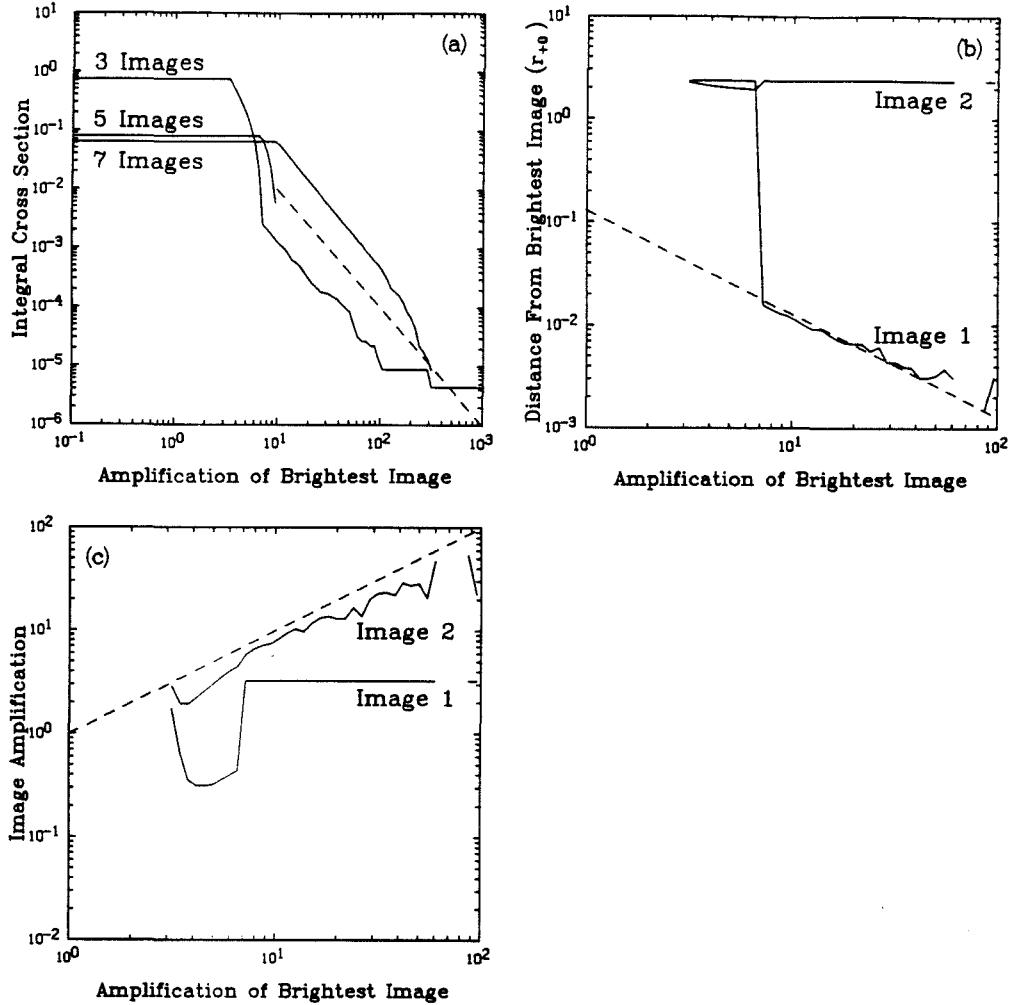
Figure 20. A detailed look at the statistics of the three-image configurations for two aligned strong lenses at redshifts of $z_1 = 0.5$, and $z_2 = 1.0$. In (a) we show the integral cross sections as a function of amplification. Note that at sufficiently high amplification, the high image multiplicities dominate the cross section. The dashed line is to show the M^{-2} dependence of the cross sections at high amplification predicted from the behavior of images near fold catastrophes. In (b) we show the average separation between the brightest image and the subsidiary images as a function of the brightest image's amplification. The second brightest image approaches the brightest image as the amplification increases. The average separation decreases proportional to M^{-1}, which is shown by the dashed line. The images are labeled in order of increasing amplification: image 1 is the faintest, and image 3 is the brightest. In (c) we show the average amplification of the images as a function of the amplification of the brightest image. The second brightest image's average amplification approaches that of the brightest image, as shown by its approach to the dashed line.

(A): Source Plane - Caustics

1 Image
3 Images
5 Images
7 Images
The Source

(B): Image Plane - Critical Lines

Positive Amplification
Negative Amplification
The Images
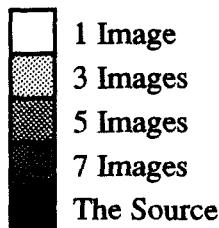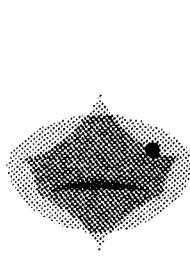
(C): Image Plane - Multiplicities
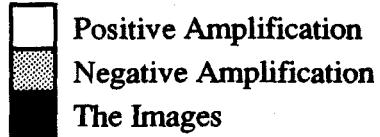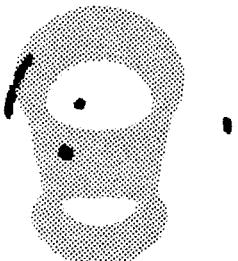
1 Image
3 Images
5 Images
7 Images
The Images
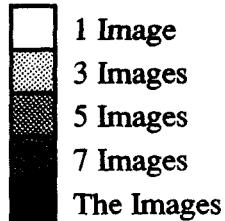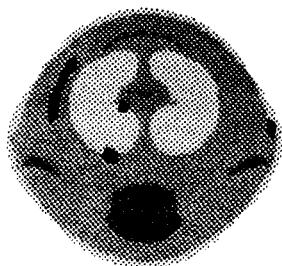
Figure 21. The caustics, critical lines, and image multiplicity diagram for two marginal lenses at redshifts of $z_1 = 0.5$, and $z_2 = 0.9$ and separation $l \simeq r_{+,0}$ are shown. In (a) we show the cross sections, in (b) the critical lines, and in (c) the image multiplicity diagram. The image multiplicity diagram shows where the images that compose a lens of a given multiplicity can lie: in a five-image lens, all five images must lie in the five-image regions of the diagram. Also shown is an example of the typical five-image configuration.
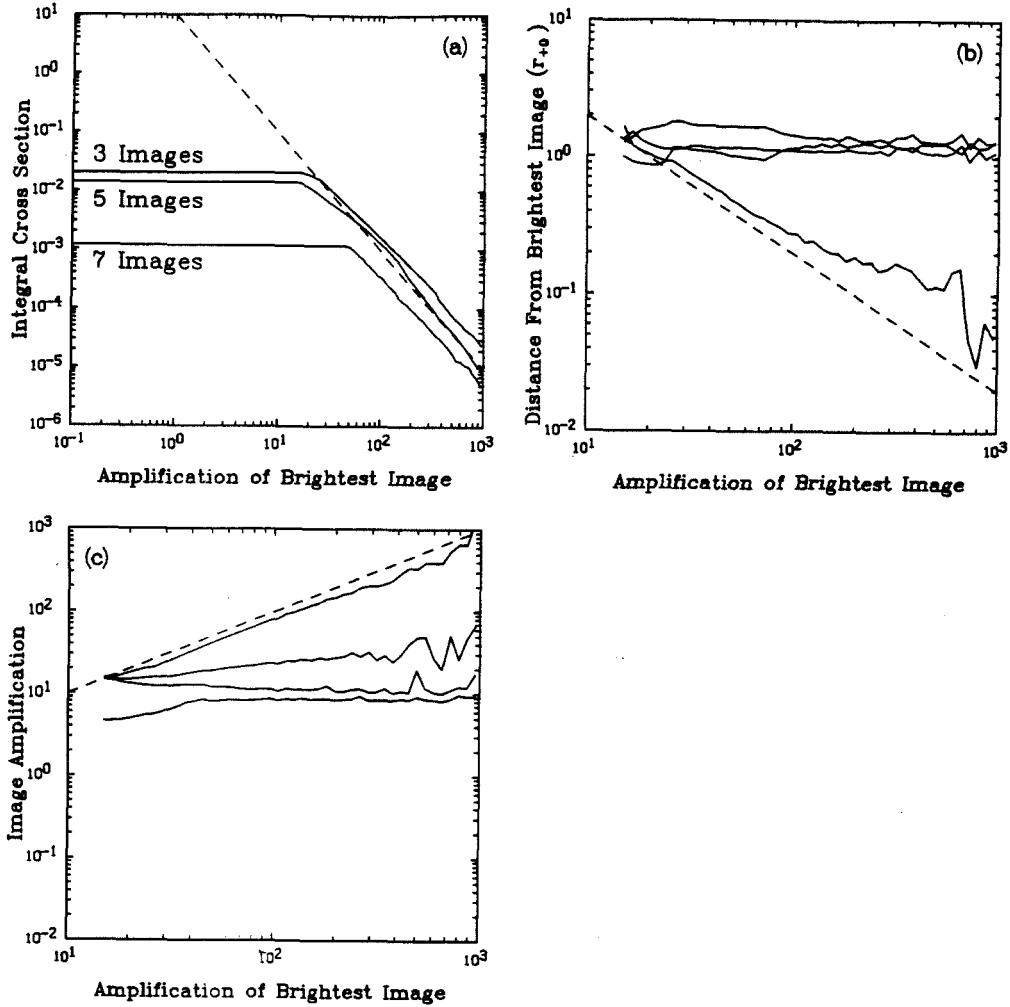
Figure 22. A detailed look at the statistics of the five-image configurations for two marginal lenses at redshifts of $z_1 = 0.5$, and $z_2 = 0.9$ and separation $l \simeq r_{+,0}$. In (a) we show the integral cross sections as a function of amplification. Note that at sufficiently high amplification, the high image multiplicities dominate the cross section. The dashed line is to show the M^{-2} dependence of the cross sections at high amplification predicted from the behavior of images near fold catastrophes. In (b) we show the average separation between the brightest image and the subsidiary images as a function of the brightest image's amplification. The second brightest image approaches the brightest image as the amplification increases. The average separation decreases proportional to M^{-1}, which is shown by the dashed line. The images are labeled in order of increasing amplification: image 1 is the faintest, and image 5 is the brightest. In (c) we show the average amplification of the images as a function of the amplification of the brightest image. The second brightest image's average amplification approaches that of the brightest image, as shown by its approach to the dashed line.

(A): Source Plane - Caustics

- ☐ 1 Image
- ▨ 3 Images
- ▨ 5 Images
- ■ 7 Images
- ■ The Source

(B): Image Plane - Critical Lines

- ☐ Positive Amplification
- ▨ Negative Amplification
- ■ The Images
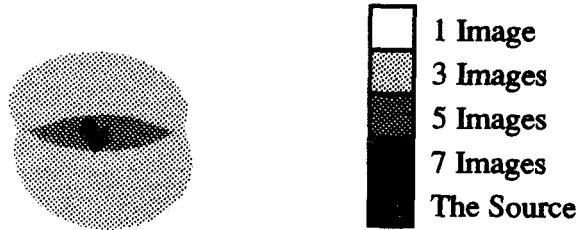
(C): Image Plane - Multiplicities
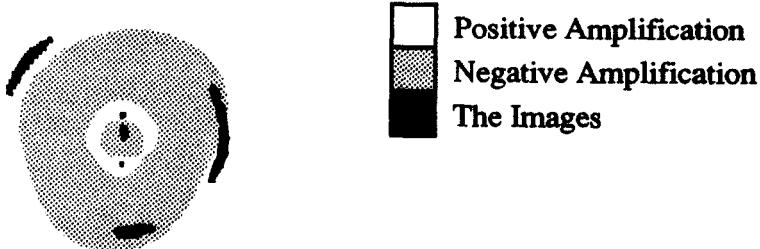
- ☐ 1 Image
- ▨ 3 Images
- ▨ 5 Images
- ■ 7 Images
- ■ The Images
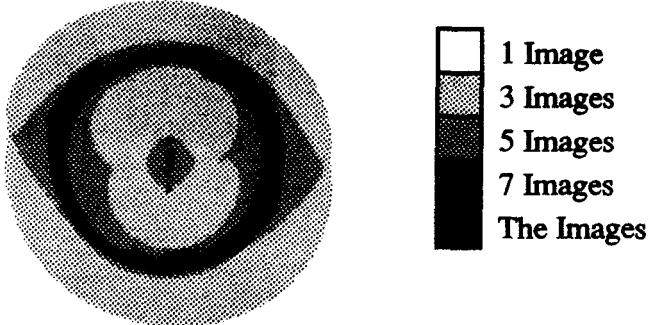
Figure 23. The caustics, critical lines, and image multiplicity diagram for two strong lenses at the same redshift ($z = 0.5$) and a separation of $l = r_{+,0}$ with ellipticities as described in §9. In (a) we show the cross sections, in (b) the critical lines, and in (c) the image multiplicity diagram. The image multiplicity diagram shows where the images that compose a lens of a given multiplicity can lie: in a five-image lens, all five images must lie in the five-image regions of the diagram. Also shown is an example of a characteristic seven-image configuration. The longest arc of the images is really two images that are connected at the critical line.

(A): Source Plane - Caustics

☐ 1 Image
▨ 3 Images
▦ 5 Images
▩ 7 Images
■ The Source

(B): Image Plane - Critical Lines

☐ Positive Amplification
▨ Negative Amplification
■ The Images
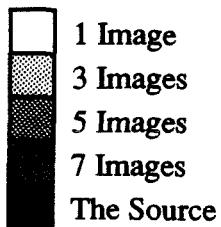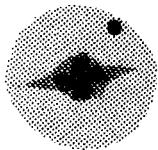
(C): Image Plane - Multiplicities
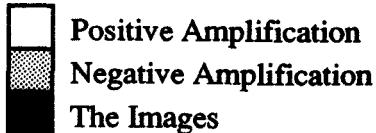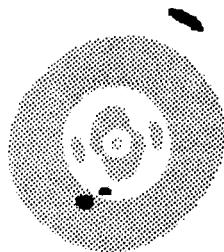
☐ 1 Image
▨ 3 Images
▦ 5 Images
▩ 7 Images
■ The Images
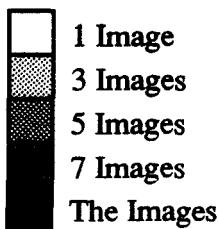
Figure 24. The caustics, critical lines, and image multiplicity diagram for two aligned strong lenses at redshifts of $z_1 = 0.5$, and $z_2 = 1.0$ with ellipticities as described in §9. In (a) we show the cross sections, in (b) the critical lines, and in (c) the image multiplicity diagram. The image multiplicity diagram shows where the images that compose a lens of a given multiplicity can lie: in a five-image lens, all five images must lie in the five-image regions of the diagram. Also shown is an example of the typical three-image configuration. (The narrow horizontal stripes are noise due to the low resolution: see §5.)

(A): Source Plane - Caustics

☐ 1 Image
▨ 3 Images
▨ 5 Images
■ 7 Images
■ The Source

(B): Image Plane - Critical Lines

☐ Positive Amplification
▨ Negative Amplification
■ The Images
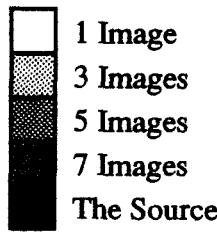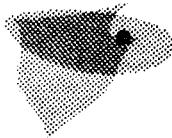
(C): Image Plane - Multiplicities
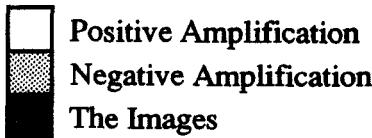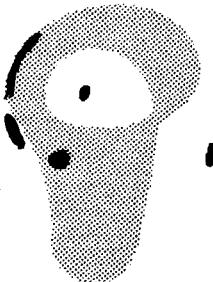
☐ 1 Image
▨ 3 Images
▨ 5 Images
■ 7 Images
■ The Images
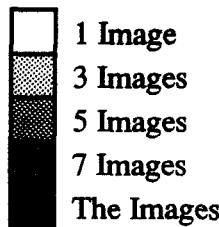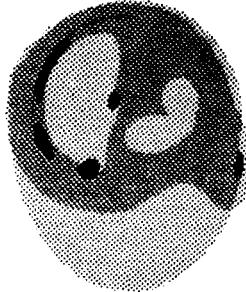
Figure 25. The caustics, critical lines, and image multiplicity diagram for two marginal lenses at redshifts of $z_1 = 0.5$, and $z_2 = 0.9$ and separation $l \simeq r_{+,0}$ are shown. In (a) we show the cross sections, in (b) the critical lines, and in (c) the image multiplicity diagram. The image multiplicity diagram shows where the images that compose a lens of a given multiplicity can lie: in a five-image lens, all five images must lie in the five-image regions of the diagram. Also shown is an example of the typical five-image configuration.

Chapter 5

The Ring Cycle:

An Iterative Lens Reconstruction Technique

Applied to MG1131+0456

Christopher S. Kochanek,
Roger D. Blandford,

Theoretical Astrophysics 130-33,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

Charles R. Lawrence,

Radio Astronomy 105-24,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

and Ramesh Narayan

Steward Observatory,
University of Arizona,
Tucson, AZ, 85721, U.S.A.

Abstract

We describe a new technique for the analysis of well-resolved gravitational lens images. This method allows us to solve for the brightness distribution of the unlensed source as well as a parametrized model of the lens. Our algorithm computes a figure of merit for a lens model based on the scatter in the surface brightnesses of image elements that, according to the model, come from the same source element. Minimization of the figure of merit leads to an optimum solution for the source and the lens. We present a successful application of the method to VLA maps of the "Einstein-ring" radio source MG1131+0456 observed by Hewitt *et al.* (1988). The inversion gives a normal galaxy-like elliptical potential for the lens and an ordinary double-lobed structure for the background radio source. The success of this procedure provides strong support for Hewitt *et al.*'s conclusion that MG1131+0456 is a gravitational lens.

1. Introduction

The discovery of three arcs in rich clusters of galaxies (Soucail *et al.* 1987, Lynds and Petrosian, 1987, Lavery and Henry, 1988) and an elliptical ring in a radio source MG1131+0456 (Hewitt *et al.* 1988), and the growing consensus that all four features are due to gravitational lensing (*e.g.*, , Paczyński 1987), has stimulated interest in the lensing of extended sources.

The qualitative features of the lensing configurations in these sources are now understood. It is well known that a source on the optic axis of a strictly circular lens creates a circular image on the "Einstein-ring." When non-circular perturbations are present (however small), the optic axis unfolds into a caustic surface representing the locus of infinitely amplified source points. Arc images with very large elongations are obtained when extended sources are located close to cusps in this caustic sheet (Grossman and Narayan, 1988; Kovner, 1988). A simple theory of geometrical optics imaging has been developed that describes the extent and asymmetry of such arcs (Blandford and Kovner, 1988). Moreover, simulations using models of clusters and background galaxies have been shown to produce images that are strikingly similar to the optical arcs (Narasimha and Chitre, 1988; Grossman and Narayan, 1988).

In this paper, we develop and test a more detailed and quantitative method of analysis of resolved gravitational images. The motivation for this study is the expectation that arc and ring images of extended sources will furnish far more information about the shape and extent of lensing potentials than images of unresolved sources such as quasars, which probe the projected lensing potential only at isolated points. Each image point provides just one datum, viz. the relative magnification. Even if the source is resolved, as is the case with VLBI observations of 0957+561 (Gorenstein *et al.*, 1984), one still has only a small number of data points to constrain a lens model.

The situation is very much improved when we consider extended sources because here one obtains a *two-dimensional distribution* of image brightness. Although the source has an unknown two-dimensional brightness distribution, parts of it are *multiply imaged*. The image thus has a great deal of redundancy, which can provide a powerful set of constraints on lens models. Furthermore, images of extended sources may actually surround the lens (*e.g.*, MG1131+0456), thus providing a more extensive probe of the lens potential than is possible with point sources, and perhaps allowing a more refined estimate of the Hubble constant if the source is variable.

We describe here an iterative procedure that we have developed to invert gravitational lens images of extended sources. This procedure has performed well in several blind trials, and we have applied it to MG1131+0456. The results are quite encouraging; the method converges to a robust and plausible solution.

This method of reconstructing the source given an image and a lens has been reported by Kayser and Schramm (1988), although they did not use it to estimate an error in the inversion nor did they combine it with an iterative method to solve for the correct inversion.

2. The Lensing Equation and a Parametrized Potential

The lensing equation for a single lens is

$$\underline{u} = \underline{x} - \nabla\phi(\underline{x}) \tag{2.1}$$

where ϕ is equal to $2D_{LS}/D_{OS}$ times the 2-dimensional Newtonian gravitational potential and $\vec{u}$, $\vec{x}$ are small angles denoting the source and image positions respectively. D_{ij} are angular diameter distances (e.g., Bourassa and Kantowski, 1975), where i, j = O(bserver), L(ens), S(ource).

The inversion procedure described in §3 is independent of the particular two-dimensional potential used to model the lens. However, we need to choose a particular parametrized model for testing and application. In this paper we use the following simple form, which is adequate for individual galaxies and clusters of galaxies that are ellipsoidal and approximately isothermal,

$$\phi(x,y) = b\left[s^2 + (1-\epsilon)x'^2 + (1+\epsilon)y'^2\right]^{1/2} \tag{2.2}$$

where

$$b = 4\pi \frac{D_{LS}}{D_{OS}}\left(\frac{\sigma_{1D}}{c}\right)^2. \tag{2.3}$$

The coordinates (x',y') are centered on the lens and are oriented so as to be parallel to the principal axes of the potential; they are related to the sky coordinates (x,y) by a translation (x_0,y_0) and a counter-clockwise rotation (θ). The quantity b, proportional to the square of the one-dimensional velocity dispersion σ_{1D} of the lens, is the deflection scale of the potential, s is the core size of the lens and ϵ is its projected eccentricity. Our model is thus specified by six parameters: x_0, y_0, θ, b, s, ϵ. For this model potential, the eccentricity of the surface density distribution is three times that of the potential.

The model employs two parameters, b and s, to describe the radial variation of the potential; this may be more general than necessary. For instance, if an arc or ring is generated by a smooth, isolated mass distribution and is thin in its radial extent, it is sufficient to expand the radial variation of the potential around the Einstein ring. The linear term, which is proportional to the total enclosed mass, determines the mean ring radius and is the only required parameter; changes in the quadratic term are essentially indistinguishable from changes in the overall size of the source, and higher-order terms are ignorably small. We used the above more general model because MG1131+0456 extends over a reasonable range of radius. In our model fits of this system the parameter s was small and poorly determined suggesting (a posteriori) that it would have been adequate to set $s = 0$. This will probably not be the case for applications to arcs in clusters of galaxies.

The angular variation of the potential in (2.2) has no dipolar contribution, because such a term is equivalent to a translation of the lens. All higher order multipoles are non-trivial, but the model is restricted to a quadrupole term, with a strength measured by ϵ. This should be a reasonable approximation if there are no substantial masses close to the Einstein ring. It will not be a good approximation when, for example, we model a cluster potential and the image cuts across individual galaxies. Different types of images may motivate alternative lens models. For example, in the case of an arc involving an isolated cusp in the caustic, a five parameter model (two position coordinates, a cusp scale, an image scale, and an orientation) should suffice.

3. Inverting an Image

We seek an algorithm that takes the observed image of a gravitational lens candidate and deduces the original unlensed source structure as well as the parameters describing the lens. We have explored a variety of procedures to carry out this inversion. The one

that we describe and apply appears to be reasonably robust despite its seemingly *ad hoc* character.

We divide the image and source planes into pixels, and associate with each image pixel the corresponding measured intensity I. Initially, we guess a set of lens parameters $(x_o, y_o, \theta, b, s, \epsilon)$ and use equation (2.1) to map each image pixel to the source plane. We then identify the nearest source pixel for each image pixel and, since intensity should be conserved by a lens (excepting a constant redshift factor), we associate the image intensity with that source pixel. Because of multiple imaging, many of the source pixels will be associated with more than one image intensity. If the lens model is correct, these intensities will agree to within the noise; if the model is wrong, the intensities will differ substantially compared to the noise. Therefore, we can associate a normalized error measure e_{ij} with each multiply-imaged source pixel (i, j),

$$e_{ij} = \frac{1}{\sigma^2(m_{ij} - 1)} \sum_{k=1}^{m_{ij}} \left(I_{ij}^k - \bar{I}_{ij} \right)^2 , \qquad (3.1)$$

where m_{ij} is the multiplicity of the source point and $\bar{I}_{ij}$ is the mean of the m_{ij} independent image intensities I_{ij}^k mapped onto that point. Note that e_{ij} will be small at singly-imaged source points. We now calculate an over-all error $\bar{e}$, averaged over the n multiply-imaged source points,

$$\bar{e} = \frac{1}{n} \sum_{i,j=1}^{n} e_{ij}. \qquad (3.2)$$

It is necessary to normalize this error by n to avoid prejudicing the method towards reducing the error by merely eliminating multiply imaged pixels.

Unfortunately, simply varying the lens parameters so as to minimize $\bar{e}$ frequently leads to unacceptable local minima, corresponding to complex or highly fragmented sources and images. We have found that the algorithm converges more efficiently if we include the following additional features in the error estimate:

(i) a <u>sparseness</u> measure S, which measures the mean number of faint (*i.e.,* , image intensity less than noise) neighbor pixels per bright pixel in the source. We penalize models with large sparseness measures.

(ii) an <u>overlap</u> O, which is the fraction of image pixels that have changed from faint (image intensity less than σ) to bright or *vice versa* in comparing the original image with the reconstructed image. Singly-imaged points do not contribute to O. We penalize models for which O is large.

(iii) a lens <u>filter</u>, which discourages unphysical lens parameters (*e.g.,* negative core radii or making the lens subcritical and unable to generate multiple images). In practice, we use a function F, which is zero when all the parameters lie within acceptable ranges and grows exponentially whenever one of them lies outside this range.

Combining all the terms, our final figure-of-merit is

$$E = \bar{e}(1 + w_S S + w_O O) + w_F F \qquad (3.3)$$

where w_S, w_O, w_F are adjustable coefficients typically chosen so that $w_S S$, and $w_O O \sim 1$ and $w_F = 1$. We emphasize that all three coefficients can be set to zero whenever an acceptable solution is near. If there are measurements at several wavelengths, or on several Stokes parameters at a single wavelength, then the individual figures of merit can be combined.

Our procedure then is to vary the lens parameters so as to minimize E. In order to test this procedure and to compare it with alternative procedures, we carried out a series of blind trials using fairly simple source models. We found that it was necessary to perform a low resolution search of parameter space to locate suitable starting solutions. However, in every case it eventually was possible to guess initial parameters sufficiently close to the true solution to ensure rapid convergence to an acceptable reproduction of the image. Moreover, local extrema were clearly distinguishable from the true solution both in terms of the value of E and from visual appearance. An example is shown in Figure 1.

4. Application to MG1131+0456

Although our original intention was to devise a procedure for analyzing images of optical arcs, the discovery of an elliptical ring in MG1131+0456 by Hewitt *et al.* (1988) allows an immediate application. The 15 GHz map of this object, made with a FWHP beam of 0.″12, has a peak brightness of 2.6 mJy, 22.6 times the 1σ noise level. The 5 GHz map, made with a 0.″39 beam, has a peak brightness of 23 mJy, 460 times the 1σ noise level. Polarization data are also available at 5 GHz.

The ring itself is only $\sim 2''$ across, and there is a significant loss of spatial information in the 5 GHz map. Moreover, the length scale of the lens parameters (see below) is only slightly larger than the 5 GHz beam. For these reasons, the 15 GHz map is a better starting point for the inversion procedure. Unfortunately, the signal-to-noise ratio of the ring itself in the 15 GHz map is only about two. The noise level in the map introduces significant errors at the starting point of the inversion. Fortunately, the multiple images of a given source point are typically separated by more than a beamwidth. Since the noise at points separated by more than a beamwidth is uncorrelated, it has little effect on the inversion.

Plausible initial guesses were made for the lens following a coarse search of parameter space by hand, and the inversion procedure optimized these values as described in § 3. The final values of the parameters are shown in Table 1. The resulting source is shown in Figure 2, along with the image that the lens model would produce of the source. The inferred source structure is reassuringly similar to that of many high redshift radio sources. Note that the source completely covers only two cusps and one fold of the tangential caustic; this leads to the observed gap in the ring at 15 GHz.

Table 1 also shows the change in the values of the parameters that lead to a 50% increase in the error function $\bar{e}$. This is calculated by computing the errors for $3^6 = 729$ nearby lens models where each parameter takes either its converged value or a nearby value ($\pm 0.″1$ for x_0, y_0, and b, 0 and $2s$ for s, and $\pm 10\%$ for ϵ and θ). An approximate second derivative $\bar{e}_{,pp}$ for the change in $\bar{e}$ as parameter p is varied *with all of the other parameters varying so as to minimize the error due to the change in p* is computed from a three point finite difference approximation. The quantity $(\bar{e}/\bar{e}_{,pp})^{1/2}$ measures the sensitivity of the model to changes in parameter p. Formally, it is the change in p for which the error measure is 50% larger than the value at the solution. The model is most sensitive to changes in x_0, y_0, and b where changes of as little as a tenth of an arc second lead to large increases in the error. The angle θ is fixed to approximately ten degrees if the error increases by less than 50%. The solution is less sensitive to changes in ϵ, with changes of order 20 to 30% leading to 50% changes in the error measure. The solution is extremely insensitive to changes in the core size s: this is a consequence of having a strong (nearly singular) lens for which $b/s \gg 1$. The absence of an image in the core of the galaxy requires $b/s \gtrsim 5$ based on flux constraints.

Figure 3 shows the result of applying the lens model to the 5 GHz map. Although the lens model was not constructed to minimize the errors in the inversion of the 5 GHz map, the results are visually superior to the 15 GHz map because of the higher signal-to-noise ratio in the ring. Because of the lower resolution in the 5 GHz map the image forms a complete ring and the source covers the tangential caustic instead of only two cusps as in the 15 GHz map. The small knot in the inverted image, which lies in the core of the lens, appears because the lens cannot duplicate the uniform broadening of the observed image produced by the large 5 GHz beam. This leads to a spurious overlap of the core of the lens by the model image of the ring, even though no flux is observed coming from that region. The observations are, however, more sensitive at 5 GHz than at 15 GHz and the inferred source structure includes a singly imaged southwestern lobe at the lower frequency.

In Figure 4 we use the 15 GHz lens model and the polarization maps at 5 GHz to make a polarization map of the source. There is a region near the critical curve where the polarized intensity is too low. Given that this region of the image plane corresponds to a region near the tangential caustic in the source plane that is small compared to the 5 GHz beam, such errors are not surprising. Of more importance is the fact that the polarization direction in the reconstructed image follows the observed image so well. Because of the averaging procedure used to generate the source, the close alignment of the vectors between the image and the inverted image provides strong support for the model.

Finally, in Figure 5 we show contours of velocity dispersion in the lens-source red-shift plane, calculated using equation (2.3) and the model parameters from Table 1. For plausible values of the lens (0.5) and source (2.0) redshifts, the velocity dispersion lies in a range typical for galaxies. We can calculate the time delay between the two compact components to be

$$\Delta t \simeq (1 + z_L) \frac{D_{OL} D_{OS}}{r_H D_{LS}} h^{-1} \left(0.26 \pm 0.07\right) \text{ years} \tag{4.1}$$

where $r_H = c H_0^{-1}$ is the Hubble radius, and $h = H_0 / 100 \text{ km s}^{-1} \text{Mpc}^{-1}$ (Blandford and Narayan, 1986). The dispersion represents the (approximate) variation in the time delay that is consistent with a 50% rise in the error measure $\bar{e}$. Despite the highly constrained nature of the model, a significant variation in the time delay is still allowed within the error estimates. In Figure 6, we show contours of time delay as a function of the source and lens redshifts for an Einstein-DeSitter cosmological model. We can also compute the amplification tensor at the position of the compact components, which would relate VLBI images of any features in the compact cores. The transformation from the compact component to the northeast to the compact component to the southwest has eigenvectors with position angles on the sky of $\theta_1 \simeq 57°$ and $\theta_2 \simeq -63°$ and eigenvalues $\lambda_1 = 0.96$ and $\lambda_2 = -1.24$ respectively. This implies that the northeast component is deamplified by a factor of 0.84 relative to the southwest component. If we write the transformation matrix in terms of a convergence κ and a shear γ, $\kappa = (\lambda_1 + \lambda_2)/2 = -.14$ and $\gamma = (\lambda_1 - \lambda_2)/2 = 1.10$. The two components have opposite parities (because $\lambda_1 \lambda_2 < 0$) with the NE component reflected relative to the SW component through the line connecting them. The SW component lies at a minimum of the time delay surface (making it the first to vary) while the NE component lies on a saddle point of the time delay surface.

The lens potential required to create this image is elongated along position angle 150° (north through east). We might therefore expect that the light distribution from the

underlying galaxy should also be elongated along this direction. Unfortunately, preliminary optical observations of the source region show extended emission elongated roughly parallel to the ring (Hewitt *et al.*, 1988, Turner, private communication). In view of this discrepancy, we have attempted to find lens models elongated along position angle $\sim 60°$. No such models were found. This result was surprising: we might expect that the ring is simply tracing the critical curve, and the underlying galaxy is elongated along position angle $\sim 60°$. Since all reasonable elliptical potentials have critical curves elongated in the same direction as the potential or surface mass density distribution, this would lead to a lens elongated in the same direction as the ring.

The reason why an acceptable model of this type was not found can be understood from Figures 7 and 8. We suppose that the two bright knots are images of the same source region, lying outside the tangential caustic and along either the minor or the major axis of the critical curve. One image of the knot will lie outside the critical curve, the other inside. A lens elongated in the same direction as the ring cannot produce images of the knot that are connected to the ring without, at the same time, filling the observed gap in the ring. This means that no satisfactory model can be constructed with a core-jet source and a simple elliptical potential elongated parallel to the ring; however, we cannot rule out models with more complex potentials, or models in which the knots are unrelated foreground sources.

Given that the observational possibilities are far from exhausted, and that so many key quantities are unknown, there seems little point in trying to produce a detailed model at present. It is worth noting, however, that while the simple elliptical potential considered here can reproduce the gross features of MG1131, it is not entirely satisfactory, even without consideration of observed optical emission. The model ring is too round, primarily because of the orientation of the critical line. To some extent this can be improved by adjusting the ellipticity (ϵ) and velocity dispersion (*i.e.*, b) of the potential. Unfortunately, improving the shape of the ring degrades the reproduction of the observed knots. Further improvements will require a more complicated model potential, not simply adjustments to the parameters.

5. Conclusions

The inversion technique described in this paper is useful in the study of extended images produced by gravitational lenses. Its primary advantage over previous methods is that no preconceived notions concerning the unlensed structure of the source are required—an acceptably low figure of merit automatically generates a good model of the source. The sparseness, overlap, and lens filter penalties, used as aids to avoid local minima in the error, might bias the source structure, but are set to zero for the final optimization.

The resulting solution uses all of the available image data including information from observations at different wavelengths and polarizations. The quality of the solution can be judged visually by direct comparison of the original image with the image generated by the solution. Furthermore, a well-defined figure of merit, given along with the inversion solution, provides a simple way for later studies to determine whether a new solution represents a significant improvement.

Our application to MG1131+0456 is made plausible by the result: a typical radio source lensed by a typical galaxy. While we are confident that our model represents the correct lensing geometry, we cannot fully reproduce the ring with our simplified potential. We have found solutions that reproduce either the ring or the compact components well,

but at the expense of introducing some distortion in the other feature. For a lens at $z_L \simeq 0.5$ and a source at $z_s \simeq 2.0$, we find velocity dispersions $\sigma \simeq 250 \pm 20$ and time delays between the compact components of $\Delta t \simeq (0.19 \pm 0.05)h^{-1}$ years. The quality of the inversion will be improved when higher sensitivity observations are available at 15 GHz, and higher resolution observations at 5 GHz are made. Application of the method to the giant optical arc in Abell 370 is underway.

Acknowledgements

References

Blandford, R. and Narayan, R., 1986, *Astrophys. J.,* **310**, 568.

Blandford, R.D., and Kovner, I., 1988, *Phys. Rev. A,* **38**, 4028.

Bourassa, R.R., and Kantowski, R., 1975, *Astrophys. J.,* **195**, 13.

Gorenstein, M.V., Shapiro, I.I., Rogers, A.E., Cohen, N.L., Corey, B.E., Porcas, R.W., Falco, E.E., Bonometti, R.J., Preston, R.A., Rius, A., and Whitney, A.R., 1984, *Astrophys. J.,* **287**, 538.

Grossman, S., and Narayan, R., 1988, *Astrophys. J. (Letters),* **324**, L27.

Hewitt, J.N., Turner, E.L., Schneider, D.P., Burke, B.F., Langston, G.I., and Lawrence, C.R., 1988, *Nature,* **333**, 537.

Kayser, R., and Schramm, T., 1988, *Astr. Astrophys.,* **191**, 39.

Kovner, I., 1988, in *The Post-Recombination Universe*, NATO ASI, ed. N. Kaiser and A.N. Lasenby, (Kluwer Academic: Dortrecht), p. 315.

Lavery, R.J. and Henry, J.P., 1988, *Astrophys. J. (Letters),* **329**, L21.

Lynds, R., and Petrosian, V., 1986, *Bull. Am. Astr. Soc.,* **18**, 1014.

Narasimha, D., and Chitre, S.M., 1988, *Astr. J.,* **97**, 327.

Paczyński, B., 1987, *Nature,* **325**, 572.

Soucail, G., Fort, B., Mellier, Y., and Picat, J.P., 1987, *Astr. Astrophys.,* **172**, L14.

TABLE 1. Fit Parameters and Errors

Parameter	$x_0('')^2$	$y_0('')^2$	ϵ	θ^3	$b('')$	$s('')$
Value	-0.70	$+0.11$	0.14	150	0.97	0.07
Variation[1]	0.15	0.16	0.04	12	0.11	0.20
Errors	$\bar{e}$	S	O			
	0.21	1.83	0.11			

[1] The variation is an estimate of the change in the parameter that will lead to a 50% increase in the error measure, when all other parameters are varied to minimize the error due to the change.

[2] The position is measured as an offset from the center of the map.

[3] Position angle (north through east) of the major axis (x axis) of the potential.
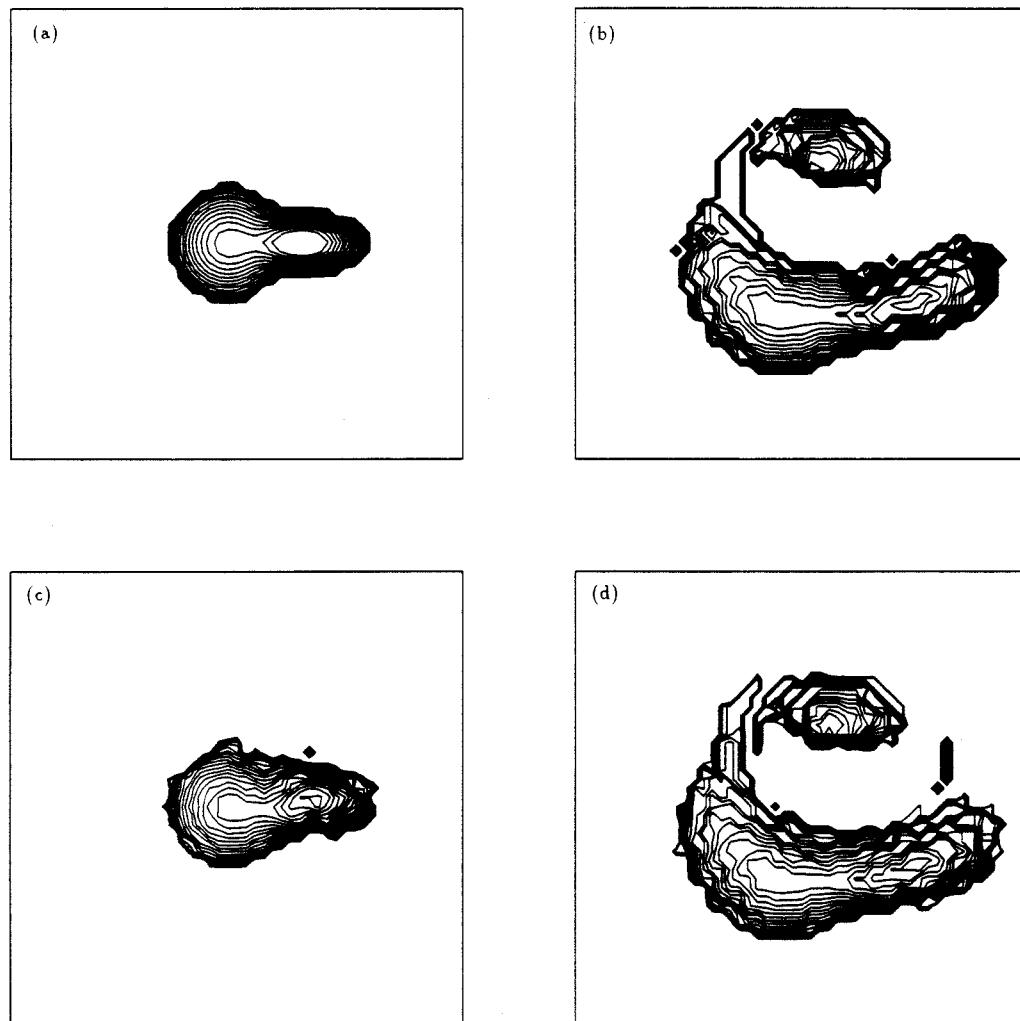
Figure 1. An example of a test problem for the convergence method. Shown are the synthetic source (a) and image (b) used for a blind test of the convergence procedure. The converged solutions for the source (c) and image (d) are shown for comparison.
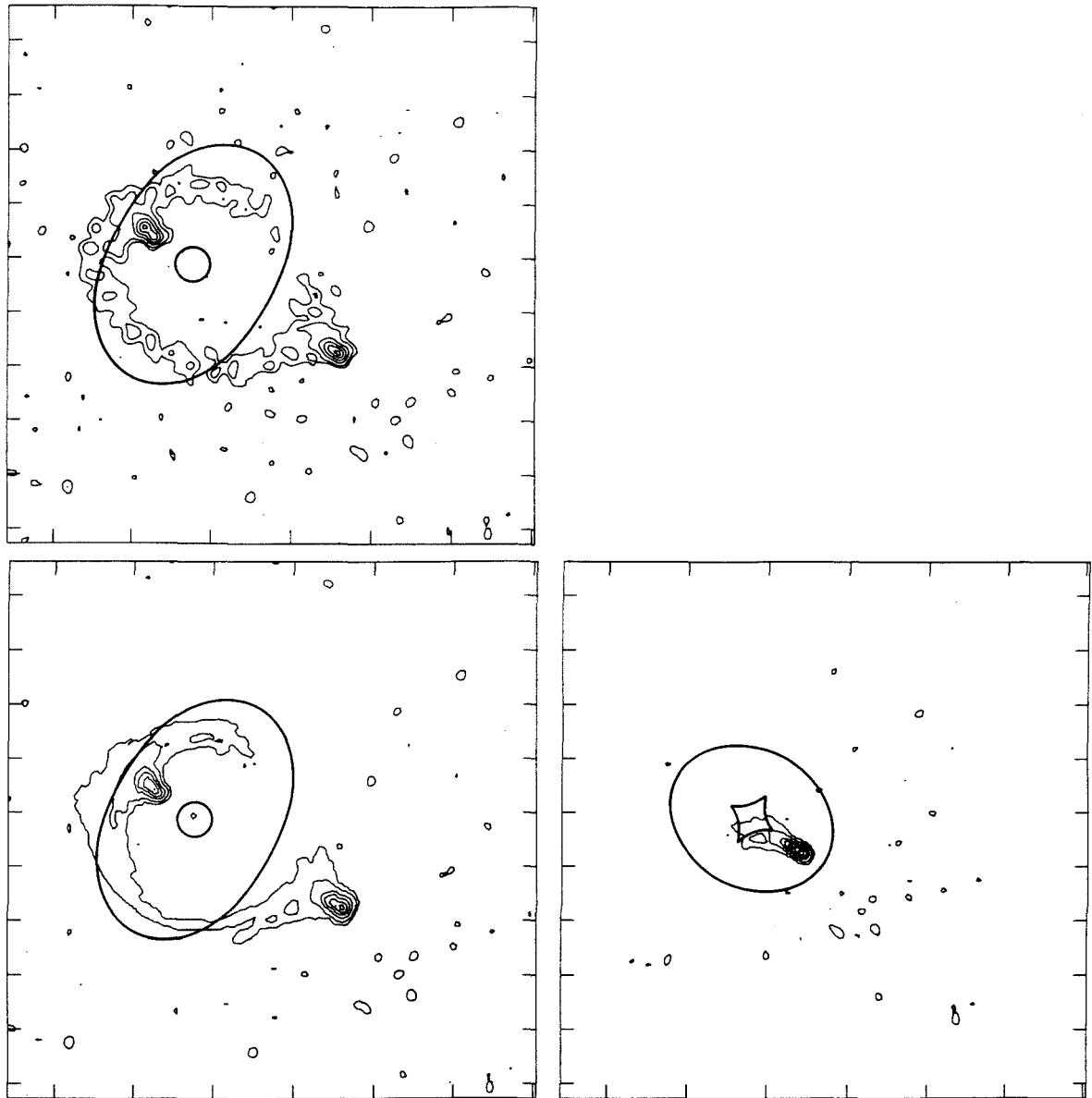
Figure 2. The 15 GHz map of MG1131+0456 (a) shown with the inversion for the source (b) and the inverted image (c). The inversions have been smoothed over a kernel three pixels wide (0."1) to eliminate numerical pixel-to-pixel noise. Contour levels are 10, 20, 40, 60, 80, and 95% of the peak flux density. The noise level is at approximately 10% of the peak flux density. All three panels show the same region of the sky, with North at top and East to the left. The sky coordinates of the map centers are $\alpha = 11^{h}31^{m}56^{s}.40$ and $\delta = 04°55'49."40$ (J2000.0). Tick marks are at 0."5 intervals on the declination axis, and at $0^{s}.05$ intervals on the right ascension axis. The critical curves are superimposed in (a) and (c) and the caustics are superimposed in (b).
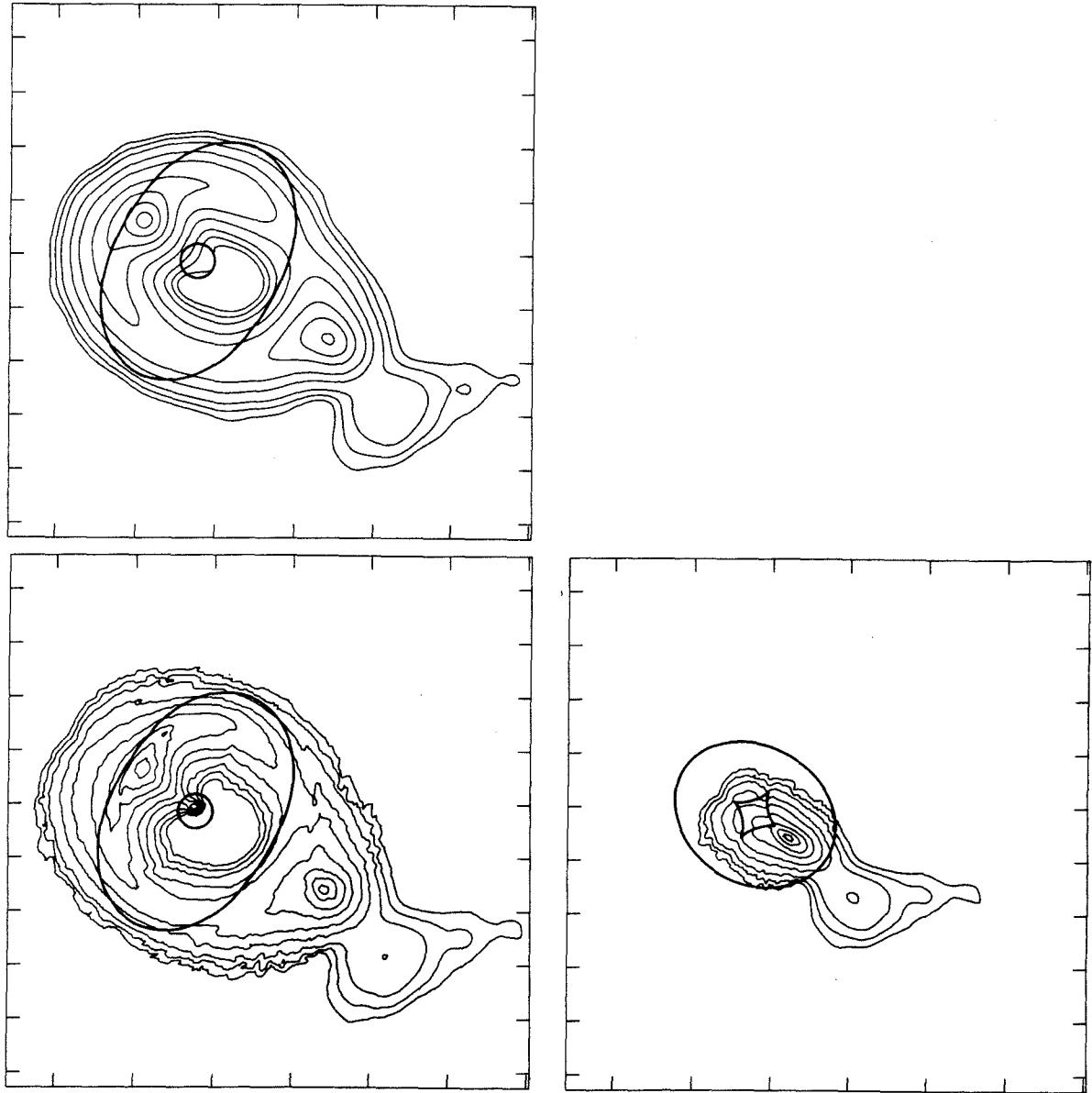
Figure 3. The 5 GHz map of MG1131+0456 (a) shown with the inversion for the source (b) and the inverted image (c). The inversion is smoothed over a kernel three pixels wide (0''1). Contour levels are 1, 2, 4, 10, 20, 40, 60, 80, and 95% of the peak flux density. The noise level is at approximately 1% of the peak flux density. Note that these inverted maps were generated using the inversion computed for the 15 GHz map. The region of sky shown is identical to that in **Figure 2.**
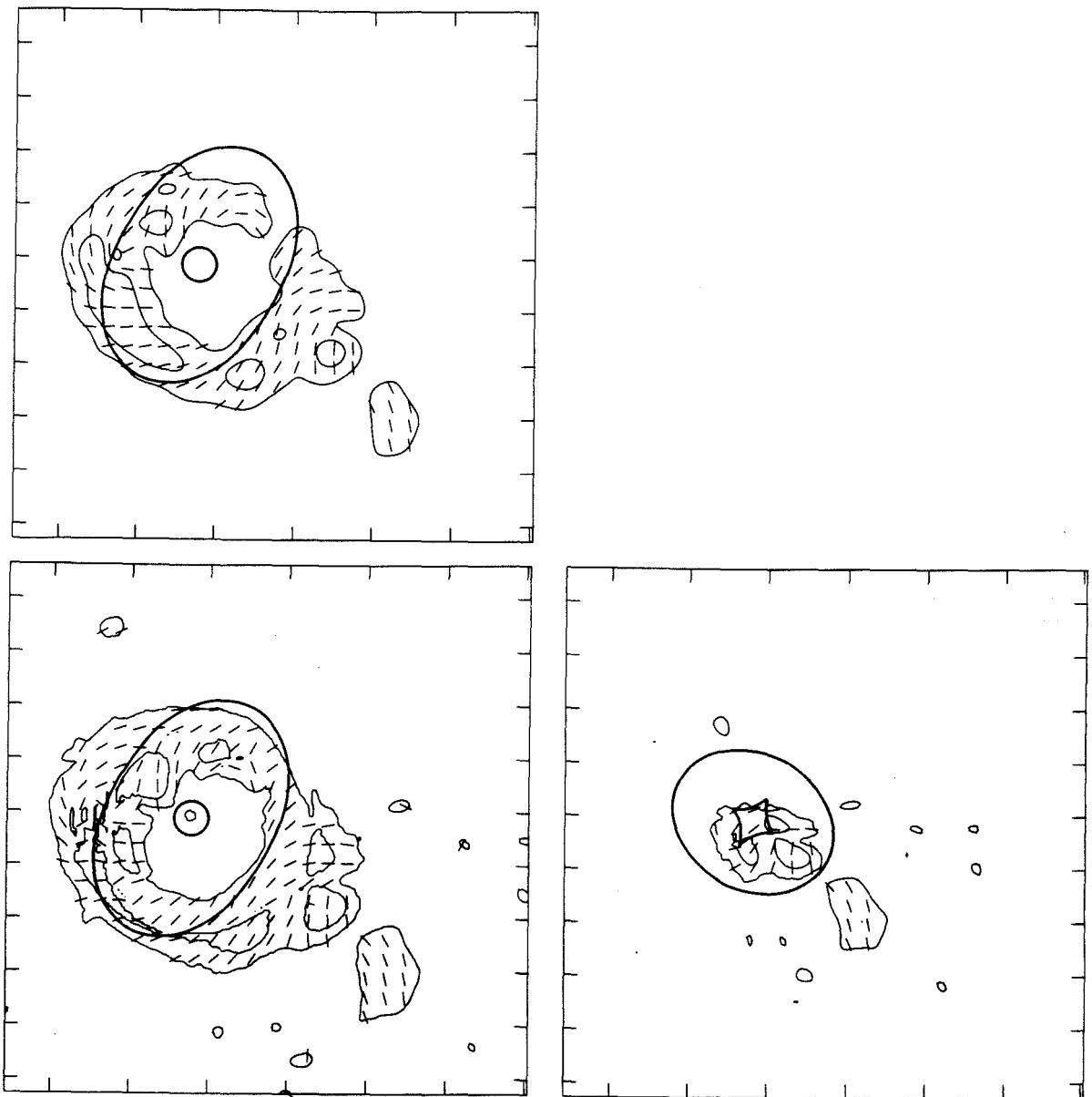
Figure 4. Polarized intensity $(Q^2 + U^2)^{1/2}$ map of MG1131+0456 at 5 GHz (a), shown with the inversion for the source (b) and the inverted image (c). The inversion is smoothed over a kernel three pixels wide ($0\overset{''}{.}1$). Contour levels are 15 and 50% of the peak polarized flux density. Superimposed line segments of constant length show the local direction of the electric field. The noise level is at approximately 10% of the peak polarized flux density. The Q and U maps were inverted separately, using the inversion computed for the 15 GHz map, and then summed and smoothed to form the polarized intensity map. In (c), the inner contours in the southwest quadrant are at the 15% level, while the corresponding contours in (a) are at the 50% level. This conversion of a ridge in the observed image to a trough in the inverted image is at least partly due to the low resolution of the polarized map. The region of sky shown is identical to that in **Figure 2.**
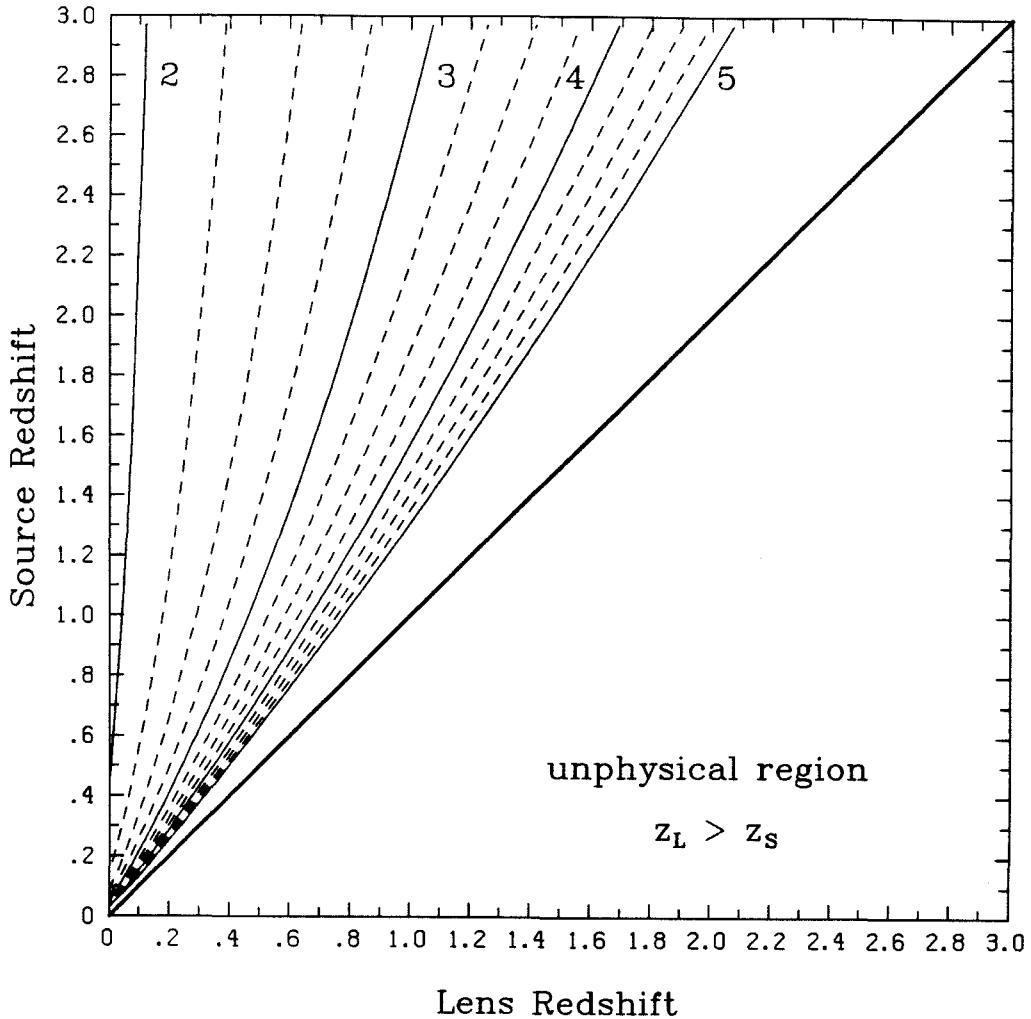
Figure 5. A contour plot of velocity dispersion in the plane of lens and source redshifts. Contour intervals are 25 km s^{-1}, and the values of the velocity dispersion on the solid lines are given in units of 100 km s^{-1}. The contours are computed using equation (2.3) relating the asymptotic deflection b to the source and lens redshifts and the velocity dispersion of the galaxy. The assumed cosmology is an $\Omega = 1$ Einstein-DeSitter cosmology. Note that the canonical values of $z_L \sim 0.5$ and $z_S \sim 2$ give $\sigma \sim 250$ km s^{-1}.
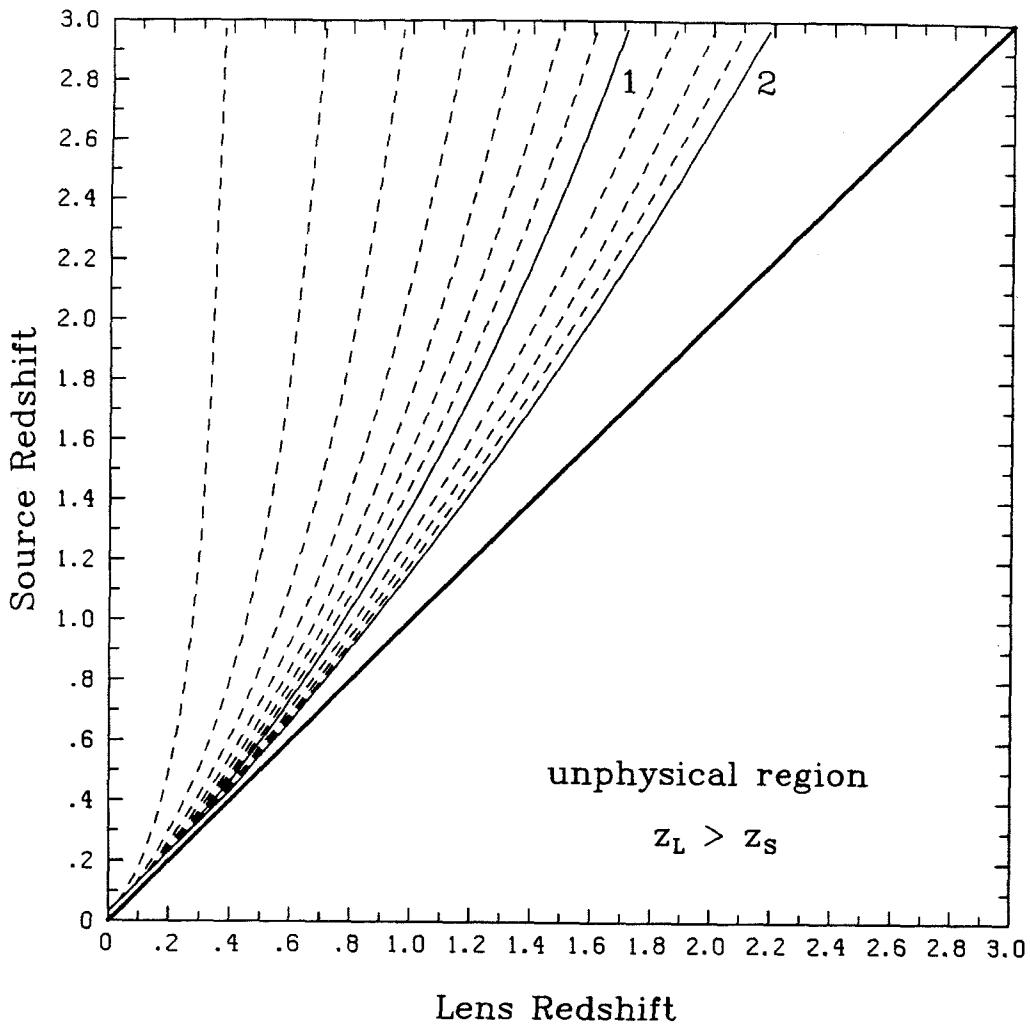
Figure 6. A contour plot of the time delay between the compact images in the plane of lens and source redshifts. The solid lines show the contours for time delays of one and two years. Below one year, the contour increments are in units of 1/8 of a year, and between one year and two years they are in units of 1/4 of a year. The contours are computed using equation (4.1) and assuming an $\Omega = 1$ Einstein-DeSitter cosmology. Note that the canonical values of $z_L \sim 0.5$ and $z_S \sim 2$ give $\Delta t \sim 0.19 h^{-1}$ years.
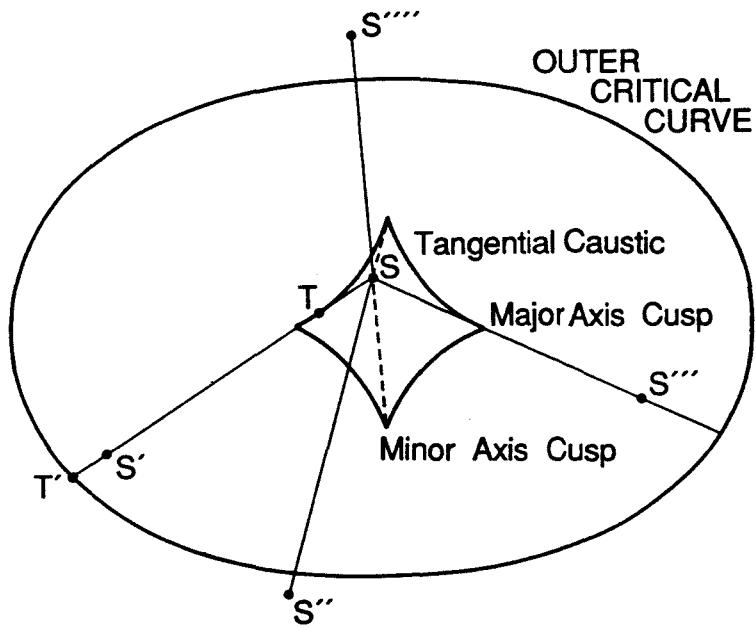
Figure 7. Graphical construction for the mapping of a source point S within a caustic onto the image plane (after Blandford and Kovner, 1988). Four tangents to the caustic passing through S are drawn. The points of tangency on the caustic (*e.g.*, T) map onto the critical curve (*e.g.*, T'). Other source points are mapped in order along the tangent with a constant magnification (*e.g.*, $S \rightarrow S'$). The four tangents produce four images (S', S'', S''', S''''). Note that a minor axis cusp maps source points onto the opposite side of the lens, and a major axis cusp maps source points onto the same side of the lens. A source near a minor axis cusp will produce three neighboring bright images on the opposite side of the lens. Point sources lying outside the caustic produce two images. The fifth (or third) image always lies in the core of the lens potential.
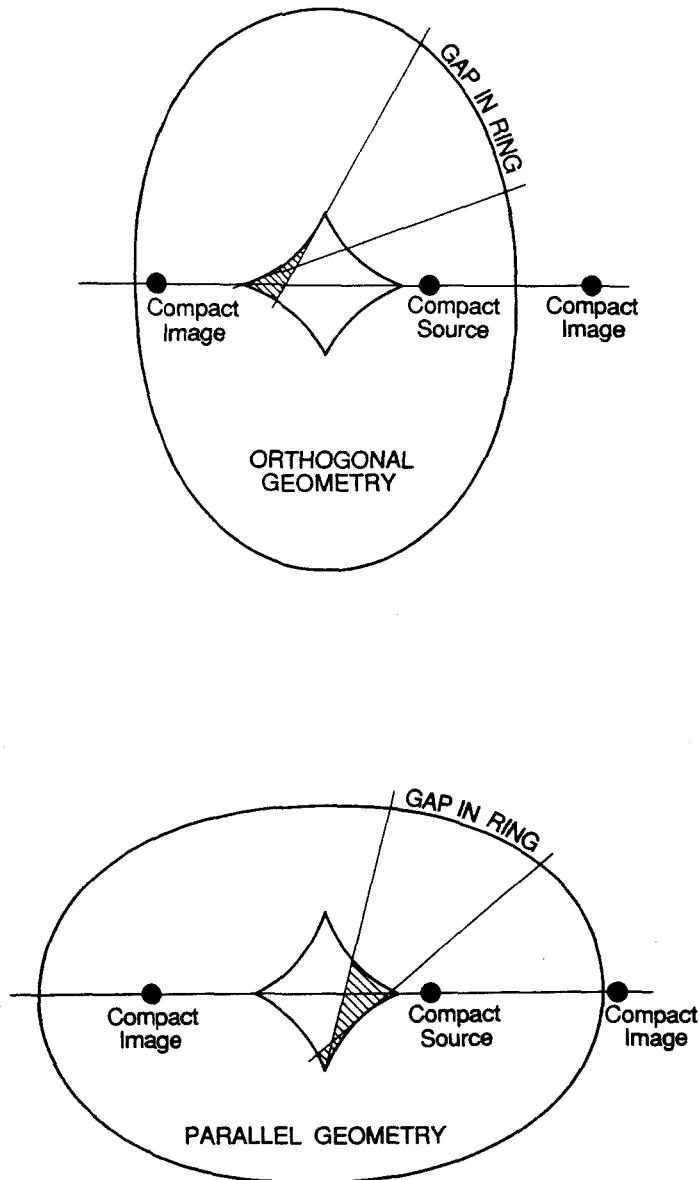
Figure 8. Schematic illustrations of the "parallel" (a) and "orthogonal" (b) models. In each case, we show the position of the knot outside a cusp, a major axis cusp in the case of the parallel model, and a minor axis cusp in the case of the orthogonal model. Also shown are the positions of the images of the knot relative to the critical curve. In the parallel case, the outer image of the knot lies closer to the critical curve than the inner knot, while in the orthogonal case, the arrangement is reversed. The hatched region shows the area of source space excluded because of the gap in the ring. The excluded region extends outside the cusp as well. The source must cover two cusps to produce the observed partial ring. In the case of the parallel model this requires the compact source to be disconnected from the source of the extended emission. This discontinuity in the source would appear as a discontinuity in the image between the compact images and the extended emission. Because no such discontinuity is observed, the parallel model is ruled out. The orthogonal model does not have this difficulty.

Chapter 6

Hollow Conical Jet Models for SS433:

A Paradigm Lost?

Christopher S. Kochanek,

Theoretical Astrophysics 130-33,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

and

John F. Hawley

Dept. of Astronomy
University of Virginia,
Charlottesville, VA 22903

Abstract

A precessing jet can be approximated as an axisymmetric flow if the precession time is short compared to the propagation time over scales of interest. An example is the jet in SS433 where the precession time is $\simeq 0.5$ years and the propagation time from SS433 to the W50 remnant is $\gtrsim 1000$ years. We perform a series of simulations of precessing jets using an R-Z (axisymmetric) finite difference hydrodynamics code. We first reproduce the results for the *filled jets* studied previously by several groups. Next we examine *hollow cylindrical jets*, which allow us to examine the effects of a hollow jet without the complications of a growing interior volume. This case may serve as a "post-focusing" model of a precessing jet. Finally we examine *hollow conical jets*, which model the behavior of a precessing jet propagating on the surface of its precession cone. The conical jets "stall" at the point where the momentum flux density in the jet becomes too low to push the ambient gas from its path. If the area of the working surface over which the jet transfers momentum to the ambient medium is A_w, the area of the jet orifice is A_j, and the initial internal Mach number of the jet is M_j, the jet stalls when $A_w = M_j^2 A_j$. The jet propagates in an extremely unsteady fashion, with large radial oscillations in the position of the jet's channel. The oscillations cause A_w to grow $\propto R^2$ where R is the outer radius of the jet, so that the jet stalls at $R \simeq M_j R_0$ where R_0 is the initial outer radius of the jet. The oscillations of the jet are caused by the competition between the supersonic vortices that make up the cocoon around the exterior of the jet, and the shocked ambient gas on the interior of the jet. While the cocoon can induce a temporary focusing of the jet, the pressure in the interior is able to prevent any permanent focusing. The bow shocks of the hollow jets are much flatter than those of the filled jets and no feature resembling the "ears" of the W50 remnant develops. It seems unlikely that the "ears" of W50 can be formed by a *hydrodynamic jet* ejected with a precession cone as large as that currently observed in the SS433 system.

1. Introduction and Motivation

High energy radio jets have inspired an extensive series of numerical hydrodynamics simulations. Although these calculations have not in the strictest sense served as *models* for any extragalactic radio jet, they have proved valuable in exploring and understanding the physics of supersonic jets. Most of these simulations have been two-dimensional, axisymmetric or slab symmetric models both with and without magnetic fields. Exotic jet geometries have yet to be examined in detail, as they are less relevant than the collimated cylinder. There are, however, a few restricted cases in which more exotic flow geometries are required. In particular, the galactic source SS433 is naturally modeled by injecting the jet into the ambient medium on the surface of a cone: a hollow, conical jet instead of a filled, cylindrical, canonical jet.

The object SS433 is a unique example of a radio jet, one for which we have extensive observations and excellent kinematic information (reviewed by Margon 1984). The time-varying Doppler shifts observed in the emission lines of the system are believed to originate in two opposing jets with velocities of $0.26c$, precessing with a period of 164 days on the surface of a cone with a half-angle of 20 degrees. VLA observations on scales of 0.1-0.5" (Hjellming and Johnston 1982, 1985) confirm this picture, revealing regions of emission arranged in a helical pattern, which are interpreted as the instantaneous pattern of ejected knots of emission on the precession cone. VLBI observations on scales of 50-300 mas (Vermeulen *et al.*, 1987) are also consistent with a series of emitting clumps of gas moving along the helical trajectory predicted by the kinematic model. These direct observations of the precessing jet, along with the kinematic model for the system, place SS433 at a distance of 5.5 kpc (1" = 0.03 pc at 5.5 kpc).

Surrounding SS433 is the asymmetric radio shell (and possible supernova remnant) W50, which has a width along the center of the jet precession cone (the major axis) of about $2°$ (Downes, Pauls, and Salter 1986). The width along the minor axis is about $1°$ ($= 97$ pc at $D = 5.5$ kpc). W50 is notable for its asymmetry, in particular the presence of "ears," or projections that coincide with the precession cone axes of the radio jets. The coincidence of alignment between the jet axis and the ears leads to the paradigm for the jet in SS433: the ears of W50 are formed by the dynamical effects of the jet. If the precession cone is drawn around the axis, the bulges have a half angle of 10 to 15 degrees compared with the 20 degree half angle of the jet as seen on the VLA scale. Hence, the paradigm also assumes that the jet is partially focused relative to its precession cone as it propagates.

There is some evidence other than the alignment of the jets and the ears to support the picture of interaction between the jets and W50. Optical filaments have been detected near the inner edge of the ears and within the precession cone (Zealey, Dopita and Malin 1980; Kirshner and Chevalier 1980). X-ray observations of the system (Watson *et al.*, 1983) reveal bright, diffuse lobes of emission lying along the major axis of W50 at distances ranging from 25 to 70 pc. The peak of the emission is well within the precession cone of the jets, although some emission is detected outside of the limits defined by the cone. The assumption is that the jets are responsible for both the optical emission, and for energizing and confining these X-ray lobes.

The specific model for the system is the hydrodynamical *hollow, conical* jet (Begelman *et al.*, 1980, Davidson and McCray, 1980, Königl, 1983). The jet drives shocks into the interior of the precession cone, heating the ambient or jet cocoon gas to X-ray luminescence. The jet terminates at the edge of W50, producing the ears. Some focusing or collimation is required to account for the ears' reduced angular size. While this overall picture is

appealing, it is not based on a rigorous examination of the physics of such a jet. Several key question remain unanswered. What is the jet geometry out in the region of 30-100 pc from SS433? How do what appear to be ballistic knots of high energy plasma seen at small radii by the VLA become hydrodynamic? How does a hydrodynamic hollow jet produce and confine the X-ray lobes and energize the ears? Do precessing jets collimate or focus themselves as they propagate through an ambient medium?

These questions can be answered in detail only through numerical simulations of the propagation of hollow conical jets. While the simulations are certainly inspired by the SS433 system, they do not represent an attempt to make a detailed model of the system; at this point we can only hope to gain a qualitative understanding of the physical behavior of a hollow jet. To establish the context for our simulations, we begin with a discussion of some of the details that frame the key issues. First, under what conditions can the jet in SS433 be modeled as an axisymmetric, hydrodynamic jet? The optical line profiles and the radio data present a picture of material ejected coherently on time scales of a few hours. If we assume the jet does not slow down appreciably in the distance to the remnant, the time required for it to flow from SS433 to W50 is approximately one thousand years ($t_p \simeq 1000$ years); in this time the jet will have precessed two thousand times. If the jet slows on larger scales (where there is no observational data on the jet's velocity), the propagation time could be considerably extended, increasing the winding number. At the maximum speed of 0.26c, the separation between loops of the precessing jet is less than 0.04 parsecs. The "bullets" or "blobs" of gas will expand either because they are overpressured with respect to the ambient medium or because of the internal velocity dispersion in the gas. The expansion due to internal pressure will occur at the sound speed of the gas, which for gas at a temperature of $10^4 K$ is $c_s \sim 10$ km sec^{-1}. While there is evidence on the radio scales that the bullets are expanding adiabatically (Vermeulen et al., 1987; Hjellming and Johnston, 1988), such expansion alone will not fill in the interbullet spacing. The sound speed is so much smaller than the propagation velocity ($c_s/v_j \sim 10^{-4}$) that the expansion of the material is much more likely to be dominated by the velocity dispersion of the jet, δv. If the jet has a velocity dispersion δv, then the time scale to fill in the gaps between successive coils of the jet is $t_f \simeq t_{precess} v_j/\delta v$, which for a one percent velocity dispersion ($\delta v/v_j = 0.01$) is $t_f \simeq 50$ years—a small fraction of the time to reach the remnant (compared to $t_f = t_{precess} v_j/c_s \simeq 5000$ years for adiabatic expansion). The current upper limit on the velocity dispersion is $\delta v/v_j < 0.04$ (Milgrom, Anderson, and Margon, 1982). Hence, unless the bullets are generated with a remarkably low velocity dispersion, a precessing jet such as SS433 can be modeled as being axisymmetric about the precession axis on scales that are large compared to the region in which the jet is generated.

Next, can we expect a hollow, conical jet to focus? Eichler (1983) proposed a simple analytic model for focusing hollow jets through interactions with the ambient medium and an *assumed* zero pressure region inside the cone. In the absence of a gravitational potential, or other external momentum sink, the ambient medium cannot provide a time independent mechanism for focusing. Momentum transfer to the ambient medium must eventually drive it out of the path of the jet. An external medium can only induce a *temporary* focusing effect. In Eichler's (1983) model it is the assumed zero pressure region that leads to focusing. Whether or not this is a viable mechanism depends upon the hydrodynamics, and the existence of a mechanism for cooling the interior gas on the dynamical time scale. In our present study, we ignore cooling and concentrate upon the gas dynamics of a traditional ideal-gas jet. If such a jet is to be focused solely by a hydrodynamic

mechanism for many dynamical times, it must occur through "self-interactions" between the jet and its cocoon. We know from numerical studies of filled jets that their dynamics are dominated by their cocoons (Norman et al., 1982 [NSW82]; Norman, Winkler, and Smarr, 1983, 1984 [NWS83,84]; Norman and Winkler, 1985 [NW85]; Kössl and Müller, 1988 [KM88]; Lind et al., 1988 [LPMB88]), and we have the same expectation for the hollow jets.

Through simulation, we can test many of the hypotheses that underlie the current paradigm for the jet in SS433. Our primary assumptions are that the jet is hydrodynamic and approximately axisymmetric on large scales. In §2 we describe the numerical method and the initial data used in the simulations, and compare some cylindrical test jets with previous calculations in the literature. In §3 we discuss some general analytic considerations governing the propagation of jets. In §4 we describe the results of our numerical experiments comparing three different jet geometries, filled jets, *hollow cylindrical jets,* and *hollow conical jets.* The jet Mach numbers and density ratios are chosen to cover the two major regimes of stable, highly supersonic jet propagation: the light, cocoon-dominated jets, and the heavy, ballistic or "naked beam" jets. Finally, in §5 we discuss the consequences of these simulations for the jet paradigm in the SS433 system.

2. The Numerical Method

We wish to consider the dynamical properties of a hydrodynamic jet propagating along the surface of a hollow, axisymmetric cone. We study three basic jet configurations: the *hollow cone,* in which the jet is injected with a fixed annular width on a cone with a $\theta_c = 20°$ half-angle opening, the *hollow cylinder,* a jet with fixed annular width injected parallel to the axis, and the now standard cylindrical *filled jet.* Although there is interest in considering the effects of non-uniform and non-stationary ambient media on jet propagation, we choose to keep our initial study as simple as possible given that we are working with a novel jet geometry. Therefore, the ambient medium is stationary, homogeneous, and isentropic.

For this study we developed a time-explicit, Eulerian finite-difference, Newtonian hydrodynamics code in cylindrical R-Z coordinates. The numerical techniques employed are the same as those described in Hawley, Smarr, and Wilson (1984), *viz.* van Leer's monotonic transport scheme. We calibrated the code using both one-dimensional (*e.g.,* standard shock tube, Sedov line explosion), and two-dimensional (*e.g.,* spherical Sedov explosion in cylindrical coordinates) test problems. We have also made direct comparisons to the published results of NWS83, NWS84, KM88, and LPMB88. These comparisons are particularly useful since, in addition to testing our code, they provide insights into the reasons behind the quantitative differences in the jets obtained by different groups. All codes in our comparison sample are 2D time-explicit. NWS83, NWS84, and LPMB88 use the same monotonic transport scheme used in our code, while KM88 use the flux corrected transport scheme. LPMB88 also employ a total rather than an internal energy equation in the hydrodynamics.

For the simulations we use units in which the ambient pressure P_a and density ρ_a are set equal to one. The length scale is the radius of the filled jet $R_j \equiv 1$. This gives a time scale $t_i = R_j(P_a/\rho_a)^{-1/2}$, which is the *isothermal* sound crossing time across the jet. LPMB88 use the same units, while NWS83, NWS84, and KM88 use the sound crossing time $t_s = \Gamma^{-1/2}t_i$. The input parameters are the jet density $\rho_j \equiv \eta\rho_a$ and the jet internal Mach number $M_j = v_j/c_j$. The jet is injected in pressure equilibrium with the ambient medium $P_j = P_a$. The hollow cylindrical jets have an inner radius of $3/4R_j$, and an outer

radius of $5/4R_j$, corresponding to a thickness to mean radius ratio of $\Delta R/\bar{R} = 1/2$. The hollow conical jets use the same inner and outer radii as the hollow cylindrical jets, but the velocities lie on a cone with a half-angle of 20 degrees. The filled jet and the hollow cylindrical jet have, by design, the same mass, energy, and momentum fluxes. The fluxes in the conical case are a factor of $1/\cos 20° = 1.06$ times larger because the grid zone size and spacing limit the adjustments possible to compensate for the angle the velocity vector makes with respect to the grid. In all cases a $\Gamma = 5/3$ ideal gas equation of state is used.

The simulations are performed on a 256 by 512 R-Z grid. In the axial direction the zone width is $1/20$ of a jet radius, while in the radial direction the zone width is $1/20$ of a jet radius until zone number 192, after which the zone widths are stretched by two percent per zone. This gives a resolution of 20 zones across the radius of the filled jet, and ten zones across the width of the hollow jets. This resolution is comparable to that used in previous jet simulations; NWS83, NWS84, and LPMB88 use 15 grid zones across their filled jets. KM88 use several resolutions up to 100 zones across the jet radius. The figures shown in this paper include the full extent of the computational region.

We employ outflow boundary conditions except at the orifice and on the symmetry axis. This is clearly appropriate for the outer radial boundary, and the downstream axial boundary, but one can select either a reflecting or an outflow condition on the upstream axial boundary. We choose an outflow boundary condition because in our picture of the model system the computational region is sufficiently far from the source of the jet that there is no interaction between this jet and the counter jet on the opposite side. The question of which boundary condition is appropriate for jet simulations has been an issue in the past; the choice of one or the other of these boundary conditions has demonstrable qualitative effects upon the resulting jet (KM88).

We calculate a pair of standard cylindrical filled jets ($M_j = 6.0$ and $\eta = 1.0$ and 0.1) for comparison with simulations by other authors (NSW83, NWS83, NWS84, LPMB88, and KM88), and to provide a baseline against which to compare the results of the hollow jets. While the various jets are qualitatively similar in terms of gross morphology (shock structure, vortex shedding), evolution, and stability, there are differences between them. For example, the rate of advance of the bow shock varies (see Figure 1). The bow shock's advance is controlled by the size of the working surface (*cf.*, equation [3.3]), which in turn depends on the size of the cocoon and the formation of vortices at the head of the jet — in short it is sensitive to all of the details of the jet's structure.

The effect of boundary conditions and numerical resolution on the propagation of jets is studied in KM88. They show that these considerations have significant quantitative effects in the locations of the bow shock and working surface, as well as the structure and vorticity of the cocoon. For example, the development of structure in the cocoon, and hence the effective area of the working surface, depends primarily on the shedding of vortices at the head of the jet. The production of vorticity in the codes will clearly depend on initial conditions, differences in algorithm (numerical viscosity), and boundary conditions. The similarity of the cocoon and shock structures at late times in comparably resolved simulations suggests that the observed differences may be due to initial transients rather than inherent numerical viscosity. Note that all the compared results were obtained with second-order, monotonic schemes, that, although different in implementation, can be expected to have roughly the same numerical diffusion. With the exception of the LPMB88 results, the jets initially show little or no structure and advance ballistically into the ambient medium; only when the flow begins to exhibit complex structures does

the jet slow down. The hydrodynamic jet of LPMB88 has considerably more vortical structure in its cocoon, and a correspondingly smaller velocity for the bow shock. Their jets are closer in appearance to the reflecting boundary condition runs of KM88, which also have more substantial cocoons and slower average bow shock velocities. The implication of these comparisons is that simulations will differ not only in the exact properties of the flow structure at any one time, but even in certain gross properties such as the position of the bow shock and working surface.

While we do not want to minimize problems of resolution, numerical methods, boundary conditions, and initial data, we believe that these simulations do provide valuable insights into the basic physics of supersonic hydrodynamic jets. All of the simulations discussed here agree on the basic morphology and stability properties of the jets. Comparative studies such as that of KM88 delineate the limits of applicability for the present numerical work. Further code comparison and validation is desirable, but the jets seem to be a poor test bed for doing detailed comparisons of numerical techniques. The flow patterns are non-stationary, and the details of the flow continually feed back into the jet to determine properties such as the rate of advance. The differences between the simulations highlight the need for well defined, inherently two-dimensional, test problems in numerical hydrodynamics.

3. Jet Kinematics and Dynamics

While the evolution of a hydrodynamic jet is not amenable to analytic treatment, there are a number of kinematic arguments that give approximate expressions for the behavior of certain large features of the jet. The analysis is based upon nothing more than conservation laws, yet it provides a reasonable understanding of the macroscopic properties of the jet.

We begin by relating the advance of the jet to the rate at which it transfers momentum to its surroundings through the simple requirement of momentum balance in the rest frame of the working surface of the jet. Equating the momentum flux in the jet with the ram pressure of the ambient gas yields

$$A_j \left[\rho_j \left(v_j - v_w \right)^2 + P_j \right] = A_w \left[\rho_a v_w^2 + P_a \right] \tag{3.1}$$

where A_j and A_w are the cross sectional areas of the jet and the working surface. The pressure terms can be neglected in high Mach number jets as they represent corrections of order $M_j^{-2} \ll 1$. The velocity of the working surface is

$$v_w \simeq v_j \frac{\epsilon}{1 + \epsilon} \qquad \epsilon \equiv \left(\frac{\rho_j}{\rho_a} \frac{A_j}{A_w} \right)^{1/2} \tag{3.2}$$

or, in terms of the Mach number of the working surface relative to the sound speed in the ambient medium,

$$M_w = M_j \left(\frac{A_j}{A_w} \right)^{1/2} \frac{1}{1 + \epsilon} \tag{3.3}$$

(NW85; LPMB88). A jet can be decelerated either by reducing its density or by increasing the size of the working surface. Using the Mach number rather than the velocity of the jet to parameterize the data insures that the primary variable controlling the rate of advance is the relative areas of the jet and the working surface.

We have no *a priori* way of computing the area of the working surface in the actual jets beyond $A_w \gtrsim A_j$. For filled jets, A_w is approximately given by the radii of the ring shock at the head of the jet (LPMB88). Since even for filled jets the ring shock is not a stationary feature of the flow, the rate of advance of the jet is not steady and depends on the details of the flow at any time.

In the hollow cylindrical jets, if the ratio of jet thickness to average radius is fairly large ($\Delta R / \bar{R} \sim 1$) the jet will form a composite bow shock from the axis to the exterior of the jet, whereas if the ratio is small, ($\Delta R / \bar{R} \ll 1$) the jet will form an annular bow shock. In either case, the total area of the effective working surface will be significantly larger than that of the equivalent filled cylindrical jet. The case we examine has $\Delta R / \bar{R} = 0.5$ so that we expect the momentum transfer to occur over a region from the axis to the outer edge of the jet, with $A_w \gtrsim 25 A_j / 16$ (because the outer edge of the jet is at $5 R_j / 4$). This implies that the rate of advance of the hollow cylindrical jet will be some 50% slower than the equivalent filled jet.

For the conical jets, geometric expansion increases the jet outer radius $R = R_0 + Z \tan \theta$ where R_0 is the initial outer radius of the jet and Z is the axial distance from the orifice. If the thickness of the jet remains roughly constant, the area of the jet will increase with R. One important consequence of this is that the input momentum flux will be spread over an *increasing* area. The Mach number of the working surface eventually becomes subsonic when the geometric dilution of the momentum flux prevents the jet from driving the ambient material out of its path. At this point, we can think of the jet as having "stalled." Neglecting the density ratio, this distance is simply the point where

$$A_w = M_j^2 A_j \qquad (3.4)$$

where M_j is the initial Mach number of the jet.

A crucial question is whether the area working surface grows proportional to R, corresponding to momentum transfer over an annulus of fixed width about the jet, or proportional to R^2, corresponding to momentum transfer over an annulus with a linearly increasing width. This must depend on how the area of the jet changes as it evolves and on the stability of the jet's channel. For the cylindrical geometries, both hollow and filled, the cross section should remain approximately constant; the jet may wiggle, waggle, and wobble, but it remains a coherent flow directed along the axis. How does the width of the conical jets vary with R? Does the thickness of the jet compensate for or enhance the geometric effects? It seems unlikely, in the absence of significant cooling, that the jet could become narrower, since lateral compressions will generate shocks in a supersonic flow, and hence compensating pressure. This would tend to make any substantial constriction a transient rather than a steady state effect. However, because our resolution across the conical jets is rather low (10 zones at the inlet) we probably cannot completely rule out such a jet constriction. For the jet to become progressively thicker, internal pressure would have to be generated by shocks faster than the pressure decrease from geometric expansion.

We turn then to a consideration of the internal jet dynamics as characterized by the changes in the jet cross sectional area A_j, average density ρ_j, and average Mach number M_j. These are related by the conservation of mass flux along the body of the jet,

$$\rho_j v_j A_j = constant \qquad (3.5)$$

and conservation of energy along flow lines

$$\frac{1}{2} \rho_j v_j^2 + \frac{\Gamma P_j}{\Gamma - 1} = constant \qquad (3.6)$$

We do not know what the evolution of the entropy in the jet will be, but the gas has only a few options. It can shock and increase its entropy, or it can evolve at constant entropy. From studies of filled jets we know that the gas shocks, and then expands adiabatically to return to pressure equilibrium. We can ask how the jets evolve in two limits: constant entropy and constant pressure. In these two cases the density evolves with Mach number as

$$
\begin{aligned}
\rho &= \rho_0 \left[\frac{(\Gamma - 1)M^2 + 2}{(\Gamma - 1)M_0^2 + 2} \right]^{-1/(\Gamma - 1)} \qquad S = constant \\
\rho &= \rho_0 \left[\frac{(\Gamma - 1)M^2 + 2}{(\Gamma - 1)M_0^2 + 2} \right] \qquad P = constant
\end{aligned}
\tag{3.7}
$$

A similar set of equations can be written to relate the cross-sectional area of the jet and the Mach number

$$
\begin{aligned}
A &= A_0 M_0 \left[\frac{(\Gamma - 1)M^2 + 2}{(\Gamma - 1)M_0^2 + 2} \right]^{-(\Gamma + 1)/2(\Gamma - 1)} \qquad S = constant \\
A &= A_0 M_0 \left[\frac{(\Gamma - 1)M_0^2 + 2}{(\Gamma - 1)M^2 + 2} \right]^{1/2} \qquad P = constant
\end{aligned}
\tag{3.8}
$$

The first of these assumptions describes the local post-shock behavior of the jet gas, while the second should more closely model the overall evolution of the jet. In the case of the filled jets, internal shocks gradually increase the entropy. One consequence is that the cocoon is hotter and lighter than the jet (NW85). The conical jets have the additional geometrical complication of a constantly increasing cross-sectional area. If the jet's entropy were constant as the area increased, the Mach number and jet density would drop. However, if the ambient medium has sufficient inertia the jet *cannot* freely expand. It will be shocked and more or less brought into pressure balance. If we assume that the conical jet evolves at a constant pressure and jet thickness ΔR, so that the cross-sectional area increases linearly with radius R, the internal Mach number of the jet decrease as $M_j(R) = M_0(\bar{R}_0/R)^{1/2}$. This is essentially the same condition as (3.4), except that here we are concerned with the conservation of momentum within the jet, and not with its transfer to the surrounding gas.

From this analysis we see that a conical jet will, in the absence of significant cooling, evolve towards lower Mach numbers. As its Mach number decreases, the jet will become unstable to local "kink" modes, developing ripples with wavelengths ranging from the jet thickness ΔR, to the distance from the axis R. These will locally resemble the instabilities of slab jets studied by NW85 so the criterion for the onset of the instability is probably similar ($M_j \lesssim 1 + \eta^{1/2}$) although we have not attempted to work out the exact condition (see also Hardee and Norman, 1988, and Norman and Hardee, 1988). Regardless, the jet will eventually become subsonic and stall, or disrupt as the jet Mach drops below the stability threshold.

The preceding discussion has assumed that the jet is surrounded by a constant pressure ambient medium. However, the properties of the medium in which the jet propagates can be just as complicated as those of the jet itself, significantly affecting jet evolution. Previous studies have found that the dynamics of filled jets are affected by self-interactions with a cocoon of spent gas. While the jet is injected in pressure equilibrium with the ambient medium, it propagates through a cocoon that not only has an unknown pressure, but is also in motion (*supersonic* motion in some regions). The initial data is, therefore, not completely consistent with the subsequent physical conditions at the inlet. If the jet is

injected into a higher pressure medium characterized by pressure $P > P_j$ then (assuming a fixed cross sectional area and pressures small compared to the jet ram pressure) the jet will quickly change its Mach number

$$M_j \longrightarrow M_j \left(\frac{P_j}{P} \right)^{1/2} \tag{3.9}$$

to come into pressure equilibrium with its surroundings. This makes it extremely difficult to inject high Mach number jets in a fully self-consistent manner—the jets will shock and reduce their internal Mach number to achieve pressure balance with the cocoon. This leads to the "inlet shocks" that extend from the edge of the jets at the inlet to a Mach disk approximately $M_j R_j$ down stream. This is not a problem with the lower Mach number, $M_j = 6$, jets, but does seem to complicate the evolution of the higher Mach number $M_j = 12$ jets. Typically the cocoon pressure is roughly twice the ambient pressure so that the jet Mach number can drop by nearly 50% at the inlet shock.

As with the jet, we can do some analysis of the cocoon through the use of conservation laws. Details are more difficult to obtain, since the cocoon is an amorphous structure, but some general relationships should hold. The cocoon is formed when the injected gas reaches the end of the jet and is decelerated by a strong shock. In fact the gas flows through a series of terminal shocks that increase the pressure and entropy. The pressure then equilibrates (to first approximation) through adiabatic PdV work. If we model this by passing the jet gas through a strong shock characterized by the Mach number of the jet, and then reduce the gas pressure to that of the ambient medium through adiabatic expansion, the resulting gas density is

$$\rho_f \simeq \alpha \rho_j M_j^{-2/\Gamma} \qquad \alpha \simeq \frac{\Gamma + 1}{\Gamma - 1} \left(\frac{2\Gamma}{\Gamma + 1} \right)^{-1/\Gamma} \tag{3.10}$$

where for a $\Gamma = 5/3$ gas, $\alpha \simeq 3.5$. This roughly matches the densities in the cocoon (to within a factor of two in either direction) away from the cores of supersonic vortices where the density can be a factor of ten smaller due to centrifugal effects. Note that for the conical jets there is a correction of $(\bar{R}_0/R)^{(\Gamma-1)/\Gamma}$ to the cocoon density under the assumption that the jet maintains a fixed width and constant pressure as it propagates from the injection radius at $\bar{R}_0$ to the terminal shock at radius R. This follows from equation (3.7).

We can combine the cocoon density and the jet mass flux to estimate the size of the cocoon assuming that the jet gas does not mix with the ambient gas (at least not initially). We must distinguish, however, between the two cases of an advancing jet, and a stalled jet. If we assume for an advancing jet that the cocoon is laid down by the working surface at rest in the "lab" frame, then, using (3.2), the equilibrium outer radius for the cocoon R_c is

$$\left(\frac{R_c}{R_j} \right)^2 \simeq 1 + \alpha^{-1} M_j^{2/\Gamma} \eta^{-1/2} \left(\frac{A_w}{A_j} \right)^{1/2} \tag{3.11}$$

where A_w is the area of the working surface. In practice, this should be an upper limit; when using outflow boundary conditions the cocoon flows back off the grid rather than remaining at rest. Further, the cocoon and the post-bow shock ambient gas will be at a higher pressure than the initial P_a. However, the functional dependence clearly shows the transition to a naked beam for dense jets, and the growth of the cocoon with Mach

number. Note that since the size of the cocoon depends on the rate of advance, and hence size of the working surface, a hollow cylindrical jet with its larger working surface generates a larger cocoon as compared with a filled jet of the same Mach number and density ratio.

If the jet is stalled rather than propagating, the cocoon inflates a vortex ring at the head of the jet. If we model such a vortex as a toroidal ring with its center at distance R from the axis and radius r_v, the mass of spent jet gas enclosed in the "vortex" is $2\pi^2 R r_v^2 \rho_f$. (Note that we have ignored the dynamics of the vortex.) The torus is inflated by the mass flux of the jet, $A_j \rho_j v_j$. If we allow the torus to grow with fixed R and time dependent r_v, then the growing torus will start to interfere with the flow of the jet when $r_v \sim R_j$. This defines a vortex formation time (in units of the isothermal sound crossing time t_i)

$$
\begin{aligned}
\frac{t_v}{t_i} &= 2\pi\alpha\Gamma^{-1/2}M_j^{-(\Gamma+2)/\Gamma}\eta^{1/2}\left(\frac{R}{R_j}\right)\left(\frac{r_v}{R_j}\right)^2 \\
&\simeq 0.3\,\eta^{1/2}\left(\frac{M_j}{6}\right)^{-2.2}\left(\frac{R}{R_j}\right)\left(\frac{r_v}{R_j}\right)^2
\end{aligned}
\tag{3.12}
$$

For the conical jets we must include the correction to the final density for the increase of jet volume with R, so that at the stalling radius $R \sim M_j R_j$

$$
\frac{t_v}{t_i} \simeq \eta^{1/2}\left(\frac{M_j}{6}\right)^{-1.6}\left(\frac{R}{M_j R_j}\right)^{0.6}\left(\frac{r_v}{R_j}\right)^2\left(\frac{\bar{R}}{R_j}\right)^{0.4}
\tag{3.13}
$$

where the latter two results assume a $\Gamma = 5/3$ equation of state. Hence jets in this Mach number regime periodically shed vortices on time scales of order $t \simeq t_i$. This basic idea of the time scale for vortex shedding applies to both filled and hollow jets. As a consequence, the filled jets advance steadily only when considered on time scales greater than t_i; on shorter time scales, the terminal shock system can slow as vortices are formed and shed at the head of the jet.

4. Numerical Results

We have carried out a program of numerical experiments that includes a baseline of filled cylindrical jets, a set of hollow cylindrical jets, and a set of hollow cones. The simulations are performed with jet internal Mach numbers $M_j = 6$ or 12 and jet to ambient medium density ratios of $\eta = 0.1$ and 1.0. Table 1 lists the numerical models that have been calculated for this investigation. Contour plots from late times in the simulations are displayed in Figures 2 through 5. The hydrodynamical behavior of the filled jets is discussed in §2. In this section we discuss the results for the other geometries.

4.1. Hollow Cylinders

We first consider the geometry of the hollow cylindrical jet. As described in §2 the area of the jet orifice is chosen to be the same as that of the filled jet so that both geometries begin with the same total momentum flux. From Figures 2 through 5 it is apparent that hollow cylindrical jets are, in some sense, similar to the filled jets. They are approximately stable, and propagate like extremely blunt filled jets. However, the hollow cylindrical jets show several qualitative departures from the behavior of the filled jets. There are three significant differences: (1) the working surface is blunter, (2) the cocoon is larger, and (3) the cocoon-jet interactions are more complex.

The first two of these differences can be understood by considering the working surface at the head of the hollow cylinder. Although the hollow cylinder's cross-sectional area is the same as the filled jet, the area of its working surface is significantly larger, extending at least from the jet's outer surface to the axis. This manifests itself as a lower rate for jet propagation (see Figures 6-7; eq. [3.2]). Relative to the similar filled jet, the $M_j = 6$, $\eta = 1.0$ hollow jet takes 30% longer to reach the edge of the grid, and the $M_j = 6$, $\eta = 0.1$ hollow jet takes 60% longer. Slower propagation means more spent jet gas is forced into the cocoon per unit length of the jet, resulting in a larger cocoon (*cf.,* equation [3.11]). Even the $M_j = 6$, $\eta = 1.0$ case, which is "naked beam" filled jet, develops a cocoon in the hollow geometry.

Since we observe larger cocoons in the hollow cylindrical geometry, it is not surprising that the cocoon-jet interactions are more complex than in the filled jet. One important feature is the "plug" of high pressure gas that forms at the the head of the jet. In the the filled jets the leading plug of shocked material causes the outward deflection of the jet along an oblique shock (LPMB88). In the hollow jets it is larger, and as a result the jet undergoes a dramatic outward "flaring" along the surface of the plug that creates large supersonic vortices in the cocoon. Initially the plug is composed of shocked ambient gas, but later it is material from weak, inwardly shed vortices drifting backwards into the jet. These vortices are suppressed by the small volume interior to the jet and by the forward motion of the plug. Except for the plug, the interior region is filled with shocked *ambient* gas, and the cocoon is restricted to the exterior of the jet.

The supersonic vortices in the cocoon of the filled jets trigger the formation of the now familiar crossed shock patterns in the jet. In the hollow cylindrical geometry, the jet is perturbed by both the vortices of the cocoon and by pressure fluctuations in the shocked ambient gas along the axis; in general, these two sources of perturbations will not act coherently. In our simulations, the cocoon is the dominant source of perturbations away from the head of the jet. This results in a partial focusing of the jet channel roughly midway between the inlet and the working surface of the jet where the largest vortex impinges on the surface of the jet channel. The material on the interior is compressed by the inward motion of the jet, eventually raising the pressure sufficiently to move the jet away from the axis. The competition between the two forces drives oscillations in the position of the jet's channel (Figure 9).

As a consequence of this complex interaction, the "simple" shock structure of the filled jets is lost (see Figure 8). Although there are a series of crossed shocks on the axis, these are formed in response to the pinching action of the jet on the interior gas. The jet itself does not have a well defined shock structure. The wall of the hollow cylindrical jet has no enforced geometric symmetry, and this permits more irregular behavior. In this sense a section of the hollow cylinder resembles a slab jet (NW85; Hardee and Norman, 1988, Norman and Hardee, 1988), which can develop large kink mode perturbations. Because the Mach number of the jet is sufficiently supersonic to avoid the normal criterion for the onset of instability ($M_j \lesssim 1 + \eta^{1/2}$), the strong perturbations can excite large "kink-like" oscillations without the jet becoming catastrophically unstable. If the Mach number of the jet is reduced to the transonic regime, the oscillations will rapidly grow and disrupt the jet.

4.2. Conical Jets

While the hollow cylindrical jets possess some superficial resemblance to the filled jets, the hollow conical jets have remarkably different properties. The most dramatic of

these is that **hollow conical jets do not propagate.** In each case the head of the jet reaches a limiting distance at which it stalls. For the Mach 12 models the $\eta = 1$ jet stalls at a distance of $14R_j$, and the $\eta = 0.1$ jet stalls at $12R_j$. For the Mach 6 jets these distances are $8R_j$ and $6R_j$. (See Figures 6 and 7.)

The geometric increase in both the interior volume of the cone and the cross-sectional area of the conical jet must account for this effect. We observe that the jet width, ΔR, does not change much over the length of the jet, so the cross sectional area of the jet increases proportional to R (Figure 11). The stalling radius seems to increase linearly with the initial Mach number, which suggests, by the analysis in §3, that the effective working surface grows as R^2. The effective working surface must extend over the entire region interior to the jet instead of being a thin annulus centered on the jet. Equation (3.8), which relates the jet's cross-sectional area to its Mach number, requires that the jet thickness must increase with R if the stalling distance is to be linear in the Mach number. As this does not agree with the observed behavior of the jet, the stalling mechanism apparently depends more on momentum transfer to the external gas than on the evolution of the jet's parameters.

We find a weak dependence of stalling radius on the density parameter η. The sense of this dependence is not consistent with the simple theory (eq. [3.2]); in our simulations lower density ratios stall at smaller distances. The result is consistent with the presence of higher pressure cocoons surrounding low η jets. Recall that the effective initial Mach number is reduced by the presence of a higher pressure cocoon (eq. [3.9]).

A second important difference between the conical and filled jets is that the conical jets are very unsteady flows. It is not possible to demonstrate fully the degree to which the conical jets depart from a steady flow pattern with a series of contour plots. We did not fully appreciate the nature of the flow until we generated a computer animated film of the propagation of conical jets, and the descriptions of the flow patterns in this section are mainly derived from observing these animated sequences.

When the conical jet is first injected into the grid, shock waves are driven into the interior of the cone. The bow shock directed towards the interior is focused and strengthened by the symmetry about the axis. Moreover, unlike the bow shock on the exterior of the cone, the pressure built up behind the shock wave can be reduced only by expanding along the axis, or by pushing the jet away from the axis. The shock-generated pressure gradient immediately deflects the jet out from the axis. After propagating several jet radii, this deflection has become large (nearly perpendicular to the axis at the head of the jet), and the head of the jet rolls off into large vortex. The accumulation of cocoon gas in these vortices and the dynamical pressure from the still supersonic cocoon gas impinging on the outer surface of the jet provides compensating pressure that gradually pushes the jet back towards the axis. This in turn recompresses the gas interior to the jet, raising the pressure to the point where the jet is again driven outwards.

In effect, the jet acts as a "wall" separating the interior and exterior gas. Pressure balance is maintained only by deflecting this jet wall. The ambient exterior gas is able to expand away from the jet towards the grid boundary, and is thus not able to provide the needed confining pressure. This role falls to the cocoon gas, and the time-dependent dynamics of the hollow jet are dominated by perturbations in the cocoon, particularly by the action of the vortices generated at the head of the jet. As for the shocks generated in the interior of the cone, they are directed out along the axis by the jet wall. These shocks catch up with the bow shock and help drive it outwards along the axis. The result, however, is an extremely flat bow shock that has no resemblance to the "ears" of W50.

These changes in the pressure gradients, both inside and outside the jet, lead to continual vortex production at the head of the jet, and cause the flow to be extremely nonsteady (Fig. 11). The dramatic radial oscillations cause the jet's momentum flux to be "sprayed" over the entire region interior to the precession cone. Thus, these oscillations contribute to the stalling process by insuring that the jet momentum is distributed over a substantially larger area than the thin shell about the precession cone. This accounts for the R^2 growth in the effective working surface. Similar radial oscillations are observed in the cylindrical jets, but they are larger for the conical jet as the geometry makes the restoring forces weaker.

The vortices are usually created with the sign natural to the side of the jet on which they are formed. The time scale for their formation and shedding is roughly in accord with the simple arguments of §3. The vortices frequently interfere with the working surface either by becoming so large that they "smother" the head of the jet, or by propagating into the path of the the jet. Large vortices do not persist inside the cone. Once outside the jet, the vortices slowly drift backwards towards the upstream boundary. Occasionally the jet is able to "tunnel" through a vortex at its head and send a brief pulse of the highly supersonic jet gas through the cocoon. When this gas reaches the contact discontinuity separating the cocoon from the shocked ambient gas it drives a shock wave outwards through the medium. This leads to a quasi-periodic system of shock waves traveling outwards along the axis behind the bow shock, and the slow forward motion of the contact discontinuity seen in Figures 6-7.

The internal properties of the jet are monitored through the use of tracer particles. Along the jet, the Mach number decreases rapidly both due to shocks and the geometric dilution of the jet. Since the velocity remains roughly constant, the jet density drops linearly with radius. The Mach number drops faster than linearly due to shock waves that keep the jet pressure more or less in equilibrium with its surroundings. We know that if the jet Mach number drops far enough the jet becomes unstable to kinking modes, and the large oscillations of the jet provide an abundance of perturbations to drive the instability. It is difficult, however, to determine whether the jet Mach number drops sufficiently far while the jet can be characterized as a coherent flow structure. The jet disrupts into a series of vortices due to the large oscillations in the radial position of the head of the jet, so that while the jet may not become unstable in a formal sense, it is difficult to resist characterizing the behavior as such.

The hollow conical jet neither focuses nor collimates. The large oscillations mean that it wanders considerably from the initial precession cone, but it does so without any directional preference. The external ambient medium plays almost no role in the dynamics of the jet; the cocoon is the primary influence on the exterior of the jet, just as it is for the filled and hollow cylindrical geometries. The Eichler (1983) focusing mechanism does not seem to be applicable to these flows. In the interior region the ambient medium acts essentially as a "shock absorber." While there is mass flow out of the interior region, the interior pressure never (on average) drops significantly. A lower interior density reduces the inertia of the gas, which makes it easier for the jet to move inwards, but it is the pressure that ultimately determines how closely the jet can approach the axis. As long as the pressure forces can become sufficiently large, the jet will be pushed back towards the precession cone, preventing any permanent focusing.

5. Consequences for SS433

The implication of the preceding sections is that hollow conical jets propagate ineffi-

ciently, if at all, because of the geometric dilution and spreading of the working surface. In addition, the jets do not focus; hydrodynamic interactions maintain significant average pressure in the interior of the cone. The simulations, however, dealt with a series of parametrized variables. In this section we ask whether or not the hydrodynamical jet can in any way be suitable as a model of the SS433 jet system.

Suppose we attempt to match the hydrodynamic simulation into the parameters of SS433 rather than simply as a series of dimensionless numbers. This is not meant to be a serious attempt to "model" SS433, but only to try to connect the results to a realistic system. To this end, we scale the jet luminosity and velocity to the approximate values for SS433. The jet kinetic luminosity is taken to be $L_{40} = L_j/10^{40}$ ergs s^{-1} based on the characteristics of the emission lines from SS433 and on the energy required to create the "ears" of W50 (Begelman et al., 1980, Königl, 1983). The jet velocity is $v_{SS433} = v_j/0.26c$. If we place the outer edge of the grid at a distance of $l = 50$ pc from the source of the jet, and assume the jet propagates on the 20 degree half angle precession cone before it reaches the left edge of the grid, we find that each zone has a width of ~ 0.1 pc. Recall that this is approximately twice the distance between successive windings of the jet. The left edge of the grid is ~ 5 pc from the origin, and the jet is injected at $\bar{R} = (R_{out} + R_{in})/2 = 1.7$ pc with width $\Delta R = (R_{out} - R_{in}) = 0.85$ pc. These values for the parameters imply that the jet density at the orifice is

$$\rho_j \sim 3 \times 10^{-4} m_p \text{ cm}^{-3} L_{40} v_{SS433}^{-3} \left(\frac{\bar{R}}{2\Delta R} \right), \tag{5.1}$$

the jet pressure is

$$P_j \sim 2 \times 10^{-11} \text{ dynes cm}^{-2} L_{40} v_{SS433}^{-1} \left(\frac{\bar{R}}{2\Delta R} \right) \left(\frac{M_j}{30} \right)^{-2} \tag{5.2}$$

and the jet temperature is

$$T_j \sim 4 \times 10^8 \text{ K } \mu \left(\frac{M_j}{30} \right)^{-2} v_{SS433}^2 \tag{5.3}$$

where $\rho = \mu m_p n$ relates mass and number densities. The chosen Mach number of 30 makes the jet in pressure equilibrium with an ambient pressure of 2×10^{-11} dynes. The conditions inside of W50 are not known, but we can derive a few crude order of magnitude estimates with which to work. The edge of W50 is quite sharp, presumably corresponding to a strong shock. The pressure inside W50 should then be at least several times the ambient ISM pressure; a value on the order of 10^{-11} dynes is not unreasonable. Further, this pressure gives a total internal energy of $\sim 10^{51}$ ergs, roughly consistent with the observed energetics of the W50. The ambient density and temperature are limited by the lack of observed X-ray emission in W50 (Watson et al., 1983). Adopting the $T_6 = T_a/10^6$ K for the temperature of the ambient medium, we obtain a density ratio of

$$\eta = \rho_j/\rho_a \sim 2.5 \times 10^{-3} T_6 \left(\frac{M_j}{30} \right)^2 v_{SS433}^{-2} \tag{5.4}$$

at the jet inlet. The isothermal sound crossing time for this set of parameters is 10^3 years. The choice of a Mach 30 jet means that, by equation (3.4), the stalling radius is approximately 50 pc. However, because the density ratio is so small the jet is well into

the cocoon dominated regime of jet propagation. Extrapolating the behavior of such a jet from the simulations we have run poses a problem of self-consistency. (Actually carrying out the specific simulation would require an enormous expenditure of computer time that would be difficult to justify for this simplistic model.) As the jet surrounds itself with spent gas, the medium through which it propagates changes. The ambient medium will play only a minor initial role in the jet evolution. If one proposes that all of W50 is a large inflated cocoon from the hollow jet, the temperature in W50 would be on order 10^{10}K, from the thermal energy associated with the large jet velocity. The cocoon density would be $\sim 10^{-5} m_p$ cm^{-3} for $P \sim 10^{-11}$ dynes (consistent with eqn [3.11] for a Mach 30 jet). Such a gas would not produce the observed X-ray and optical emission in W50; this would have to be due to entrained ISM. The present hollow jet would be propagating essentially through its own spent cocoon with density ratio on order unity.

Can a purely hydrodynamical conical jet account for the "ears" of W50? In the absence of a significantly asymmetric background pressure or density distribution, a protrusion such as the ears requires the directed deposition of energy. A jet provides such direction in principle, but that directionality will be lost at large distances from a *stalled* jet. We must therefore require that the jet stall on a scale comparable to W50, and that it focus or collimate sufficiently so as to direct its momentum and energy flux primarily in the direction of the ears. If we consider only the morphological resemblance of the simulations to that of SS433 we immediately find several differences. First, without significant focusing, the conical jets do not form any structure reminiscent of the ears of W50. Although the jet can oscillate inwards to much smaller radii than that of the precession cone, this inevitably results in an increase of the pressure in the interior, driving the jet out away from the axis. The cocoon is preferentially inflated outside the precession cone leaving the interior filled with shocked ambient material. These interior shocks are driven out along the axis, but they produce an extremely flat bow shock (relative to the filled jets) rather than a protrusion along the axis. The periodic compression of material inside the precession cone could serve as a natural source for the diffuse X-ray emission observed by Watson *et al.* (1983). However, in the final analysis the failure of the conical jets to propagate, and the morphological differences between W50 and any structure observed in the simulations suggests that a hydrodynamical model without additional physics may be inappropriate for the propagation of the SS433 jet on the largest scales.

Is SS433 a purely hydrodynamic jet? As discussed in §1 there is no clear answer to these questions based on the current state of the observations. Our simulations indicate, however, that the simple hydrodynamic jet we have modeled does not resemble the W50 system. If our simple hydrodynamic model is untenable, what are the alternatives? We can categorize several options:

(1) the jet is focused hydrodynamically, but additional physical processes or different initial data must be used to modify the dynamics,

(2) jet focusing or collimation is achieved by other than hydrodynamic means,

(3) despite appearances the jet is not *directly* responsible for the 'ears', or

(4) the jet we see today has changed from the jet that generated the ears of W50.

The first category can be divided into adding new physical processes, and changing the assumptions of our model. If the goal is to focus the jet, the additional physics should cool the gas inside the precession cone. The aim must be to reduce the interior pressure sufficiently rapidly to allow a pressure gradient to focus the jet as in the Eichler (1983) model. Regardless of the physical mechanism proposed to accomplish this end, care must be taken that the energy lost is accounted for. For example, the observed luminosity of

the area inside the precession cone $L \lesssim 10^{36}$ ergs s^{-1} strongly limits the radiative cooling times. The internal energy inside the precession cone of our simple parameterized jet is

$$E_{internal} \sim 10^{49} \left(\frac{P}{10^{-11} \text{ dynes cm}^{-2}} \right) \left(\frac{l}{50 \text{ pc}} \right)^3, \qquad (5.5)$$

which limits the radiative time scale to $t_R \gtrsim E/L \simeq 5 \times 10^5$ years. This is much longer than the dynamical time of the jet ($\sim 10^3$ years), which is the rate at which shocks reheat the gas. This suggests that radiative processes cannot save our hollow jet. Other cooling mechanisms may be similarly constrained.

Another approach is to ask what initial conditions or assumptions in the simulations can be changed without adding additional physics. It is possible that the assumption of axisymmetry, which forces the axial focusing of shocks, significantly distorts the results. If the shocks were not coherently focused at the axis, the interior pressure would not be as effective in preventing focusing. The cone would not oscillate in phase; compression of the interior on one side of the cone could be compensated for by expansion on the other side. It is not clear whether this would result in focusing or in even more chaotic motion of the jet channel. This idea could be studied in a manner analogous to the studies of slab jets by looking at the propagation of two slab jets injected at an angle to each other (*cf.*, Norman and Hardee, 1988). Alternatively, the interior pressure might be reduced by altering the steady jet injection to a periodic injection. If the jet has 'holes' in it that allow the interior gas to leak through, the interior pressure can be reduced. Moreover the interior gas may be able to leak through the jet without significantly perturbing the jet; this would help to damp the radial oscillations that lead to the R^2 dilution of the jet momentum flux density. Considering that the observations of the jet show that it has a "blob-like" structure on the smallest scales, this effect must be present in the real jet to some degree. This might be studied by using a "two-phase" jet model in which most of the momentum and energy is carried in pulses of dense, cool gas and the interpulse part of the jet is filled with tenuous material.

In the second category, we mention the possibility of a magnetically dominated jet. Such a jet might be able to collimate in a manner reminiscent of the magnetically confined model for overpressured filled jets. Direct numerical simulations (LPMB88,Norman *private communication*) have succeeded in propagating such overpressured jets. An added difficulty for the SS433 jet is that magnetic hoop stress must overcome the large transverse ram pressure of the jet rather than the thermal pressure forces present in the filled jet simulations. Magnetohydrodynamic simulations of hollow, conical jets are needed to explore this idea.

Models in the third category are hard to justify because of the near-perfect alignment of the jet axis and W50. However, Katz (1986) mentions such a model in which the ears are created by radiation flux from the central object, rather than from the hydrodynamic jet. The greatest radiation flux from a thick accretion disk would naturally be aligned with the jet.

A related possibility is that the jet we see today is not the jet that generated the ears. The jet's precession cone may be slowly widening; the ears would have formed when the cone was much narrower. In any case, removing a causal connection between the ears at the current epoch has the virtue of explaining why there is little evidence for energetic jet interactions with the W50 remnant (apart from the ears!) It avoids the discrepancy between the large implied jet kinetic luminosity $L_j \sim 10^{40}$ ergs s^{-1}, and the observed luminosity $\sim 10^{36}$ ergs s^{-1}.

To conclude, it is hardly surprising that the SS433 jet is more complicated than an axisymmetric, conical, hydrodynamic jet. Our aim in this paper has been to draw attention to that fact. As there clearly exist a large number of possible modifications to our simple model, the hydrodynamic jet paradigm cannot be considered dead; it is, however, more tightly constrained. Considering the powerful effects of the "turbulent" gas dynamics on the evolution of the jet, it is unlikely that a model of the SS433 jet is complete or robust in its conclusions without including the full multidimensional, non-linear effects of supersonic hydrodynamics.

Acknowledgements

References

Begelman, M.C., Sarazin, C.L., Hatchett, S.P., McKee, C.F., and Arons, J., 1980, *Astrophys. J.*, **238**, 722.

Davidson, K., and McCray, R., 1980, *Astrophys. J.*, **241**, 1082.

Downes, A. J. B., Pauls, T., and Salter, C. J., 1986, *Mon. Not. R. astr. Soc.*, **218**, 393.

Eichler, D., 1983, *Astrophys. J.*, **272**, 48.

Hardee, P.E., and Norman, M.L., 1988, *Astrophys. J., in press.*

Hawley, J. F., Smarr, L. L., Wilson, J. R., 1984, *Astrophys. J. Suppl.*, **55**, 211.

Hjellming, R., Johnston, K., 1982, in *Extragalactic Radio Sources, IAU Symp. No. 97*, ed. Heeschen, D. and Wade, C. M. (Dordrect: Reidel), p. 197.

Hjellming, R., Johnston, K., 1985, in *Radio Stars*, ed. R. M. Hjellming, and D. M. Gibson (Dordrect: Reidel), p. 309

Hjellming, R., Johnston, K., 1988, *Astrophys. J.*, **328**, 600.

Katz, J.I., 1986, *Comments Astrophys.*, **11**, 201.

Kirshner, R.P., and Chevalier, R.A., 1980, *Astrophys. J. (Letters)*, **242**, L77.

Königl, A., 1983, *Mon. Not. R. astr. Soc.*, **205**, 471.

Kössl, D., and Müller, E., 1988, *Max Planck Institut preprint.*

Lind, K. R., Payne, D. G., Meier, D. L., Blandford, R. D., 1988, *Caltech preprint.*

Margon, B., 1984, *Ann. Rev. Astr. Astrophys.*, **22**, 507.

Milgrom, M., Anderson, S. F., Margon, B., 1982, *Astrophys. J.*, **256**, 222.

Norman, M. L., Smarr, L. L., Winkler, K.-H. A, and Smith, M. D., 1982, *Astr. Astrophys.*, **113**, 285.

Norman, M. L., Winkler, K.-H. A., Smarr, L. L., 1983, in *Astrophysical Jets*, ed. A. Ferrari and A.G. Pacholczyk (D. Reidel), p. 227.

Norman, M. L., Winkler, K.-H. A., Smarr, L. L., 1984, in *Physics of Energy Transport in Extragalactic Radio Sources*, ed. A. Bridle and J. Eilek (Green Bank: NRAO), p. 150.

Norman, M. L., Winkler, K.-H. A., 1985, *Los Alamos Science*, No. 12, p. 38.

Norman, M.L., and Hardee, P.E., 1988, *Astrophys. J., in press.*

Vermeulen, R. C., Schilizzi, R. T., Icke, V., Fejes, I., Spencer, R. E., 1987, *Nature*, **328**, 309.

Watson, M. G., Willingale, R., Grindlay, J. E., and Seward, F. D., 1983, *Astrophys. J.*, **273**, 688.

Zealey, W.J., Dopita, M.A., and Malin, D.F., 1980, *Mon. Not. R. astr. Soc.*, **192**, 731.

TABLE 1

Models Studied in the Parameter Survey

Jet Mach Number	Density Ratio	Jet Geometry
6.0	0.1	Filled Cylinder
6.0	0.1	Hollow Cylinder
6.0	0.1	Hollow Cone
12.0	0.1	Hollow Cone
6.0	1.0	Filled Cylinder
6.0	1.0	Hollow Cylinder
6.0	1.0	Hollow Cone
12.0	1.0	Hollow Cone
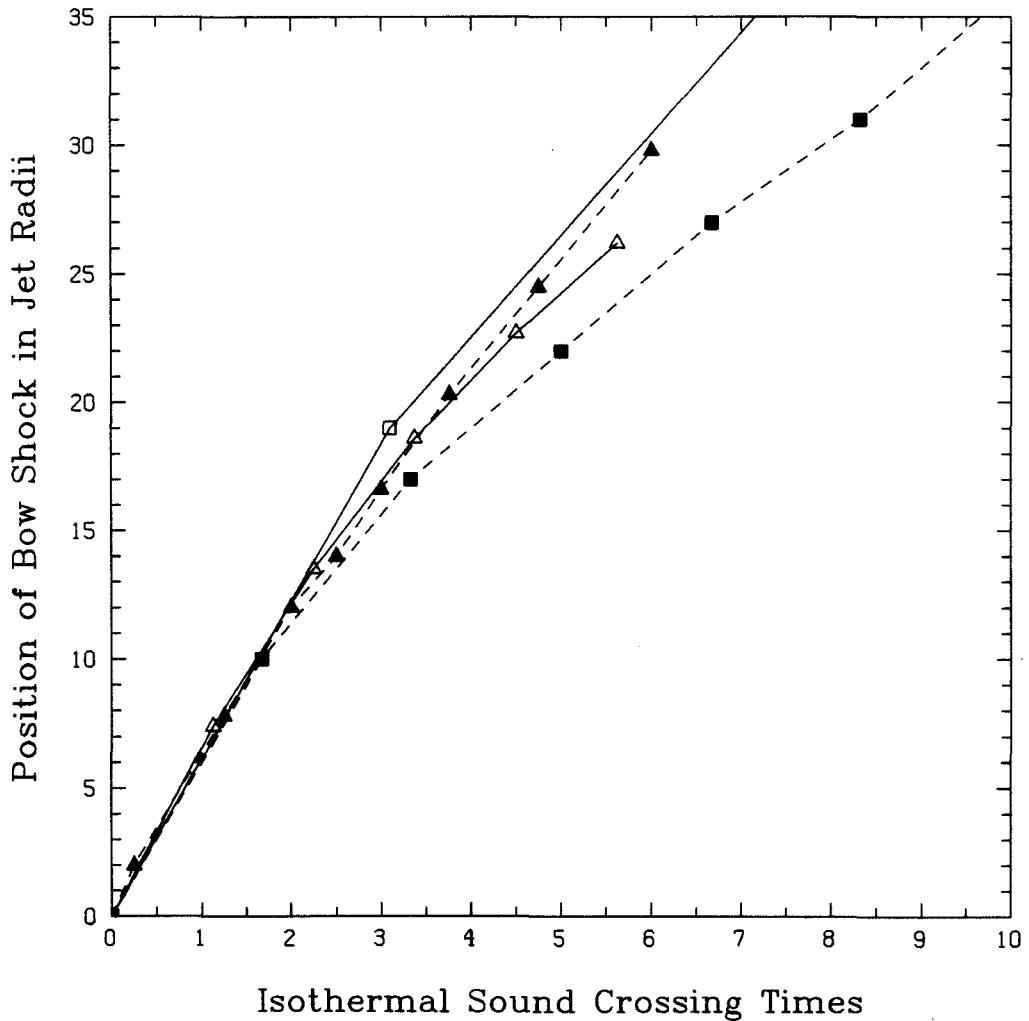
Figure 1. Position of bow shock for four different jet calculations. NSW85 $M_j = 6$, $\eta = 0.1$, 15 grid zones across jet, solid line and open squares LPMB88 $M_j = 6.1$, $\eta = 0.1$, 15 grid zones across jet, dashed line and solid squares KM88 $M_j = 6.0$, $\eta = 0.1$, 40 grid zones across jet, dashed line and solid triangles this work $M_j = 6.0$, $\eta = 0.1$, 20 grid zones across jet, solid line and open triangles.
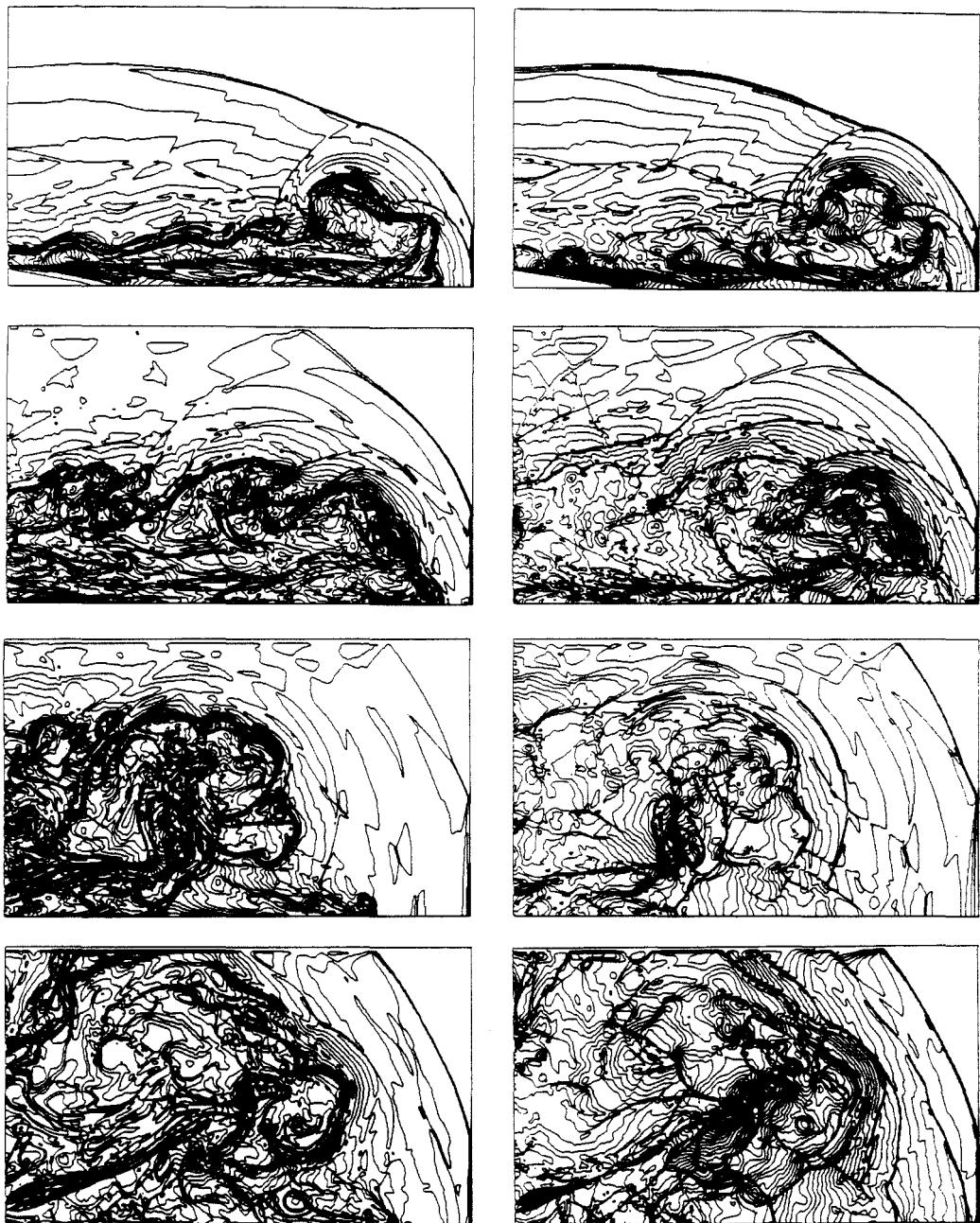
Figure 2. Density and pressure contours of the density contrast $\eta = 0.1$ jets examined in this study at their final time level where the bow shock is near or slightly past the downstream boundary. Thirty contours distributed uniformly in the logarithm of the density and pressure are displayed. On the right is the density, and on the left the pressure. The cases are (from top to bottom) the filled, cylindrical, and conical $M_j = 6$ jets, and the conical $M_j = 12$ case. The time levels of the contours are $t/t_i = 5.6$, 9.0, 11.3, and 6.8 respectively.
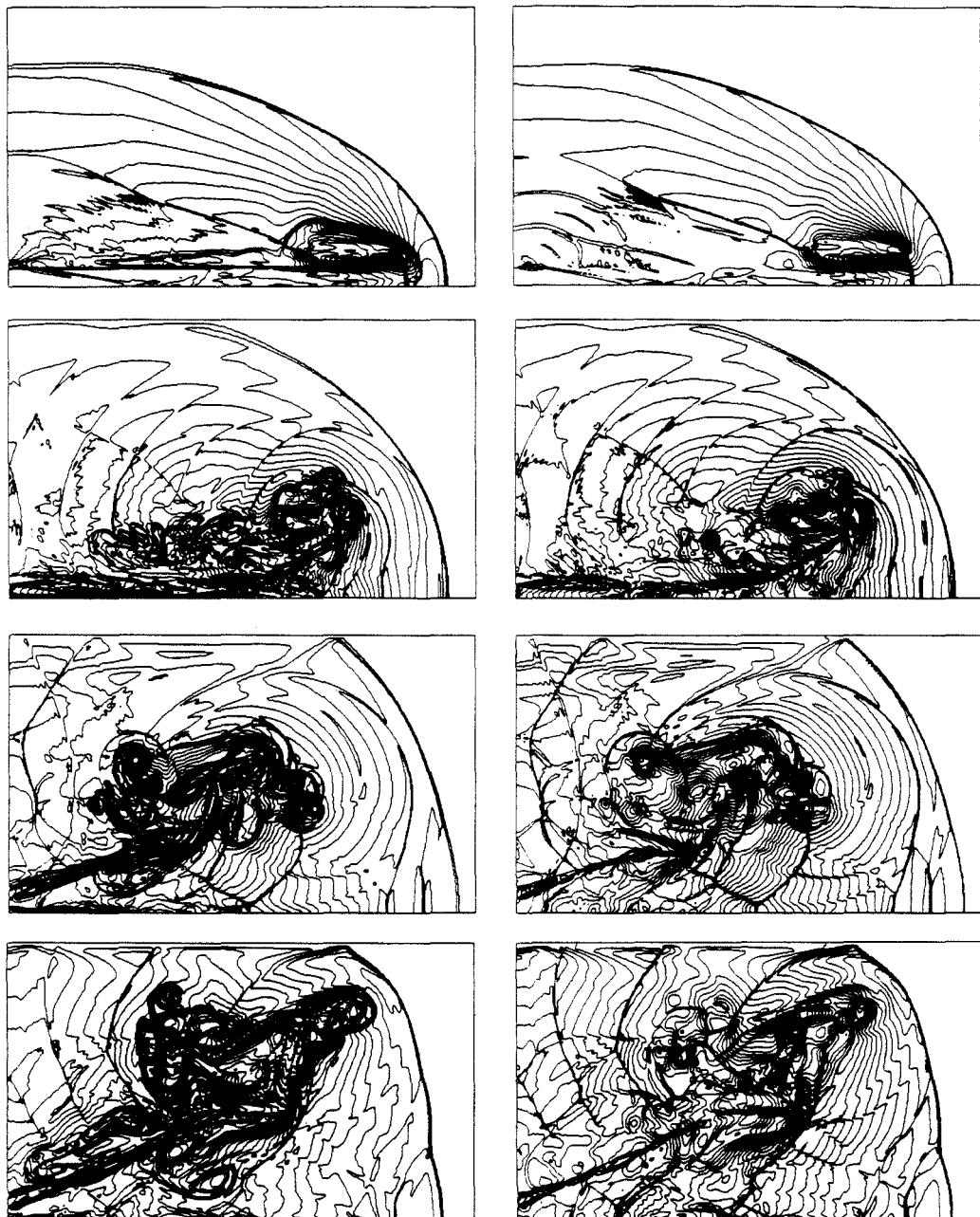
Figure 3. Density and pressure contours of the density contrast $\eta = 1.0$ jets examined in this study at their final time level where the bow shock is near or slightly past the downstream boundary. Thirty contours distributed uniformly in the logarithm of the density and pressure are displayed. On the right is the density, and on the left the pressure. The cases are (from top to bottom) the filled, cylindrical, and conical $M_j = 6$ jets, and the conical $M_j = 12$ case. The time levels of the contours are $t/t_i = 6.8$, 9.0, 11.3, and 6.8 respectively.
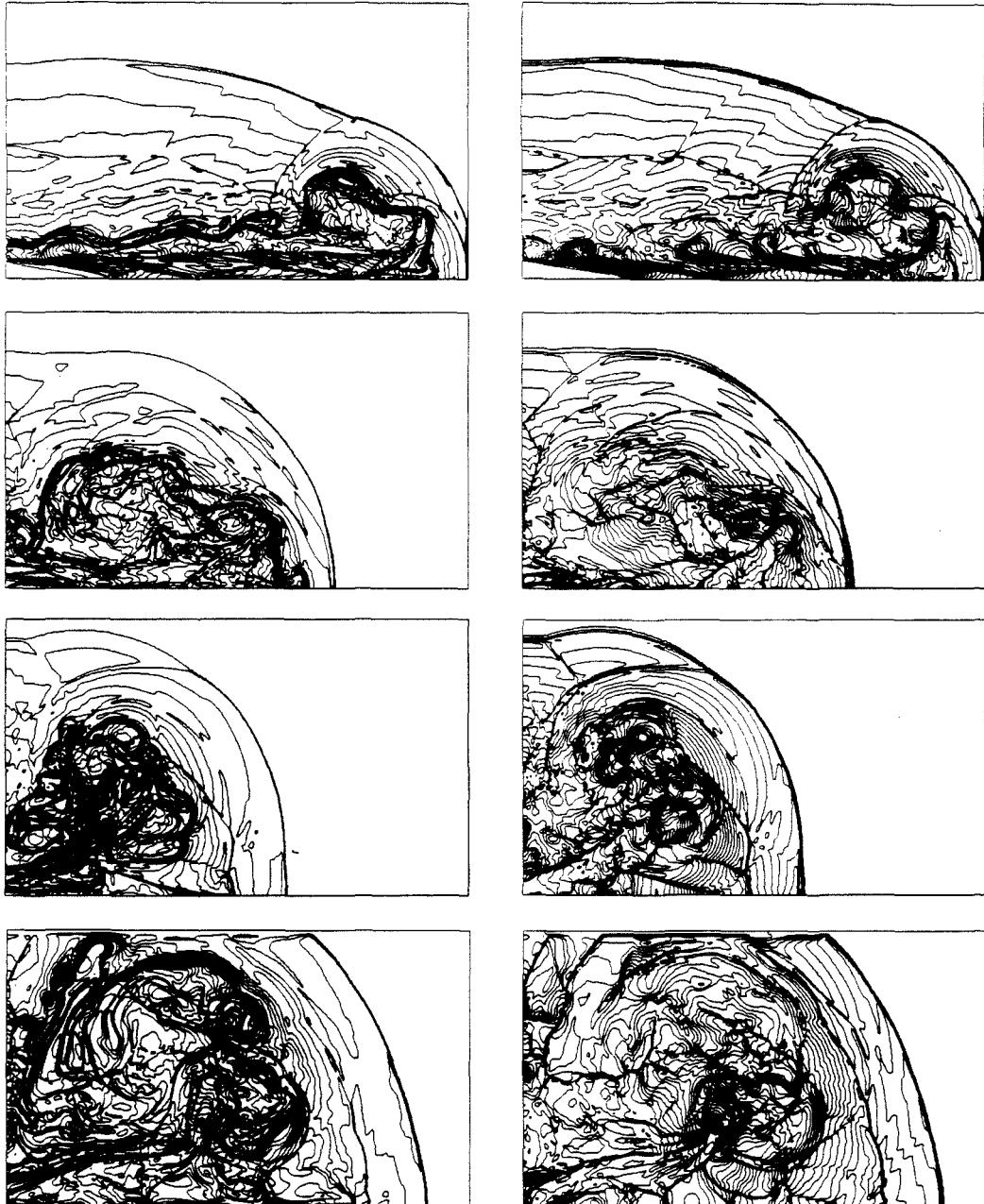
Figure 4. Density and pressure contours of the density contrast $\eta = 0.1$ jets examined in this study at the same time level $t/t_i = 5.6$. Thirty contours distributed uniformly in the logarithm of the density and pressure are displayed. On the right is the density, and on the left the pressure. The cases are (from top to bottom) the filled, cylindrical, and conical $M_j = 6$ jets, and the conical $M_j = 12$ case.
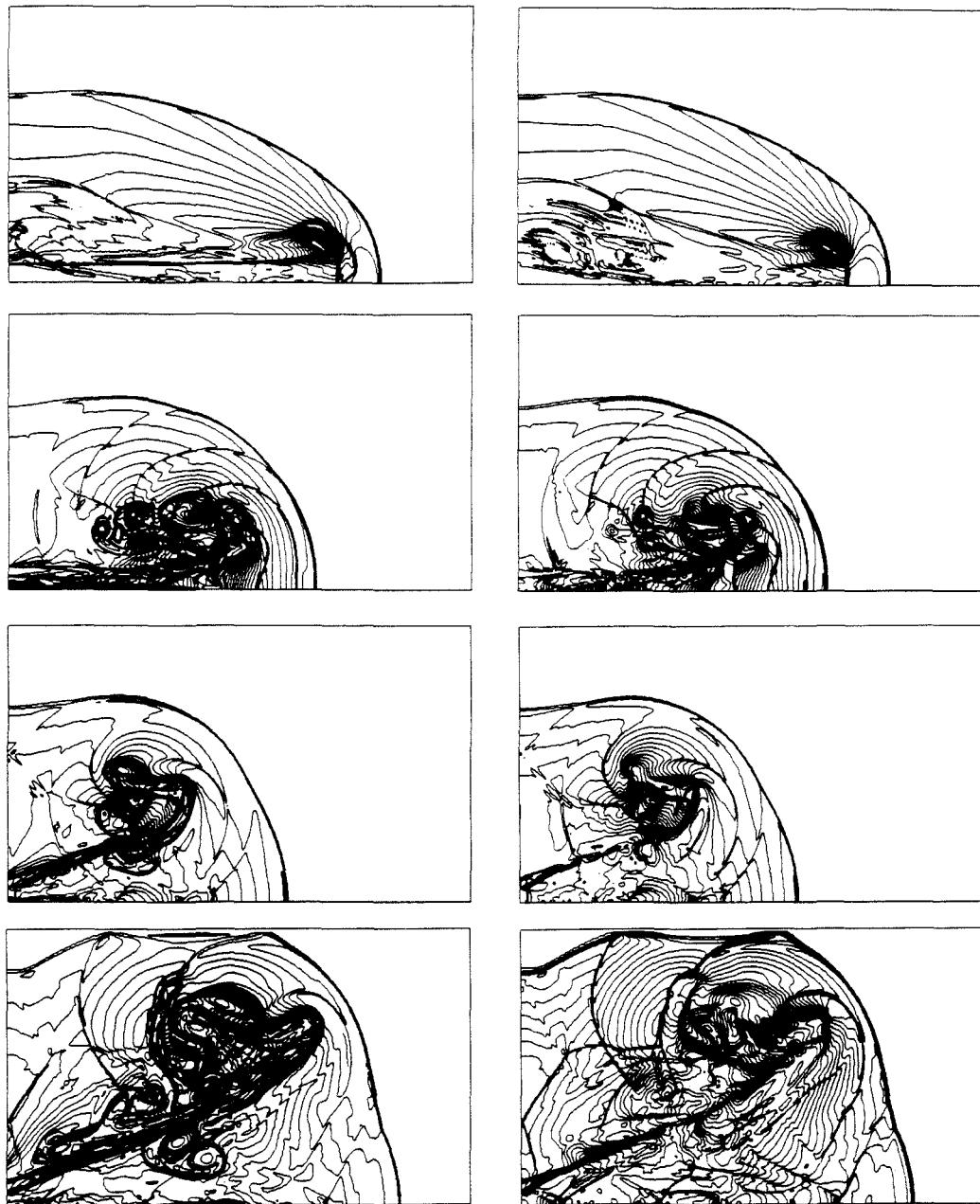
Figure 5. Density and pressure contours of the density contrast $\eta = 1.0$ jets examined in this study at the same time level $t/t_i = 6.8$. Thirty contours distributed uniformly in the logarithm of the density and pressure are displayed. On the right is the density, and on the left the pressure. The cases are (from top to bottom) the filled, cylindrical, and conical $M_j = 6$ jets, and the conical $M_j = 12$ case.
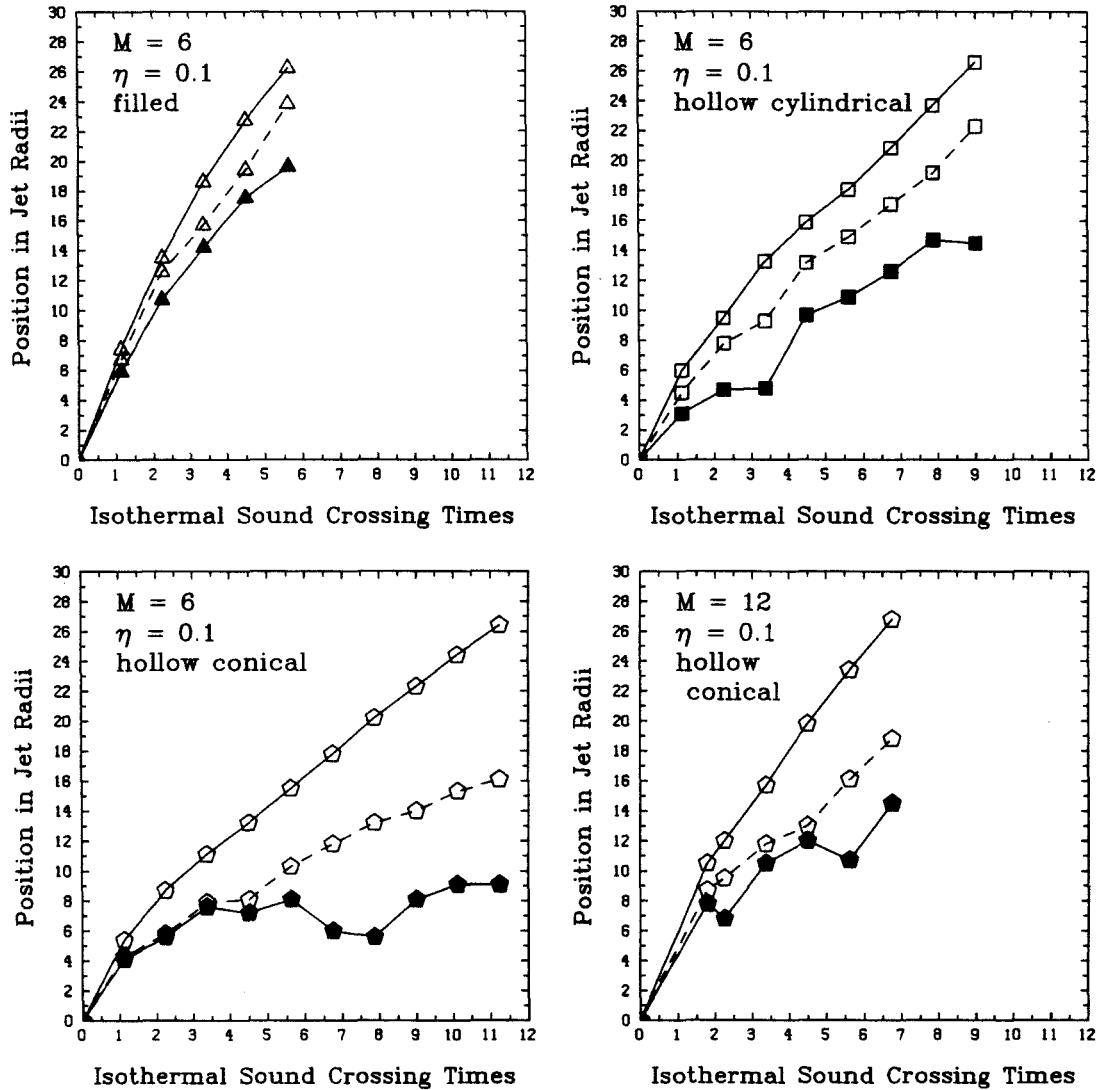
Figure 6. Position of the bow shock (open symbol and solid line) contact discontinuity (open symbol and dashed line), and terminal shock (filled symbol and solid line) for the four cases with density ratio $\eta = 0.1$.
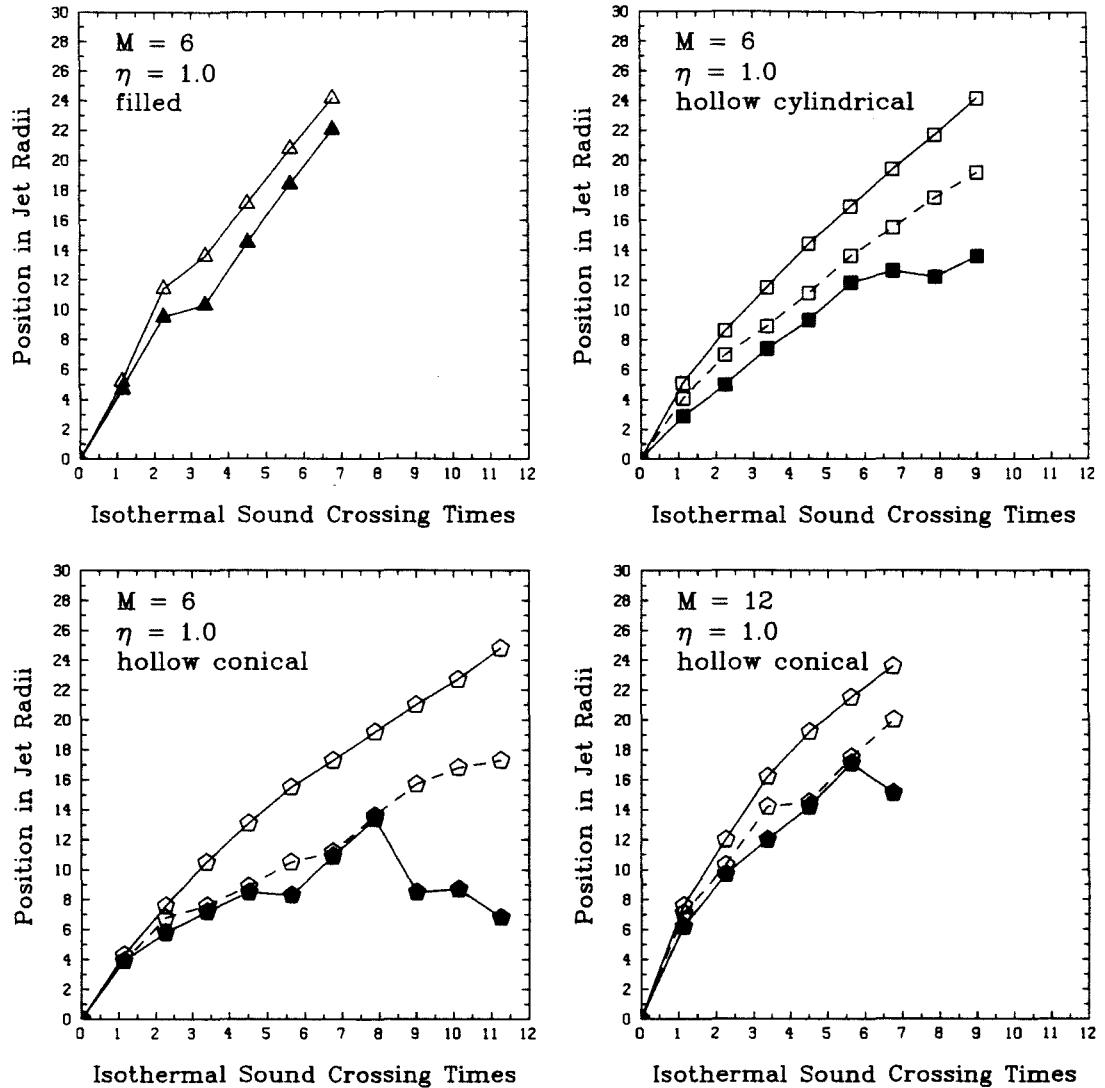
Figure 7. Position of the bow shock (open symbol and solid line) contact discontinuity (open symbol and dashed line), and terminal shock (filled symbol and solid line) for the four cases with density ratio $\eta = 1.0$. Note that in the cylindrical $M_j = 6$ case, the contact discontinuity and the terminal shock are essentially coincident.
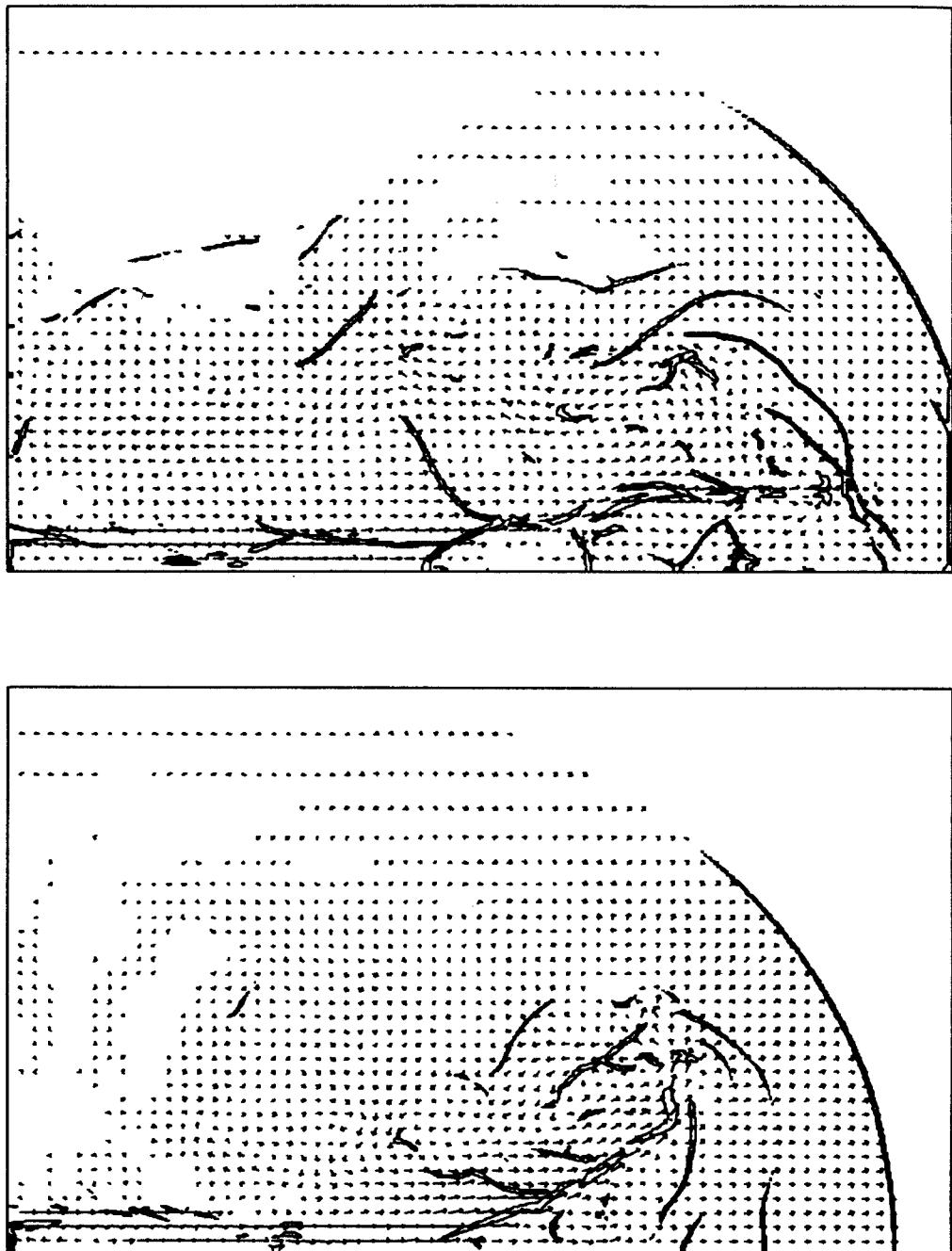
Figure 8. Velocity and shock structure for the two hollow cylindrical cases at the final time slice, $t/t_i = 9.0$. The $M_j = 6$, $\eta = 0.1$ case is shown in (a), and the $M_j = 6$, $\eta = 1.0$ case is shown in (b). Vectors indicate direction and magnitude of the fluid velocity, with the length proportional to the square root of the velocity. Only every eighth velocity vector is shown in each direction. Contours are contours of large negative div · v indicating the position of shocks.
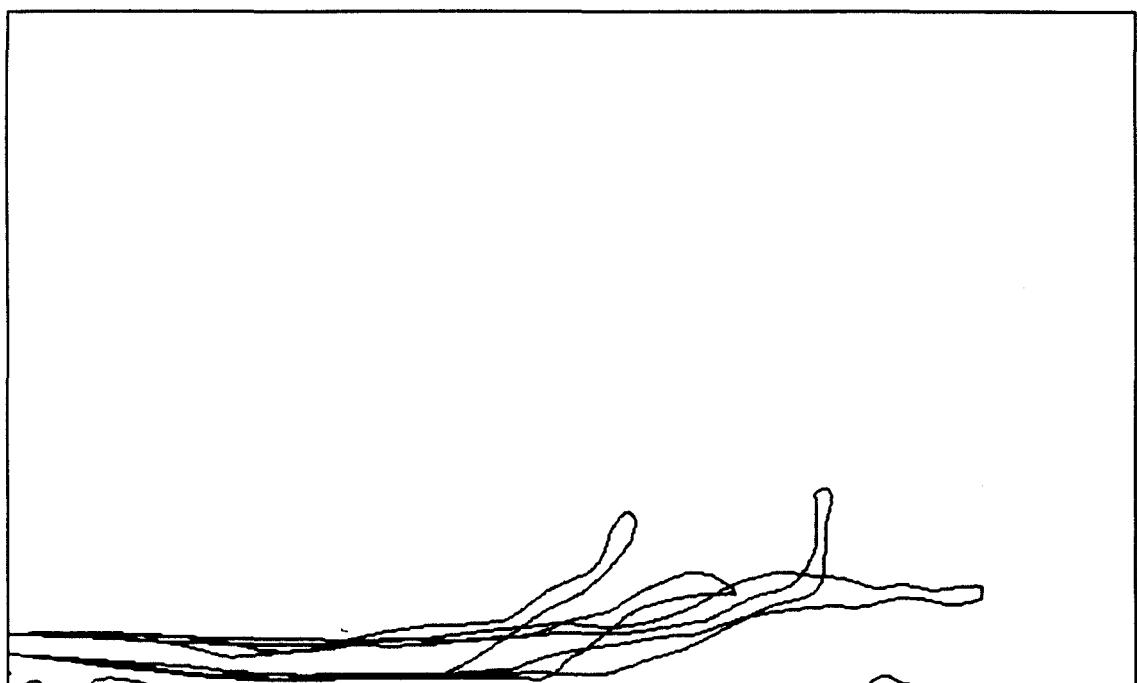
Figure 9. The position of the jet channel for the $M_j = 6$ $\eta = 0.1$ hollow cylindrical jet showing its varying position with time. The channel is defined by the region in which the total gas velocity is greater than 60% of the initial jet velocity. Contours are shown at $t/t_i = 5.6, 6.8, 7.9$ and 9.0. The inward pinching of the jet channel is clearly visible.
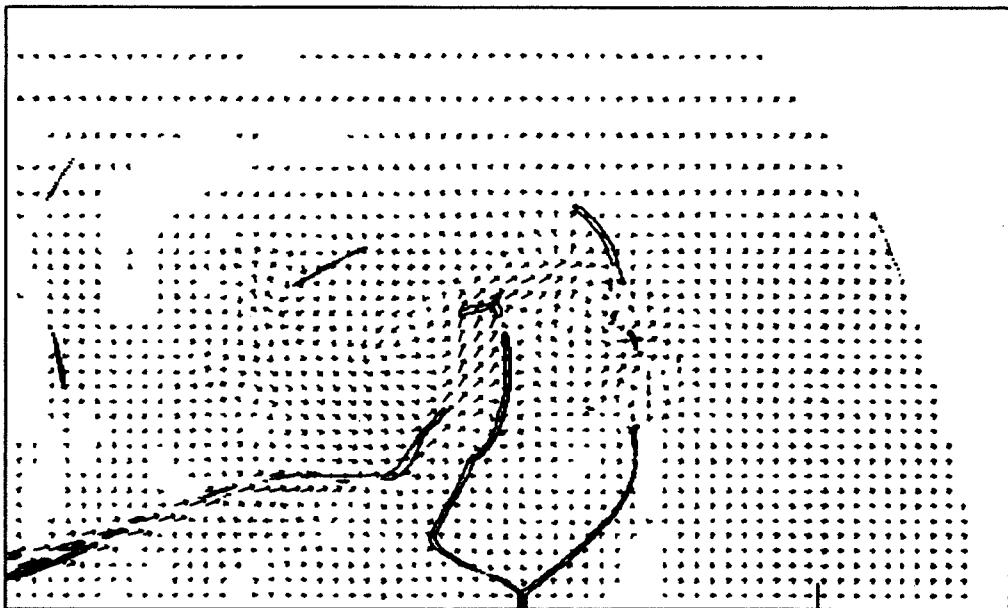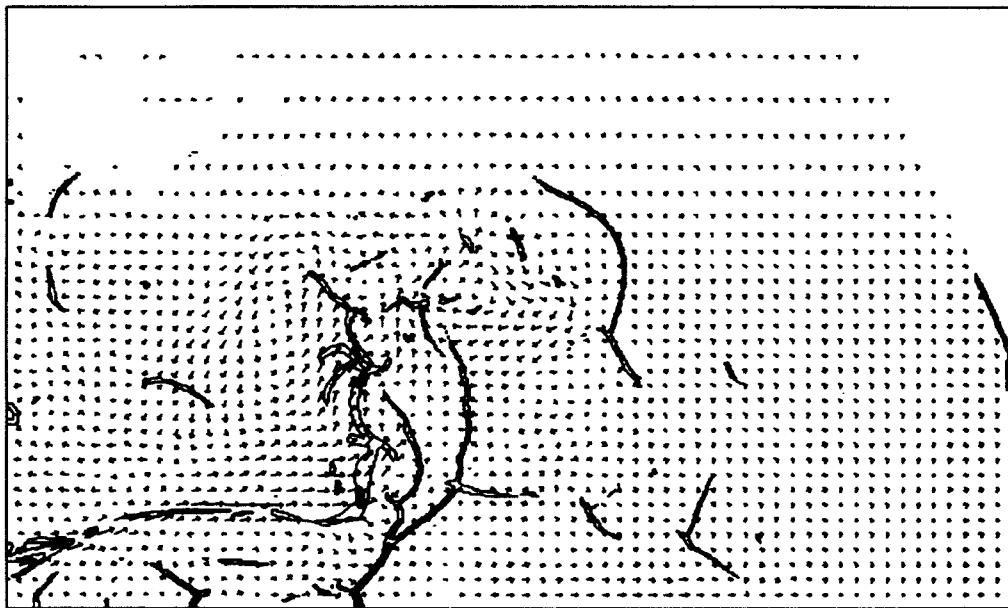
Figure 10. Velocity and shock structure for the two $M_j = 6$ hollow conical jets at the final time slice, $t/t_i = 11.3$. The $M_j = 6$, $\eta = 0.1$ case is shown in (a), and the $M_j = 6$, $\eta = 1.0$ case is shown in (b). Vectors indicate direction and magnitude of the fluid velocity, with the length proportional to the square root of the velocity. Only every eighth velocity vector is shown in each direction. Contours are contours of large negative div · v indicating the position of shocks.
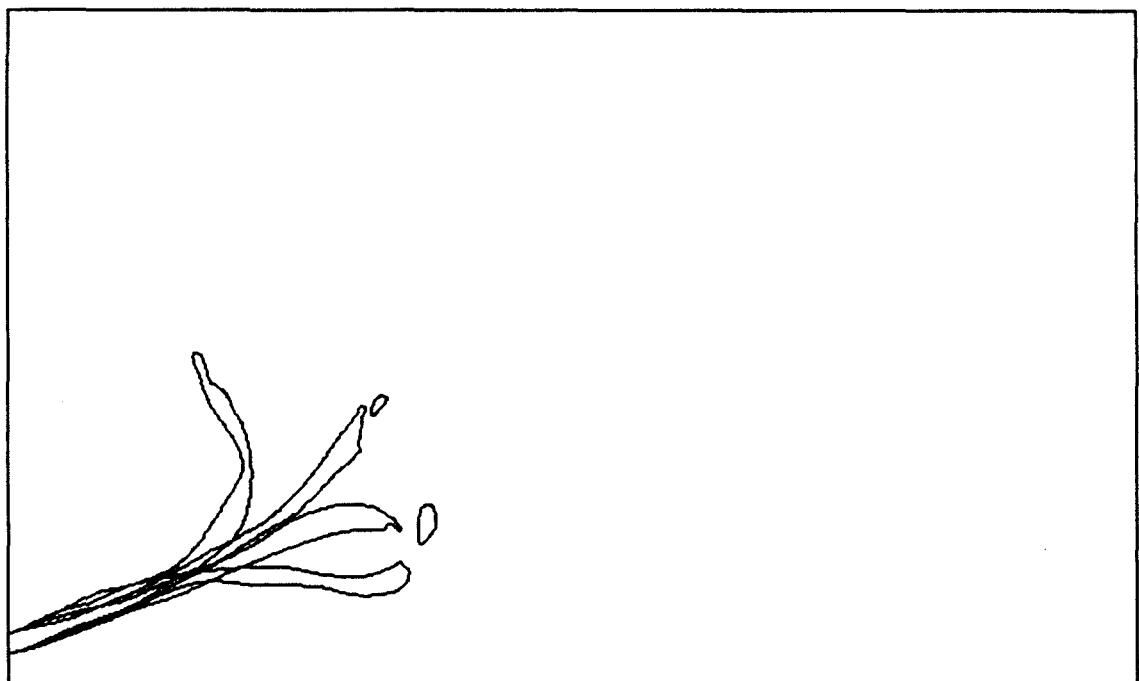
Figure 11. The position of the jet channel for the $\eta = 0.1$ hollow conical jet showing its varying position with time. The channel is defined by the region in which the total gas velocity is greater than 60% of the initial jet velocity. Contours are shown at $t/t_i = 7.9$, 9.0, 10.1 and 11.3. The radial oscillations in the jet channel are clearly visible.

Chapter 7

The Tidal Disruption of a Star

by a Massive Black Hole

Charles R. Evans,
and
Christopher S. Kochanek

Theoretical Astrophysics 130-33,
California Institute of Technology,
Pasadena, CA 91125, U.S.A.

Abstract

We describe a three-dimensional, numerical calculation of the tidal disruption of a low mass main sequence star, with mass M_* and radius R_*, on a parabolic orbit around a massive black hole ($M_h = 10^6 M_*$) with pericentric separation $R_p = 10^2 R_*$. The post-disruption evolution is followed until hydrodynamic forces become negligible and the liberated gas becomes ballistic. The distribution of these orbits is analyzed. About half of the debris escapes from the hole with velocities up to $7200(M_*/R_*)^{1/2}$ km s^{-1} (M_* and R_* in solar units), while half remains bound to the hole. We give the rate at which bound mass returns to pericenter after orbiting the hole once. We discuss several processes that determine the timescale to circularize the debris orbits and allow an accretion torus to form. This timescale and the timescales for radiative cooling and accretion inflow determine the onset and duration of the subsequent flare in the AGN luminosity.

1. Introduction

The tidal disruption of stars that pass sufficiently close to a massive central black hole ($M_h \simeq 10^6 M_\odot$) provides one mechanism by which low luminosity active galactic nuclei (AGN) might be fueled. While other processes in dense galactic clusters, including star–star collisions, may provide more efficient average fueling (Hills, 1975, 1978, Frank, 1978, Young et al., 1977), the disruption of a star by a massive black hole should lead to an intense, but brief accretion flare whose signature might reveal the presence of the black hole (Rees, 1988, Phinney, 1989). Crucial aspects of the physics of stellar disruption were first understood by Lacy et al. (1982); reviews of the scenario are given by Rees (1988) and Phinney (1989).

Rees (1988) emphasizes that there are several features of stellar disruption that require further analysis: (1) what is the orbital distribution of the stellar gas after disruption, (2) how is the accretion torus formed, (3) how radiatively efficient is the accretion process and how long does the flare last, and (4) how do the above factors depend upon the closeness of the approach? In this paper we address the first of these questions, and some aspects of the second and third, by modeling in detail a typical star-disrupting encounter.

The strength η of a tidal encounter is defined to be the square root of the ratio between surface gravity and tidal acceleration at pericenter:

$$\eta = \left(\frac{R_p^3}{GM_h R_*} \frac{GM_*}{R_*^2} \right)^{1/2} , \qquad (1)$$

where R_p, R_*, and M_* are the pericentric distance, stellar radius, and stellar mass, respectively. The quantity η is also the ratio between the duration of the encounter and the hydrodynamic timescale (Press and Teukolsky, 1977).

We model the star as a $\Gamma = 5/3$ polytrope, structurally similar to low mass main sequence stars. Carter and Luminet (1983, 1985, Luminet and Carter, 1986) find that $\Gamma = 5/3$ polytropes disrupt when $\eta \lesssim 1.2$. For somewhat larger pericentric separations, the star avoids total disruption but may be captured by the hole, and lose part of its envelope, as orbital energy is removed to excite oscillations in the star. Very close encounters ($\eta \lesssim 0.1$) may be qualitatively different due to relativistic effects and compression into the orbital plane before disruption. Carter and Luminet suggest that the compression may trigger significant nuclear energy release in the core, modifying the post-disruption dynamics. The range $1.2 \gtrsim \eta \gtrsim 0.1$ encompasses 80% of the disruptive encounters if the loss-cone is full. A lower overall rate of stellar disruption occurs, but with a higher fraction near $\eta = 1$, if the loss-cone is partly depleted and stars diffuse in at the edge due to two-body encounters (Young et al., 1977, Ozernoy and Reinhardt, 1978) or, more importantly, due to collective effects (Begelman et al., 1980, Roos, 1981, Norman and Silk, 1983) in the cluster.

In this paper we take the mass ratio $M_h/M_* = 10^6$ and consider only the case $\eta = 1$. We define $R_p/R_* \equiv \kappa$ so that for this case $\kappa_2 \equiv 10^{-2}\kappa = 1$. The use of a polytrope allows the results to be scaled with the stellar mass and radius for fixed η and κ. We measure M_* and R_* in solar units throughout. The pericentric distance is $24\,\eta^{2/3}(R_*/M_*)(M_*/M_6)^{2/3}$ times the black hole's Schwarzschild radius, which allows a Newtonian treatment of the black hole tidal field for $M_6 \equiv 10^{-6}M_h \lesssim 1$ and $\eta \sim 1$. With more massive holes, $M_h \gtrsim 10^8 M_\odot$, main sequence stars do not tidally disrupt without also plunging through the event horizon.

Tidal disruption has previously been difficult to model due to its three-dimensional (3D) nature. The affine model of Carter and Luminet (1983, 1985, Luminet and Carter, 1986) approximates the star as a time-dependent ellipsoid, with a fixed density profile. Such a model is of limited utility if the true matter distribution deviates from the assumed shape and density profile. However, a recently introduced numerical hydrodynamic technique, called smooth particle hydrodynamics (SPH), makes no restrictive assumptions on the hydrodynamic degrees of freedom and allows us to study the 3D encounters with full generality.

2. Numerical Method

The SPH method has been described elsewhere (Gingold and Monaghan, 1977, Benz et al., 1988, Hernquist and Katz, 1988). Our implementation is most similar to that of Hernquist and Katz (1988) in that a tree-based algorithm is used for both the self-gravity and hydrodynamic calculations, although the tree structure differs from that of Hernquist and Katz. We will describe the details of the method and the results of a set of tests elsewhere. Here we describe an axisymmetric analogue of the 3D tidal encounter for which we compare simulations using both SPH and finite difference hydrodynamics (FDH) (Evans, 1986). In this test the tidal gravity is given by the *axisymmetric* quadrupolar potential $GM_h R(t)^{-3} r^2 P_2(\cos\theta)$. The radius $R(t)$ is chosen to be the radial position of a star along the $\eta = 1$ parabolic orbit. We find that the axisymmetric tidal encounters are qualitatively similar to the true 3D encounters; in particular, the star disrupts when $\eta \lesssim 1$. Figure 1 shows the mass distribution of specific energies in the gas following an $\eta = 1$ axisymmetric disruption. The detailed agreement of the SPH and FDH methods on this non-trivial, multi-dimensional model for disruption demonstrates the capabilities of the SPH algorithm for problems of this type.

We calculate the disruption in the accelerating (but nonrotating) frame given by the initial parabolic approach orbit. In this frame the star's center of mass remains nearly at rest, allowing the post-disruption expansion of the gas to dominate the calculated dynamics. With 4×10^4 particles per star, energy is conserved to within 15% of the binding energy (0.2% of the expansion energy!), and angular momentum is conserved to one part in 10^5.

3. Hydrodynamics of Disruption and Gas Orbits

The cluster velocity dispersion σ gives incoming stars small specific energies, $|\epsilon_{init}| \simeq 5 \times 10^{13}\sigma_{100}^2$ ergs g^{-1}. Tidal disruption reduces the orbital energy by the binding energy of the star, $\epsilon_b \simeq 2 \times 10^{15}(M_*/R_*)$ ergs g^{-1}, which is ~ 40 times ϵ_{init}. Consequently, a parabolic approach orbit is assumed with stellar radius to pericentric separation ratio $R_p/R_* \equiv \kappa = 10^2$, tidal parameter $\eta = 1$, and mass ratio $M_h/M_* = \eta^{-2}\kappa^3 = 10^6$. For comparison, the specific kinetic energy at pericenter is $\epsilon_o = GM_h R_p^{-1} = 2 \times 10^{19}\kappa_2^2\eta^{-2}(M_*/R_*)$ ergs g^{-1}, or of order 10^4 times $|\epsilon_b|$.

The tidal bulge raised on the star by the black hole becomes an order unity distortion near pericenter. The bulge rotates in an attempt to stay aligned with the direction to the black hole but lags by $\simeq 60°$ at pericenter (see Figure 2). The magnitude of the tidal bulge, the phase lag and the transient nature of the encounter provide a tidal torque on the star. The specific torque, of order $GM_h R_*^2 R_p^{-3}$, acts over the duration of the encounter, $t_o \equiv R_p^{3/2}(GM_h)^{-1/2}$, to give a specific spin angular momentum of $\ell_s \sim 10^{-4}\kappa_2^{-2}\ell_o$, where $\ell_o = (2GM_h R_p)^{1/2}$ is the orbital angular momentum. The surface velocities implied by

such a rotation rate are $v_s \sim \eta^{-1} v_e$, where $v_e = (2GM_*/R_*)^{1/2}$ is the stellar escape velocity. The large surface velocities and the order unity tidal bulge combine to overcome the stellar self-gravity and lead to the disruption of the star (see Figure 2).

The variation of specific energy in the released gas is determined by the relative depth of a mass element in the potential well of the black hole. The spread in specific energy of the gas, $\Delta\epsilon$, is given by the change in the black hole potential across a stellar radius (Lacy et al., 1982):

$$\Delta\epsilon \simeq (GM_h/R_p)(R_*/R_p) \simeq 10^{-2}\kappa_2^{-1}\epsilon_o. \tag{2}$$

This is much larger than the binding energy, ϵ_b, and kinetic energy, $\frac{1}{2}v_s^2$, generated by spin-up near pericenter. In order of magnitude we have $\epsilon_o \gg \Delta\epsilon \simeq 10^{-2}\kappa_2^{-1}\epsilon_o \gg \epsilon_b \simeq 10^{-4}\kappa_2^{-2}\eta^2\epsilon_o$. The mean orbital binding energy of the stellar debris is ϵ_b, but since $\Delta\epsilon \gg \epsilon_b$ very nearly half of the mass of the star remains bound to the hole, and half is ejected (Lacy et al., 1982) with estimated velocities of $v_\infty \sim (2\Delta\epsilon)^{1/2} = 6000\eta^{-1}\kappa_2^{1/2}(M_*/R_*)^{1/2}$ km s^{-1}. From the simulation ($\eta = 1$, $\kappa_2 = 1$) we find the peak velocity (see Figure 3) in the ejecta to be $7200(M_*/R_*)^{1/2}$ km s^{-1} and the total kinetic energy of the unbound material is $8.7 \times 10^{49} M_*^2/R_*$ ergs.

The kinetic energy of relative expansion rapidly dwarfs both the adiabatically decreasing internal energy and the stellar self-gravity. The freezing of the energy and angular momentum distributions locks the debris into Keplerian trajectories around the black hole, until such time as hydrodynamic forces again become important. We expect the mass distribution to have a spread in energy of $\simeq 2\Delta\epsilon$, leading to an estimated mass distribution $dM/d\epsilon \simeq M_*/2\Delta\epsilon \simeq 5.2 \times 10^{15}\eta^{-2}\kappa_2^{-1}R_*$ g^2 ergs^{-1}. In the simulation, we observe that the spread in energy is $1.6\Delta\epsilon$, and that the mass distribution is nearly constant, with $dM/d\epsilon \simeq 5.8 \times 10^{15}R_*$ g^2 ergs^{-1} (see Figure 3). The most tightly bound gas has total energy $\simeq -\Delta\epsilon$, which makes the minimum period before return to the hole $T_m = 2^{-1/2}\pi\kappa_2^{3/2}t_o \simeq 0.1\,\eta\kappa_2^{3/2}R_*^{3/2}M_*^{-1/2}$ yr. Using the Keplerian relation $d\epsilon/dT = (1/3)(2\pi GM_h)^{2/3}T^{-5/3}$ and that $dM/d\epsilon$ is nearly constant, the estimated rate at which mass returns to the black hole after one post-disruption orbit is

$$\frac{dM}{dt} \simeq \frac{1}{3}\frac{M_*}{T_m}\left(\frac{t}{T_m}\right)^{-5/3} \simeq 3.3\left(\frac{t}{T_m}\right)^{-5/3}\eta^{-1}\kappa_2^{-3/2}\left(\frac{M_*}{R_*}\right)^{3/2} M_\odot \text{ yr}^{-1}. \tag{3}$$

Figure 4 displays the mass infall rate obtained by the simulation, which confirms this power law dependence [note the error (Phinney 1989) in Rees (1988)]. The simulation shows that the first material returns slightly before T_m, and that the peak rate, at $t \simeq 1.5T_m$, is $1.4(M_*/R_*)^{3/2}M_\odot$ yr^{-1}. The debris returns at a rate exceeding the Eddington rate $\dot{M}_{edd} \simeq 0.02\,\varepsilon_{0.1}^{-1}M_6\,M_\odot$ yr^{-1}, for radiative efficiency ε, for $\simeq 2.2\varepsilon_{0.1}^{3/5}R_*^{3/5}M_*^{-1/5}$ years in this $\eta = 1$, $\kappa_2 = 1$ case.

4. Post-Disruption Behavior and Formation of an Accretion Torus

The returning gas does not immediately produce a flare of activity from the black hole. First a sufficient fraction of the material must enter circular orbits and form an accretion torus. Only then will viscous effects release enough binding energy to power an Eddington-level emission flare ($L_{edd} = 1.3 \times 10^{44}M_6$ ergs s^{-1}). It is difficult to assess how rapidly the torus forms once stellar debris begins to return to pericenter, but our knowledge of the distribution of debris orbits allows us to outline several potentially important processes.

As the star disrupts, thermal pressure accelerates material out of the orbital plane to a velocity $\lesssim v_e$ before the expansion, (which predominantly occurs within the orbital plane) adiabatically cools the gas. The predicted (and observed) spread in inclination angles is $\delta i \sim v_e/v_o \sim 0.01\eta\kappa_2^{-1}$, where v_o is the velocity at pericenter. These orbits are focused back into the original orbital plane at apocenter. There the convergence velocity is $\sim v_a\delta i$, where v_a is the orbital velocity at apocenter. This causes the formation of a pancake shock that weakly redistributes the orbital parameters and damps out some of the vertical motion.

A second effect occurs as the stream approaches pericenter, where the radial focusing of orbits acts as an effective nozzle. During the disruption adiabatic expansion and general relativistic precession produce relative changes in the apsidal angles and eccentricities of the debris orbits. Hydrodynamic effects dominate and produce, for example, an observed spread in apsidal angles of $\simeq 1.4°$. If these hydrodynamic effects were absent, the highly sheared stream would return to pericenter with a width as narrow as $h \simeq \sqrt{2}(1-e)^{5/2}R_*/(6\pi)$, or $h \sim 4 \times 10^{-6}R_*$ for $e = 0.98$. In this case Liouville's theorem guarantees that the material returns to its original density at pericenter. A spread in eccentricities, δe, about the mean eccentricity, e, implies a widening of the stream, estimated to be $h \sim R_p\delta e/(1-e)$. At the peak of the re-infall rate, the spread in the eccentricities has an upper limit of $\delta e \lesssim 10^{-4}$. For $e = 0.98$ we then find a limit on the width $h \lesssim R_*/2$. Spreading by this amount reduces the density of gas returning to pericenter to $\gtrsim 10^{-5}$ of the original stellar density. The flow converges with Mach number $M \simeq h\nabla \cdot V/c_s \sim \eta^{-1}(h/R_*)(\rho/\rho_*)^{-1/3}$ where ρ_* is the mean density of the original star. For our orbital parameters this implies a Mach number of $\lesssim 10$ near the peak of the mass infall rate. If the Mach number is high enough, oblique shocks will form in the stream that will alter the orbital velocities and help circularize the orbits.

A third effect is due to relativistic precession. As the debris passes the black hole, its apsidal angle precesses by $\Delta\phi = 3\pi GM_hR_p^{-1}c^{-2}$, for a Schwarzschild black hole. After passage the outflowing gas is on an orbit that will collide with the infalling stream near apocenter, giving rise to an angular momentum redistributing shock much like those in cataclysmic variable systems. The relative velocities are of order $\Delta v/|v| = e\Delta\phi/((1-e)^2 + (\Delta\phi/2)^2)^{1/2}$. For $\eta = 1$ orbits around a $M_6 = 1$ black hole ($\Delta\phi \simeq 10°$), the redistribution is fairly weak and several passes are required to circularize the orbits. More efficient redistribution occurs if the precession is larger (*i.e.*, for a more massive black hole or for a smaller pericentric separation). On the other hand, if the black hole is a Kerr black hole and the orbital plane is not perpendicular to the rotation axis, the outgoing stream may miss the incoming stream due to precession of the orbital plane about the black hole's spin axis.

Debris will ultimately settle into a torus about the black hole. The angular momentum of the infalling gas corresponds to that of a circular orbit with radius $r_K = \sqrt{2}R_p$. The initial binding energy of the debris, however, is smaller than the binding energy of the circular orbit by $\sim \kappa^{-1}$. Hence the torus must extend to radii much larger than r_K for its binding energy to be consistent with that of the debris. Near r_K the Thomson optical depth is extremely high, $\tau \sim M_*\sigma_T/R_K^2m_p \sim 10^7M_*R_*^{-2}\kappa_2^{-2}$, so that if material is added on a timescale shorter than the cooling timescale the torus will be radiation pressure supported. Once the torus is formed, it will evolve under the influence of viscosity, radiative cooling, winds and the time-dependent mass inflow.

The three primary timescales are the orbital circularization timescale, the radiative cooling timescale, and the viscous timescale. Each of the three hydrodynamic redistribu-

tion processes we discussed above requires at least several orbits to produce circularization (for $\eta \sim 1$ and $M_6 \sim 1$). Thus the timescale for forming the torus is $\sim 10\, T_m \sim 1$ year. If the torus radiates at a fraction f of Eddington luminosity, the Kelvin-Helmholtz cooling time for an extended torus of mass M is $(GM_h M/R_K)/f L_{\rm Edd} \sim 3 f^{-1} M/M_\odot$ years. If the formation of the torus proceeds via hydrodynamic effects that are stronger than those outlined above (and such effects are difficult to rule out), then a radiation pressure supported torus forms. However if the mechanisms we have outlined are indicative of the timescale for circularizing the orbits, then the cooling operates at a competitive rate and the qualitative structure of the torus (i.e., thick or thin) may vary with radius (Begelman 1985). The viscous timescale near r_K is $t_v \sim \alpha^{-1} t_o$ where the viscosity parameter α is only weakly constrained, $10^{-6} \lesssim \alpha \lesssim 1$. If $t_v \ll 1$ year the torus is drained as fast as it forms, while if $t_v \gg 1$ year a long-lived torus of mass $\simeq M_*/2$ builds up. In the intermediate range, with $\alpha \sim 10^{-4}$ and $t_v \sim 1$ year, all three timescales are comparable. Finally, to understand the evolution of such an accretion torus, plus its associated AGN flare, the time dependence of the external mass inflow rate must also be counted.

5. Conclusions

We provide numerical details of how a low mass star is tidally disrupted during an encounter with a massive black hole, examine the distribution of debris orbits and calculate the rate at which debris returns to the vicinity of the hole. The numerical results closely match analytic estimates for the case considered here. This is less likely to be true for either more distant encounters in which the star does not fully disrupt, or for closer encounters in which shocks develop as the star is vertically compressed. The post-disruption evolution depends upon the balance between the timescales for orbital circularization, cooling, viscous accretion and debris infall. A full understanding of the circularization process will require detailed hydrodynamic calculations.

Acknowledgements

References

Begelman, M.C., Blandford, R.D., and Rees, M.J. 1980, *Nature,* **287**, 307.

Begelman, M.C., 1985, in *Astrophysics of Active Galaxies and Quasi-stellar Objects,* J.S. Miller, ed., (University Science Books: Mill Valley) p. 411.

Benz, W., Bowers, R.L., Cameron, A.G.W., Press, W.H., 1988, *CFA preprint.*

Carter, B., and Luminet, J.P., 1983, *Astr. Astrophys.,* **121**, 97.

Carter, B., and Luminet, J.P., 1985, *Mon. Not. R. astr. Soc.,* **212**, 23.

Evans, C.R., 1986, in *Dynamical Spacetimes and Numerical Relativity,* ed. J. Centrella, (Cambridge Univ. Press: Cambridge), p. 3.

Frank, J., 1978, *Mon. Not. R. astr. Soc.,* **184**, 87.

Gingold, R.A., and Monaghan, J.J., 1977, *Mon. Not. R. astr. Soc.,* **181**, 375.

Hernquist, L., and Katz, N., 1988, *IAS preprint.*

Hills, J.G., 1975, *Nature,* **254**, 295.

Hills, J.G., 1978, *Mon. Not. R. astr. Soc.,* **182**, 517.

Lacy, J.H., Townes, C.H., and Hollenbach, D.J., 1982, *Astrophys. J.,* **262**, 120.

Luminet, J.P., and Carter, B., 1986, *Astrophys. J. Suppl.,* **61**, 219.

Norman, C.A., and Silk, J., 1983, *Astrophys. J.,* **266**, 502.

Ozernoy, L.M., and Reinhardt, M., 1978, *Astrophys. Sp. Sci.,* **59**, 171.

Phinney, E.S., 1989, to appear in the Proc. IAU Symposium 136 *The Galactic Center,* ed. M. Morris, (Kluwer Academic, Dordrecht).

Press, W.H., Teukolsky, S.A., 1977, *Astrophys. J.,* **213**, 183.

Rees, M.J., 1988, *Nature,* **333**, 523.

Roos, N., 1981, *Astr. Astrophys.,* **104**, 218.

Young, P.J., Shields, G.A., and Wheeler, J.C., 1977, *Astrophys. J.,* **212**, 367.
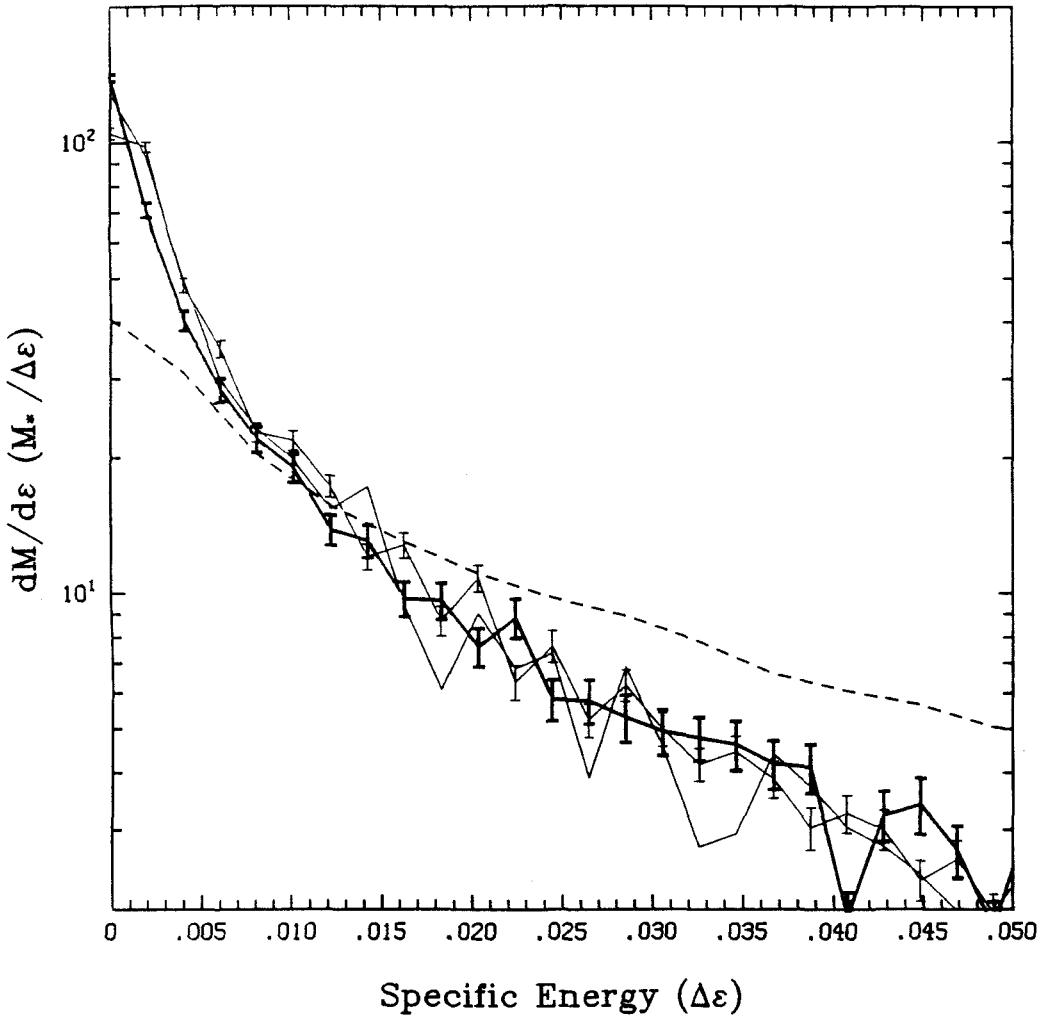
Figure 1: Mass distribution of specific energies in the liberated gas that results from tidal disruption in the axisymmetric test problem at time $t = 1.06$ (in units of dynamical time $t_o = (R_p^3/GM_h)^{1/2}$ at pericenter). The SPH technique, which is used for the three-dimensional calculations, is compared in this two-dimensional test to a finite difference hydrodynamics (FDH) calculation, and to the quasi-analytic Carter and Luminet model. SPH resolutions of 10^4 and 2×10^4 particles per star are shown (light solid lines), as is a FDH run with a resolution of 150 radial zones by 45 angular zones (heavy solid line). The Carter and Luminet model (dashed line) shows strong deviations from the results of the numerical models. The error bars accompanying the SPH results indicate statistical fluctuations induced by binning the particles.
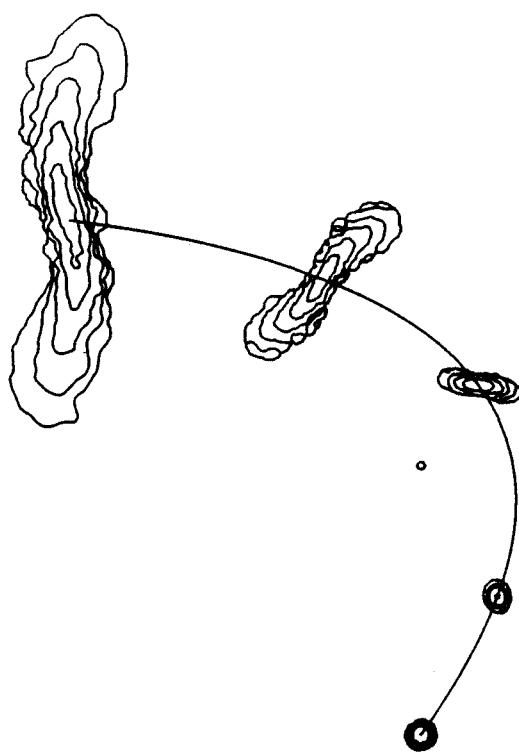
Figure 2: Density contour plots of star and liberated gas at several points along the orbit. **The apparent size of the star is increased by a factor of 15.** The star is shown at times $t = -3.33, -1.44, 0.46, 2.36$, and 6.15 in units of t_o. The small circle at the center indicates the Schwarzschild radius of the black hole.
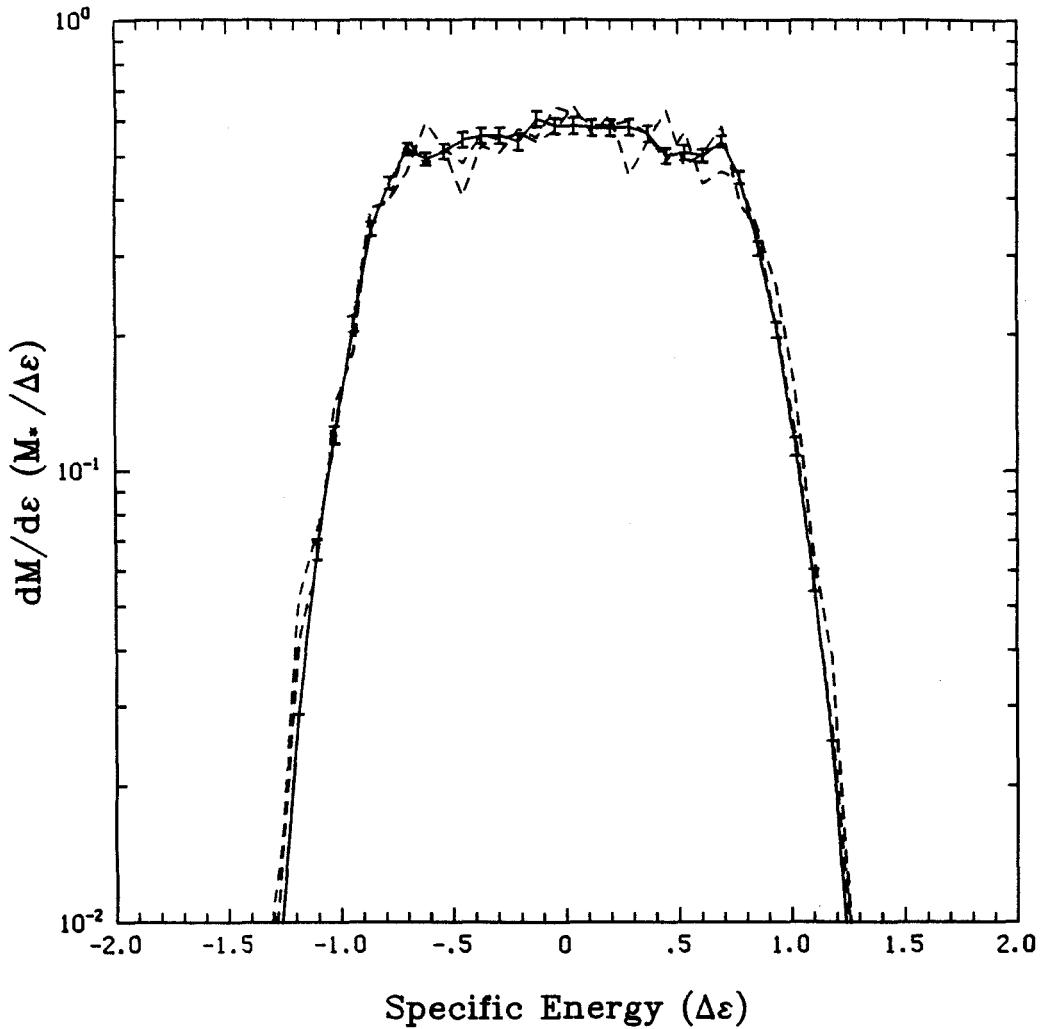
Figure 3: Differential mass distributions with respect to specific energy for debris in the 3D disruption simulation. The energy is measured in units of $\Delta\epsilon \equiv \epsilon_o \kappa^{-1}$. The results for three particle resolutions of 10^4, 2×10^4, and 4×10^4 particles per star are shown. The error bars indicate statistical fluctuations induced by binning the particles.
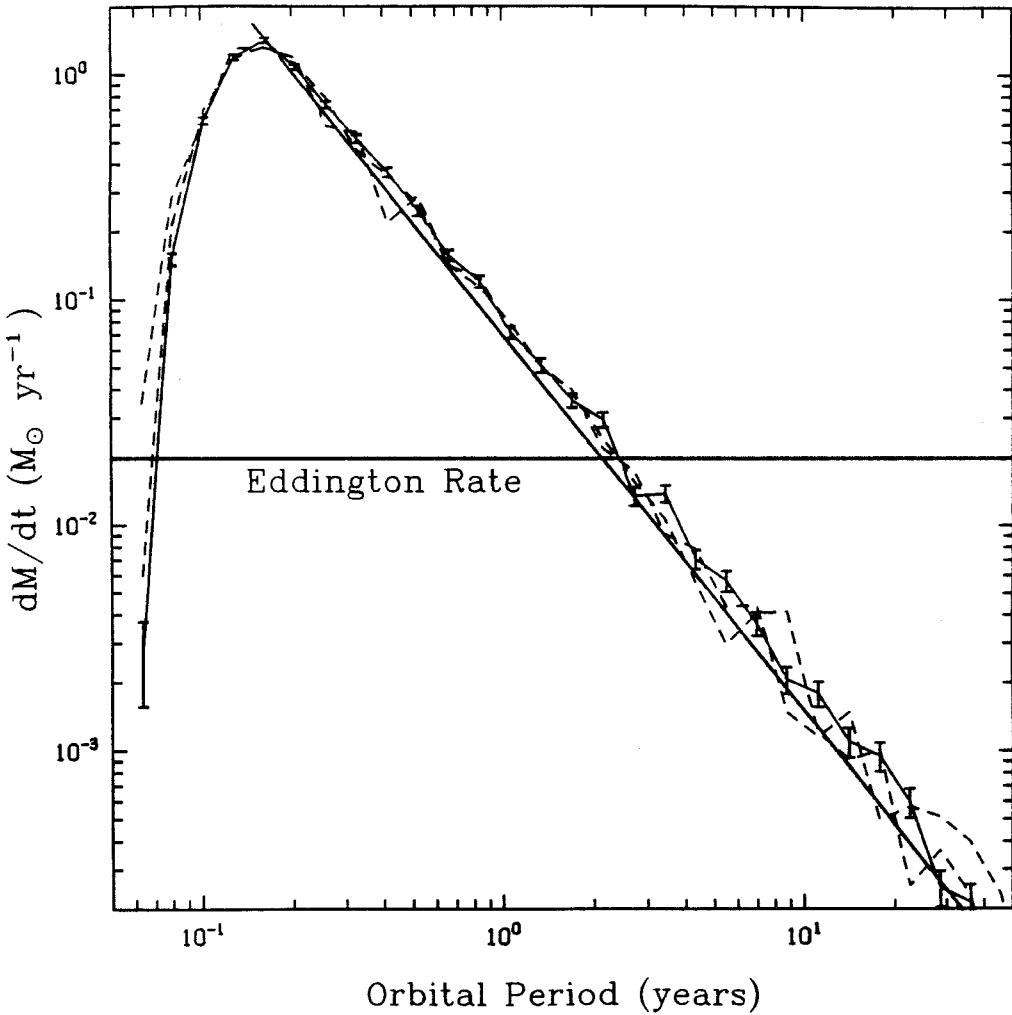
Figure 4: The rate at which stellar debris returns to the vicinity of the black hole versus time. The infall rate is given in units of solar masses per year. The Eddington accretion rate for a $M_h = 10^6 M_\odot$ black hole, with radiative efficiency $\varepsilon = 0.1$, is indicated. The heavy line shows the estimated infall rate given by equation (3) in the text. The results from three different SPH simulations are shown, with particle resolutions of 10^4, 2×10^4, and 4×10^4 particles per star. The error bars indicate statistical fluctuations induced by binning the particles.